

A CRITICAL STUDY OF ASSESSMENTS MADE DURING
A TRIAL INTERVIEW SITUATION.

Thesis submitted in fulfilment of the
requirements for the degree of Master of
Science (in Psychology)

by

A.L.Pons.

ACKNOWLEDGEMENTS.

I should like to acknowledge the facilities made available to me by Dr.S.Biesheuvel, Director of the National Institute for Personnel Research.

I am also indebted to Mr.A.G.Arbus who, in his capacity as Research Officer, planned and organised the selection procedures of which the Trial Interview was an integral unit. I have greatly appreciated his helpful criticisms and guidance.

I acknowledge with gratitude the careful consideration which my work has received from Professor E.H.Wild. His advice and constructive criticisms have contributed much to the study.

I also wish to acknowledge the invaluable criticisms and suggestions contributed by Mr.J.S.Maritz of the Statistical Section of the National Institute for Personnel Research.

I am indebted to Miss. S.Glen for the drawing of the numerous frequency distributions and to Mr.L. van Zyl for the typing.

Finally I wish to thank Mrs.J.Esprey, also of the National Institute for Personnel Research, for her painstaking efforts to ensure accuracy in the many calculations.

THE TRIAL INTERVIEW.

Administrative Trainee Selection (1948-1949).

CONTENTS.

CHAPTER I. INTRODUCTION.

- A. The purpose of the selection procedures.
- B. The psychometric method of assessing test performance in the present battery.

CHAPTER II. THE TRIAL INTERVIEW.

- A. The object of the Trial Interview.
- B. Method of administration.
- C. Method of assessing performance.
 1. General.
 2. The interviewee's report.
 3. The rating scales.

CHAPTER III. RESULTS.

- A. The Overall Assessment.
 1. The characteristics of the frequency distributions of the tests included in the battery.
 2. Varying standards used by the raters.
 3. Non-homogeneity of the group.
 4. Administration of the testing procedures.
 5. Subjective influences.
 6. Considerations in terms of 1. to 5.
- B. The Assessment of Individual "Personality" and "Ability" items.
 1. The characteristics of the frequency distributions of "Personality" Total, "Ability" Total and "Personality+Ability" Total.
 2. A statistical explanation of multimodality.
 3. The reasons for the high correlations between individual items in the "Personality" and "Ability" fields.
 4. Selection and definition of rated items.
- C. A Comparative Study of the Overall Assessments and the Assessments of "Personality" and "Ability" Items.
- D. The Inconsistency of Assessments.
- E. A Study of the Interviewees' Reports.
 1. Inadequate interviews.
 2. Method of separation.
 3. The assessments of candidates whose interviews are inadequate.
 - a.) Assessments of "Personality" and "Ability" items.
 - b.) Overall Assessments.
 4. The frequency distributions of Adequate and Inadequate interviews.
 - a.) Adequate interviews.
 - b.) Inadequate interviews.

5. The correlation of Trial Interview Assessments with the remainder of the battery's tests.
 - a.) The Overall Assessments.
 - b.) The assessments of individual "Ability" and "Personality" items.

CHAPTER IV. A STUDY OF THE WEIGHTS OF THE BATTERY'S TESTS.

- A. Before Purification.
- B. After Purification.
- C. The Weights of Various Sets of Assessments Made During the Trial Interview.
- D. The Significance of the "Ability" Total's Weight.

CHAPTER V. A SUMMARISED REVIEW OF THE STUDY OF THE TRIAL INTERVIEW'S ASSESSMENTS.

CHAPTER VI. SUGGESTED INNOVATIONS IN THE TEST AND IN THE METHOD OF ADMINISTRATION.

ANNEXURE A. Product Moment Correlation Coefficient.

ANNEXURE B. Contingency Coefficient.

ANNEXURE C. The Significance of Difference Test.

ANNEXURE D. Pivotal Condensation.

ANNEXURE E. The New Test.

