AD NUMBER

ADB951468

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; 09 JUL 1946. Other requests shall be referred to Adjutant General's Office [Army], Washington, DC.

AUTHORITY

AGO ltr, 13 Nov 1979

PRS REPORT 717

# COMPARISON OF THE RATING CHECK LIST (RCL) AND FORCED CHOICE LIST (FCL) METHODS OF OBTAINING RATINGS

SEPTEMBER 1945.

PJ-4073.



PERSONNEL RESEARCH

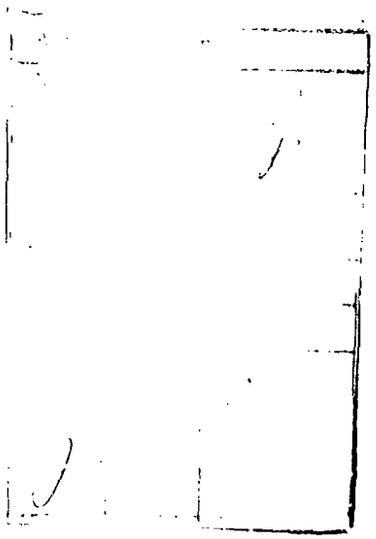
SECTION

PR AND P BR--PRRS BUR--AGO

## 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.

Reprint date 19 Apr 51
PR-4073



PRS Report No. 717
PRS-SSU
PR-4073
EAR/ss
9 July 1946

# COMPARISON OF THE RATING CHECK LIST (RCL) AND FORCED CHOICE LIST (FCL) METHODS OF OBTAINING RATINGS
September 1945

## I. THE PROBLEM

In the development of the Officer Evaluation Report (OER) for the officer retention program, two types of item showed particular promise--the RCL, or rating check list, and the TCL, or three-choice list. The RCL consists of descriptive phrases such as "commands the respect of his men" to which the rater responds by indicating on a five-point scale the degree to which the man rated displays the trait. The TCL is a forced choice technique, identical in principle with the self-descriptive items used in the Biographical Information Blank (see PRS Report No. 703). As used in OER, three adjectives were grouped and the rater responded to all three terms, choosing one as most descriptive, one as least descriptive, and one as intermediate in its application to the man rated. In the construction of an Officer Efficiency Report, FCL-3, the TCL form was modified by the use of a different method of scaling (see PRS Report No. 679, Tab J in "Report and Recommendations, Project PR-4073") and by utilizing four items in each group. The term "forced choice list" (FCL) was used to describe the items. In the present program, FCL will be used to refer to the general forced choice form of item.

The major purposes of this paper are to compare the validity of the RCL and FCL methods of obtaining ratings and to determine whether there is any advantage in combining the two techniques. The RCL form is well liked by raters, but it suffers the disadvantage that it is possible for the rater to see how he is marking the ratee, making it possible for him to raise or lower the final score at will. The FCL form is less well liked by the rater, but it becomes difficult for him to determine where he is placing the ratee in the scale represented by the total score.

Combining the two techniques (as described in Section II below) might make the scale better liked than FCL by the rater. Such a combination should permit obtaining a maximum number of responses to a given group of items with consequent gain in efficiency and reliability. It is also possible that making RCL judgments prior to FCL judgments might improve the latter. Combining the two forms has the disadvantage of increasing the time to complete the rating. Whether this expenditure of extra time is warranted depends on the degree to which validity is increased over either technique used alone.

## II. CONSTRUCTION OF THE RATING FORMS

### A. The Military Report Form, MR, OCS-1.

The RCL and FCL techniques were first combined in MR-OCS-1, constructed for use in a study of procedures for the selection of officer candidates. The results of its use in this connection are described in PRS Report No. 711. Here only its method of construction will be reviewed as a means of clarifying the design of the present study. In MR -OCS-1, adjectives and descriptive phrases were arranged in groups of five as in the following typical item:

Group X

1) attentive
2) easily confused
3) listless
4) responsible
5) sincere

Most descriptive
Least descriptive

The rater is first instructed to consider each adjective or phrase separately, and indicate how well it applies to the man rated, using this scale:

KEY
1. Applies to an EXCEEDINGLY HIGH degree
2. Applies to an UNUSUAL or OUTSTANDING degree
3. Applies to an AVERAGE degree
4. Applies to a LIMITED degree
5. Applies to a SLIGHT degree or NOT AT ALL

This is the RCL technique. After indicating his description on the report form or answer sheet for the five adjectives in a group, the rater then indicates which of the five is most descriptive of the man rated and which is least descriptive (the FCL technique). In all, seven responses are obtained for each group, five RCL responses and two FCL responses.

The general technique for constructing a group of phrases is described in Sep studies No. 702 and 703. Essentially, it consists in grouping two pairs and a "neutral" item. The pair is obtained by bringing together two adjectives which, in theory at least, appear of equal merit or value to the rater but differ with respect to prediction of officer success. Hence, he can describe a man equally favorably by the selection of either but he finds it difficult to detect which is more favorable with respect to officer success. The likelihood is thereby reduced that he can deliberately mark the blank to increase or decrease the total score.

In the Military Report, one relatively desirable pair and one relatively undesirable pair were selected. The "neutral" term was selected to be between the two pairs with respect to its appearance of merit and also with respect to its ability to predict officer success.

The scale values needed in constructing the groups were obtained from PRS Report No. 702 and a supplement to Report No. 711 which is in the confidential files.

Three further considerations were involved in construction of the groups: (1) no adjective was used more than once, thus limiting the number of well constructed pairings which could be made; (2) no adjective or phrase was used which could not be applied to enlisted men; and (3) effort was made to include in each group an adjective or phrase related to one of the three factors which emerged repeatedly in the studies leading to the construction of OER: attention to duty or conscientiousness, leadership or force, and stability. This latter principle could not be followed completely, but was utilized to the maximum possible degree.

Five items were included in each group, rather than fewer, since experience with the Biographical Information Blank had suggested that the inclusion of a neutral item increases the probability of finding discriminating items.

Thirty-eight such groups were constructed and included as Section I of the Military Report. Section II consists of an over-all rating of competence obtained by ranking the ratee in relation to a typical group of 20 newly commissioned second lieutenants.

Since the Military Report was to be collected by mail from non-commissioned and commissioned officers and a simplified procedure was considered desirable, an answer sheet was not used.

B. The Officer Efficiency Report, OER-B.

After the Military Report was constructed, it was decided to adapt it for experimental use in the study of Officer Efficiency Reporting Procedures, PR-4073. It was considered desirable to undertake preliminary evaluation of this technique for officer efficiency reporting to determine whether future work should utilize the combined RCL-FCL techniques.

Adapting the Military Report required minor changes in directions--elimination of references to enlisted men, substitution of references to officers, change of the reference group for the over-all rating from newly commissioned second lieutenants to Army officers in general, and adaptation of the procedure to machine scoring.

C. Officer Evaluation Reports, OER-C and OER-D.

To obtain a direct comparison between RCL and FCL techniques and between each of these and the RCL-FCL technique, with item content controlled, OER-C and OER-D were prepared. Section I of both OER-C and OER-D contains the same items (adjectives or descriptive phrases); in OER-C they are grouped as in OER-B, the difference being that in OER-C the rater responds only to the forced choice situation (the FCL technique); in OER-D the items are likewise the same, but the rater responds only by the RCL technique. OER-C and D were

adapted for machine scoring. Section II, the over-all rating, was included in OER-B, but not in OER-C or OER-D. It was inadvertently omitted from the latter two forms. The error is not serious since it is Section I that it is desired to compare.

## III. POPULATIONS AND GENERAL PROCEDURE

The data for this study were collected as an incidental duty of the field representatives conducting the rater training study under Project PR-4073, the officer efficiency report project. The installations and groups of officers involved are described in PRS Report No. 674. The populations for this study consist of

364 officers for whom OER-B was completed

111 officers for whom OER-C was completed

123 officers for whom OER-D was completed.

The above groups do not overlap. No officer was rated on more than one form, nor did any rater use more than one form.

The criterion consisted of ratings by associates secured as described in PRS Report No. 670. The ratings were combined into the criterion index described in the study cited.

## IV. DEVELOPMENT OF SCORING KEYS

It will be recalled that OER-B utilizes a combination RCL-FCL technique; OER-C, the FCL technique; and OER-D, the RCL technique. It is assumed for this study that keys developed on the basis of responses to OER-B can be applied to OER-C and OER-D and that, allowing for the inevitable differences in length of OER-B, OER-C, and OER-D any differences in validity and owing to the method of rating employed.

The group rated with OER-B, as noted above, consisted of 364 officers, practically all of company grade. Sixty-four (64) of the answer sheets were removed by a systematic method after all answer sheets were arranged in order of the criterion index score. This group was used for cross-validative purposes. The remaining 300 were utilized for an alternative analysis (100 high, 100 middle, and 100 low). The standard of item selection for scoring were (a) a sliding scale of percent of those rated who were descriminated by a particular alternative, the percent scale used being the same as that described in PRS Report No. 703 and (b) the alternative differentiated the high and middle, and the middle and low groups in the same direction.

Three-level (+, 0, or -) or two-level (+ or -) scoring was utilized according to the break which would yield the highest correlation of the item with the criterion. These "r's" were not computed, but their relative size for

different points of cut was estimated from the distribution of high-low differences. The resulting key will be referred to as the statistically determined key.

A key was also developed on the basis of scale values of the alternatives. These scale values are given in PRS Report No. 702 and the supplement to Report No. 711. This key will be referred to as the Predetermined Key.

## V. RESULTS

A. Comparison of Validities of Predetermined and Statistically Determined Keys in OER-B 1. The scores compared are the FCL score, the RCL score and the total of the two scores. The correlation of these scores with the criterion index for the 300 officers used in the alternative analyses and the 64 officers in the cross-validation group are

| | Predetermined Key | | Statistically Determined Key | |
|---|---|---|---|---|
| Criterion Index Correlated with: | Alternative Analysis Gp | Cross-Valid. Gp | Alternative Analysis Gp | Cross-Valid. Group |
| FCL Score | .365 | .468 | .502 | .425 |
| RCL Score | .419 | .469 | .513 | .535 |
| RCL + FCL Score | .434 | .510 | .540 | .562 |

There are three features of special interest in the above tabulation. It is apparent that, while a key can be predetermined with considerable success, the statistically determined key is somewhat more valid. There is but one instance where the validity of the predetermined key exceeds that of the statistically determined key--the FCL score for the cross-validation group. While the predetermined key is not considered further in this study, the finding is of interest in indicating how well a key may be guessed from scale values.

The second finding of interest is the general superiority of the correlation in the cross-validation group for both keys (there is again but one exception). This implies that shrinkage is not occurring from the validation to the cross-validation group and that the differences between the two groups are to be attributed to sampling errors.

A third finding is that the summation of the RCL and FCL scores gives the highest validities for both groups and both keys. Whether this result indicates the desirability of utilizing both techniques simultaneously will be considered below.

B. Comparison of OER-B, C and D.

The major problem in this study was to compare the FCL, RCL, and a combination of the two techniques of obtaining ratings. The data were obtained incidentally during the training run of Project PR-4073 so the groups given

OER-B, C, or D could not be matched in grade and arm of service as would have been desirable, since the validity of ratings in the Army is known to be somewhat influenced by these factors. A comparison of the grade distribution for three groups is given below:

| | OER-B | | OER-C | | OER-D | |
|---|---|---|---|---|---|---|
| Major | 1 | 0.28 | | | | |
| Captain | 35 | 9.94 | 2 | 1.89 | 6 | 4.96 |
| 1st Lt | 189 | 53.69 | 56 | 52.83 | 61 | 50.41 |
| 2nd Lt | 127 | 36.08 | 48 | 45.28 | 54 | 44.63 |
| TOTAL* | 352 | 99.99 | 106 | 100.00 | 121 | 100.00 |

* The difference in N's between this and other tables is owing to information on grade being omitted for some cases.