The books and publications referred to in the text are listed in the Bibliography, page 101.

C H A P T E R I

INTRODUCTION.

A. The purpose of the selection procedures.

At the end of 1948 the National Institute for Personnel Research (N.I.P.R.) organised a series of testing procedures for the selection of administrative trainees for a large industrial organisation. The fruitful results obtained by the British War Office and the Civil Service Selection Boards in the application of approved techniques for similar purposes supplied the basis on which the battery of tests was constructed. While having regard for this established approach as well as the controlling influence imposed by the stipulations of the contract, the N.I.P.R. was able to devise a battery which provided scope for its specific interest in research.

The industrial organisation's demands were, briefly, that the N.I.P.R. should, after the completion of the testing programme, submit recommendations for the acceptance of twenty-five to thirty applicants on the proposed training scheme, the ultimate purpose of which was to create a trained and experienced reserve of employees capable of executing high-level administrative duties; particular emphasis in the selection of the trainees was consequently placed on a superior standard of achievement throughout the series of tests.

The battery of tests endeavoured to ensure complete coverage of all aspects considered relevant. Broadly these were - (i) innate mental abilities.

(ii) family, educational and occupational (if any) background.

(iii) interpersonal relations in group situations.

(iv) interpersonal relations of a face-to-face nature.

- (v) personality make-up.
- (vi) the manner and effectiveness of dealing with administrative problems in actual situations and, subsequently, in the form of considered reports.
- (vii) clinical interview where the amassed information was reviewed in gestalt form to confirm both the favourable considerations and possible flaws or weaknesses. A short statement of the general principles considered while selecting and designing the tests to be used in the battery follows in Section B. (*)

B. The psychometric method of assessing test performance in the present Battery.

The psychometric method of assessment is applicable to the whole of any battery in which each test and test situation is assessed by means of a definite score or system of points. The subsequent weighting of each test score included in the battery allows the final recommendations to be placed in an order of merit more discriminating than (Ref. 1 pages 149 - 151) would be possible as a result of purely subjective methods of assessing performance and potential.

(*) A detailed report of the principles involved, methods used and the results of the project as a whole is given by A.G. Arbous (Ref. 1). The present paper is concerned with one aspect of these selection procedures. The writer was not involved in the planning and preparation referred to in Chapters 1 and 2.

The work recorded in Chapter 3 Section A was undertaken by the Industrial Research Team of which the writer is a member. The writer's individual research begins in Chapter 3 Section B.

Procedures incorporating tests to investigate each of the aspects listed in A must naturally be comprehensive. Moreover, each test situation (over and above the actual scores obtained from the pencil and paper tests) was designed so as to provide psychometric assessments rather than purely descriptive comments. The testing procedures accordingly demanded that the candidates complete each of the following:-

- (i) Test A(F), a measure of general mental ability.
- (ii) Test M, a measure of abstract reasoning ability.
- (iii) Test H, a measure of number ability.
- (iv) Test G, a measure of ability to perceive essentials in a given mass of detail.
- (v) A written project; ability to deal with administrative problems on paper.
- (vi) A Trial Interview; an inter-personal situation of a face-to-face nature.
- (vii) Two group discussions (Assigned Leadership and Leaderless Group discussions); assessment of personality attributes, potential and performance in a group.
- (viii) A biographical inventory.

At the conclusion of the testing procedures a report was written on each candidate at which stage he could be appraised in relation to his total performance. The following main headings were dealt with in the report:-

1. Full names, age, etc.,
2. Family record (i.e. orphaned, divorced, health, etc.).
3. Educational record.
4. Occupational record.
5. Motivation as evident in 3 and 4.
6. Test results.
7. Temperament and personality qualities.
8. Recommendation (accept, borderline or reject).

Apart from the factual information listed, general comments followed where specifically relevant.