The groups are predominately 1st and 2nd lieutenants. The OER-B group contains more captains. The difference is not considered sufficiently great to vitiate comparison between the three forms. In previous studies it has been found that combining officers of widely different grades has tended to reduce the validity. Hence the difference would tend to reduce the validity of OER-B.

The distributions by Arm or Service by these groups are

| Arm or Service | OER-B | | OER-C | | OER-D | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Air Corps | 33 | 9.65 | -- | -- | -- | -- |
| Cavalry | 7 | 2.05 | 2 | 1.94 | 3 | 2.70 |
| Chemical Warfare | 17 | 4.97 | 6 | 5.82 | 5 | 4.50 |
| Combat Military Police | 4 | 1.17 | -- | -- | 2 | 1.80 |
| Corps of Engineers | 25 | 10.23 | 15 | 14.56 | 14 | 12.61 |
| Dental Corps | 1 | 0.29 | 1 | 0.97 | 1 | .90 |
| Field Artillery | 48 | 14.03 | 21 | 20.39 | 17 | 15.32 |
| Infantry | 140 | 40.94 | 40 | 38.83 | 56 | 50.45 |
| Medical Administrative Corps | 2 | 0.58 | -- | --- | -- | -- |
| Medical Corps | 1 | 0.29 | -- | --- | -- | -- |
| Ordnance | 5 | 1.46 | 3 | 2.91 | 1 | .90 |
| Q.M. Corps | 4 | 1.17 | -- | --- | -- | -- |
| Signal Corps | 1 | 0.29 | -- | --- | -- | -- |
| Transportation Corps | 44 | 12.86 | 15 | 14.56 | 12 | 10.81 |
| TOTAL | 342 | 99.98 | 103 | 99.98 | 111 | 99.99 |

The groups are fairly comparable with respect to Arm or Service, being composed largely of Infantry, Field Artillery, Engineers, and Transportation Corps. The greatest difference is the absence of Air Corps from the OER-C and D. Since the validity of rating scales has tended to be lower for the Air Corps (probably owing to less acquaintanceship of the officers furnishing the criterion data), this difference would tend to reduce the correlations for the OER-B groups.

The validity for the OER-B, C and D forms can be compared only for the FCL and RCL scores since the over-all rating was not included in its OER-C and D forms. The correlations for the three forms are

| | OER-B | | | |
|---|---|---|---|---|
| | Item Analysis Group | Cross Validation Group | OER-C | OER-D |
| N | 300 | 64 | 111 | 123 |
| FCL Score | .502 | .425 | .466 | --- |
| RCL | .513 | .535 | --- | .345 |
| FCL + RCL Score | .540 | .562 | --- | --- |

The scoring keys for OER-C and OER-D were developed, as noted previously, from the responses to OER-B. Hence, the FCL score for OER-B and C is based on identical items as is the RCL score for OER-B and D. The FCL key contains 48 items; the RCL key, 93 items. On the basis of length of rating scale the RCL key would be expected to yield the highest criterion correlations. This is the case for the items scored within the OER-B form, although the difference for the larger group is slight (r's equal .502 for FCL score and .513 for RCL score). When the FCL form and the RCL form are given separately, however, its difference favors the FCL form (r's equal .466 for FCL and .345 for RCL). These findings suggest that when used alone the FCL technique is superior to the RCL technique. That the two techniques yield similar results when used in combination may well be owing to the forcing of more serious consideration of the individual items when it is known that a choice must be made after rating on each single trait.

Combining the two techniques yields the highest validity for both the item analyzed group (.540) and the cross-validation group (.562). The gain is considerable when compared with the validity for the RCL technique (.345) or for the FCL technique (.466) used alone, but negligible when compared with the validity for the RCL score obtained when the techniques are used in combination (.513 for the item analyzed group and .535 for the cross-validation group).