Tests (i) to (v) above conform readily with the requirements of the psychometric methods of assessment. Though these assessments do not throw any light on section 7 of the report, tests (vi) and (vii) were designed for this purpose. The dynamic organisation of personality, however, renders more difficult a psychometric assessment of its attributes; for the same reason it is also difficult to evaluate the performance in situations where personality is an active force. The disadvantage inherent in the practice of assigning values to specific and apparently individual aspects of personality (implicitly so by virtue of their assessment) was considered to be outweighed by the following benefits:-

- (i) the inclusion within the battery of a measure of a candidate's personality make-up remedies the deficiency presented by an array of purely objective pencil and paper test scores. The validated inclusion of such a measure ensures that due cognisance is taken of the significance of personality qualities in conjunction with other more readily recognised abilities. (Ref. 2 pages 9 - 11). This is of special importance when considering the purpose of the selection procedures.
- (ii) The Assessments and accompanying comments made available by two of these situations (test (vii)) are the test results from which the Final Ratings are obtained; these are measures of overall potential in the administrative field in which special note is taken of personality attributes. (In due course this Final Rating was found to correlate highly (.600) with the criterion of success and to contribute considerable weight (.322) to the

battery's predicting efficiency).

The results of the Trial Interview (test (vi)) were not included either in the Final Rating or as individual test results in the battery because of the view that the assessors' consistency had been impaired by subjective influences (page 28). The present paper is concerned with the analysis which led to this view and the subsequent study of the assessments and its findings.

C H A P T E R II

THE TRIAL INTERVIEW

A. The object of the Trial Interview.

The battery of tests includes two situations in test (vii) (Leaderless group discussion and Assigned Leadership group discussions) which enable the team of assessors to observe for almost three hours each candidate's performance within a group of six or eight fellow applicants. The assessors' attention is focussed mainly on the personality attributes for the purpose, not only of getting to know more about each individual in general terms, but also of deriving an otherwise unobtainable value which can contribute a weight to the battery's predicting score.

It was anticipated that the Trial Interview (test (vi)) would strengthen the conclusions or reveal possible weaknesses in the assessments made in the above two group tests (test (vii)). In so doing the benefits mentioned in Chapter I, section B would be within more certain reach. Moreover, it was considered that observations of the candidate's behaviour in an interpersonal situation of a face-to-face nature while in a position of authority would be significantly indicative of possible success or failure as an administrator. Also, the need was felt to determine more satisfactorily a candidate's ability to make shrewd assessments of human potentialities when faced with the task of employee placement. In view of the practical experience gained in England (Chapter I, section A) as well as on theoretical grounds, this ability was considered to be one of the many which administrators should possess.

The purpose of the Trial Interview can accordingly be summed up as follows. Usually the gestalt impression of a candidate is gained from a review of the quantitative assessments obtained from the pencil and paper tests and an

appraisal of his personality make-up during the clinical interview. The Trial Interview (as well as the two group discussions) was included in the battery for the purpose of improving this situation by obtaining a reliable quantitative measure of personality attributes. The final assessment made on each candidate would therefore be derived not only from quantitative ratings representative of both pencil and paper scores (as was the case originally) but from the personality assessments as well.

This would enable the N.I.P.R. to supply Management with a listed table of recommendations in order of merit thus replacing the original three broad and insensitive categories of recommendation, viz., accept, borderline and reject (Chapter 1, section B 8). Once this is achieved and due regard is given to both the criterion of success and Management policy, the level of acceptance in terms of the final battery score can be determined satisfactorily.

B. Method of administration of Trial Interview.

The candidates are prepared by being told that they will have to interview a female applicant for the post of secretary/assistant to a senior executive (full details of the Trial Interview test are recorded in reference 1 appendix pages 60- 74). Each candidate is given sufficient time (about thirty minutes) to examine closely the informative data supplied:-

- (i) specifications and requirements of the new post for which application is being made.
- (ii) testimonials from the applicant's past employers.
- (iii) testimonials from the Dominee of the applicant's home town.

He is allowed to prepare notes on method of approach and treatment of the pending interview if he considers it necessary. The material given to each candidate, though

comprehensive, is presented in such a way that the applicant's personality and occupational inadequacies are not readily apparent; moreover the situation as devised includes several blind alleys into which the less vigilant interviewer will be led. The candidate, assuming the role of senior official, is accommodated in an office where he interviews the applicant for the purpose of establishing the extent of her suitability for the post.

It is important to stress the fact that the candidate is left in no doubt as to what is required of him, viz., he has to interview the applicant (being prepared in terms of the information supplied) and determine, in the light of the corroborative (or otherwise) data elicited during the interview, whether she is to be offered the post. The interviewer may conduct the proceedings in either English or Afrikaans.

The applicant whose identity is cloaked under the nom-de-plume of Miss P.C. van Jaarsveld, is in actual fact a trained and experienced member of the N.I.P.R. staff. Her duties as sole observer necessitate a complete report and assessment without having the opportunity of consulting other raters in order to clear up points about which she may be doubtful and uncertain. Moreover, the very nature of the test situation makes it impossible for her to control in any direct way the manner in which the interview is conducted, its thoroughness and effectiveness. Her instructions, as regards imparting the correct though less satisfactory information about her work record, motivation and sense of responsibility both at school and in later life are as follows. Miss van Jaarsveld's first answer to any question must conform to the information supplied in the letter of application. For instance, to the interviewer's

possible question - "Why did you fail Junior Certificate in 1938?" - her initial reply must be - "Besides not caring for Arithmetic, I had an unpleasant Maths teacher, who invariably made life unpleasant for me. I was unable to do myself justice and failed Arithmetic." However, if he probes further and asks confirmation of this in another way using a different approach she must reply with the correct answer - "Actually, I prevailed on my father to take me away from boarding school and I lived with my aunt. As a result I had too many parties and my schoolwork suffered."

The situation which has been developed in this way is revealing. It is designed to enable a reliable estimate (confirmation or otherwise of the assessments made during the two group discussions) to be obtained of an applicant's ability to evaluate human material; likelihood of accepting facts at their face value; method, manner and deportment when in a position of authority, as well as several additional "ability" and "personality" aspects listed in section C below.

At the conclusion of the interview the candidate is instructed to draft a letter to his immediate superior recommending the acceptance (or otherwise) of Miss van Jaarsveld for the post. This letter was rated by three independent assessors. The agreement between these assessors was so poor that the ratings were not used. (Ref. 1 pages 51 - 52).

C. Method of assessing Performance.

1. General.

The following four aspects of the Trial Interview situation allow the assessments and comments on the candidate's personality and performance in the test to cover a wide field.

- (i) Interviewer's letter to his superior.

- (ii) Interviewee's report.
- (iii) Quantitatively, by means of rating scales used by the interviewee.
- (iv) Subjectively in relation to all other data gathered during the testing procedure.

Though the value of (i) and (iv) above is considerable in relation to the total overall picture, the present study is especially concerned with (ii) and (iii) (Chapter II, section A).

2. The interviewee's report is designed to present noteworthy comments on three aspects emerging from the Trial Interview situation:-

- a. Temperament and Personality.
- b. The method, thoroughness and effectiveness of the interview.
- c. The interviewer's letter to his superior. The report and the outcome of the interview may be viewed in its entirety more satisfactorily when the interviewee has ascertained whether the contents of the drafted letter are representative of the interview proper or not. In some cases an adequate and searching interview was followed by a mediocre and incomplete statement of facts in the letter. Some cases were found where hopelessly inadequate interviews were followed by recitals based purely on the contents of the material originally supplied, or by statements produced by fertile and ingenious imaginations.

The comments embodied in the interviewee's report are of a supplementary nature to the rating scales described below. A profile apparent when viewing the distribution of ratings assumes significance when accompanied by explanatory notes. This arrangement of perspective is of particular

use in anomalous instances where a very low assessment may be indicative of the inhibiting functions of extreme nervousness on potential rather than of low ability as such.