From the results thus far it appears that the combination of the RCL and FCL techniques is superior to either technique used alone. One further point needs investigation before judging whether this superiority is sufficient to warrant the additional time required to complete a scale utilizing the combination technique, namely, the validity of the scale with the over-all rating added.

The only available comparison that can be made is between the OER-B and the FCL-3. The latter scale contains 50 items of the FCL type and an over-all rating. Although the FCL-3 does not contain the same grouping of items as OER-B, the content covered is the same. Hence, the comparison, while not exact, is legitimate.

In the validation of FCL-3 (PRS Report No. 674) data on 1,284 2nd lieutenants, 2,015 1st lieutenants, and 1,299 captains were collected. These populations were combined with the grades weighted 36%, 54%, and 10% respectively, making the population comparable with that utilized for OER-B. The weighting is a rounding of the percentages given on page 6 of this report.

The criterion index vs. FCL-3 correlation for the weighted population is .49. The criterion index vs. OER-B total score is .60. This latter figure was obtained as follows: For 267 of the 300 cases in the item-analyzed group given OER-B, an over-all rating of exactly the same kind as used in FCL-3 was available. The means and sigmas for FCL-RCL score and criterion index are practically identical for the group of 300 and the group of 267, as shown below, indicating that no systematic selection of those who omitted the over-all rating.

| | N = 300 | | N = 267 | |
|---|---|---|---|---|
| | Mean | Sigma | Mean | Sigma |
| RCL - FCL Score | 96.18 | 37.32 | 96.39 | 37.66 |
| Criterion index | 28.61 | 9.19 | 28.71 | 9.07 |

The R of .60 between the criterion and FCL-RCL score and the over-all rating was calculated from the following matrix:

| N | FCL-RCL Score | Over-all Rating | Criterion Index |
|---|---|---|---|
| 267 | --- | .732 | .540 |
| 300 | | --- | .572 |
| 300 | | | -- |

Utilizing integral raw score weights the correlation is .598. The raw score weights calculated from above sample turned out to be identical with the determined for FCL-3a. The weights, like the scoring keys, are secret.

While the FCL-RCL technique (OER-B) is definitely superior to the FCL technique (FCL-3a) in this comparison, the difference in the number of items and consequent difference in rating time should be taken into account. Considering each point on the over-all rating scale as an item, there are 161 scored items. On the same basis there are 69 items in FCL-3. The comparison desired is between OER-B and FCL-3--2-1/2 times as long.

The reliability of FCL-3a was estimated as .82 on the basis of an r of .70 (somewhat lower than actually found in the study of FCL-3 cited above) between the two parts. With this reliability, the validity of an FCL-3a of 2-1/2 times present length is estimated as .52. The validity of OER-B (.60) is still superior.

## VI. DISCUSSION AND CONCLUSIONS

The results of this study indicate that, when both are used alone, the FCL technique is superior to the RCL technique, that a combination of the two techniques is superior to either used alone, and that utilizing the combination improves the validity of the RCL score. While the present study is not conclusive, owing to the limited number of cases, the improvement in validity brought about by the use of the combination technique is sufficient to warrant serious consideration.

The major objection to the use of the FCL-RCL technique is the considerably greater time required by the rater. Objections on the basis of the expenditure of time were frequently voiced. The objections were made the more vigorous by the feeling the rater developed of answering the same questions repeatedly.

The fact that the validity of RCL improves considerably when it is used in combination with FCL suggests a compromise position which would considerably reduce the time required, namely the use of an FCL section and an FCL section of different items, a procedure similar to that employed in the construction of the original Officer Evaluation Report (OER) developed for use in the male officer integration program. In OER the RCL section preceded the FCL section. Because of the apparent influence of the FCL procedure in the RCL procedure, this arrangement should be reversed.

The recommended procedure has one other considerable advantage. It will almost certainly reduce the weighting of the over-all rating in the total score, thus making it still more difficult for the rater to determine just where he is placing the individual rates on the final scale.

## VII. TECHNICIAN RESPONSIBLE FOR STUDY

E. A. Rundquist.