3. The rating scales provide the means whereby the purpose of the test may be achieved. The other related aspects serve the functions of supplementation and corroboration; with them as reference a meaningful interpretation of the rating scales is possible where necessary. Moreover, in the event of significantly varying standards being used by the raters, notes for the basis of a discussion to remove the discrepancy are available.

Two rating scales are used. In the first one five individual "Personality" items and five individual "Ability" items are rated on a six-point scale. The second one consists of an overall assessment on a nine-point scale.

The grouping of items under headings of "Personality" and "Ability" must to some extent be arbitrary and therefore not absolutely satisfactory. While recognising this limitation, justification for the broad classification suggested to the assessors by Arbous (Ref. 1) is found in the views of various leading authorities. An examination of the points outlined by Tredgold (Ref. 3 pages 66 - 78) in his study of the normal mind reveals that the listed "Personality" items can be readily associated with the factors or variables characteristic of the conative and affective fields. Similarly, though not quite so convincingly, the "Ability" items are identified with the variables of the cognitive field. Moreover Tredgold's views are not unsupported and isolated. The Heymans and Wiersma theory developed by Biesheuvel (Ref. 4 pages

1 - 15) provides similar corroboration with the additional recommendation of having been satisfactorily tested empirically.

Therefore, though the original grouping of "Personality" and "Ability" items was made for the convenience of the assessors, the writer believes it was largely justified. But at no stage is it suggested or implied that the ratings of one or more of these items is a measure of Personality as such or Ability as such. Moreover, the use of the terms "Personality" items and "Ability" items (later "Personality" Total and "Ability" Total as well) with inverted commas throughout the text will serve as a constant reminder of this.

(a) The assessment of individual items covers two fields as follows:-

- (1) Personality make-up as revealed in an interpersonal situation of an intimate nature with emphasis on:-
 - (i) self confidence.
 - (ii) presence (i.e. poise, bearing, social effectiveness)
 - (iii) pleasantness of manner.
 - (iv) ability to impose his personality on interviewee.
 - (v) ease of establishing social contact.
- (2) Ability, not only in conducting the interview but also in evaluating the personality qualities of the interviewee and differentiating between essentials and non-essentials in the data supplied. The facets of this broadly defined field on which special attention is focussed are:-
 - (vi) perseverance in getting at facts.

- (vii) effective use of questions.
- (viii) effective use of data supplied.
- (ix) effectiveness of control of situation.
- (x) appreciation of requirements of situation.

The rating scale used together with a verbal description of the relative values is as follows:-

Very Inferior	Inferior	Low Average	High Average	Superior	Very Superior
0	1	2	3	4	5

(b) One overall assessment by the interviewee is made on the following scale for each interviewer:-

Exceptional	Superior	Very Good	High Average	Average	Low Average	Below Average	Poor	Very Poor
A+	A	A-	B+	B	B-	C+	C	C-

The assessor is instructed to place an "X" under the appropriate letter.

There is an important difference between the assessments of the individual items in (a) and the overall assessments in (b). The assessments made on the scale in (a) represent the raters' evaluations of the listed items in terms of the extent to which the attributes or qualities are manifested during the Trial Interview situation. The Overall Assessments, on the other hand, though based on the observed performance in the test, are made bearing in

mind the purpose for which the selection procedures are being conducted, i.e.,

"From what I have observed, which position on the scale represents this candidate's potential ceiling, with particular reference to personality attributes, in relation to the demands which he will have to meet as an administrator?"

The desirability of obtaining uniformity of standards was further considered by instructing the assessors that a rating of B+ was to signify that, in terms of the applicant's performance in the situation, he just qualified for acceptance on the training scheme. Similarly a rating of B indicated that he just failed to qualify for acceptance.

The interpretation of administrative duties which was used throughout the testing procedures can best be clarified by quoting the definition made known to and used by all the assessors (Ref. 1 Annexure page 57):-

"Administrative jobs refer to positions which are concerned with the formulation of policies or drawing up of plans. The incumbents are expected to possess knowledge, experience and wisdom, and to be able to make shrewd assessments of the values and potentialities of men, machines and systems. The responsibilities of the positions are heavy for in them rests in the first instance the success of the undertaking as a whole, by planning the effective co-ordination of constituent elements, and being able to think into the future."

Assessments with the use of rating scales are subject to several well-known sources of error. The interviewees' pre-test briefing included a review of these.

C H A P T E R III

RESULTS.

Altogether 219 interviews were conducted. These were allocated to the interviewees as follows:-

Mrs. B	104
Mrs. P	69
Rest (4 Raters)	<u>46</u>
	<u>219</u>

A. The Overall Assessment.

1. The characteristics of the frequency distributions of the tests included in the battery.

The scores and assessments obtained by the candidates in the battery's tests and situations were distributed with the following results.

- (a) Test A(F) (Fig. 1)
- (b) Test M (Fig. 2)
- (c) Written Project (Fig. 3)
- (d) Overall Assessment of the interviewer by the interviewee (Fig. 4)
- (e) Final Rating on two group tests (Fig. 5).

The tests listed on page six but not included in the battery were omitted on account of being considered unsuitable for the reasons given by Arbous (Ref. 1 pp 52, 65 - 69).

The frequency distribution of the overall assessment (O.A.) is bimodal. The pronounced modes are separated by two class intervals. The remainder of the frequency distributions are definitely unimodal though skewness is apparent particularly in the Written Project (W.P.) and the Final Rating (F.R.). Furthermore the ten frequency distributions of the assessments of the individual items are also unimodal (Sets of Figures 6 and 7). Owing to the

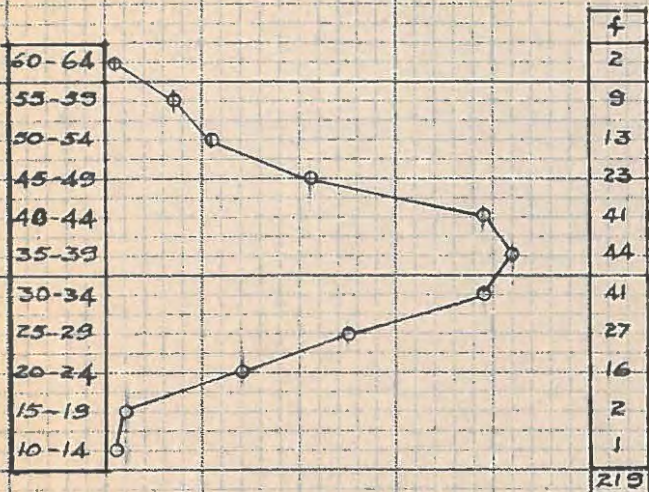


FIG. 1: TEST A(F)

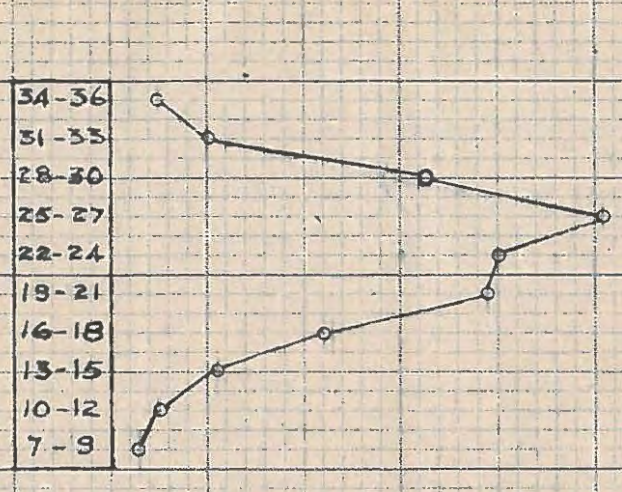


FIG. 2: TEST M

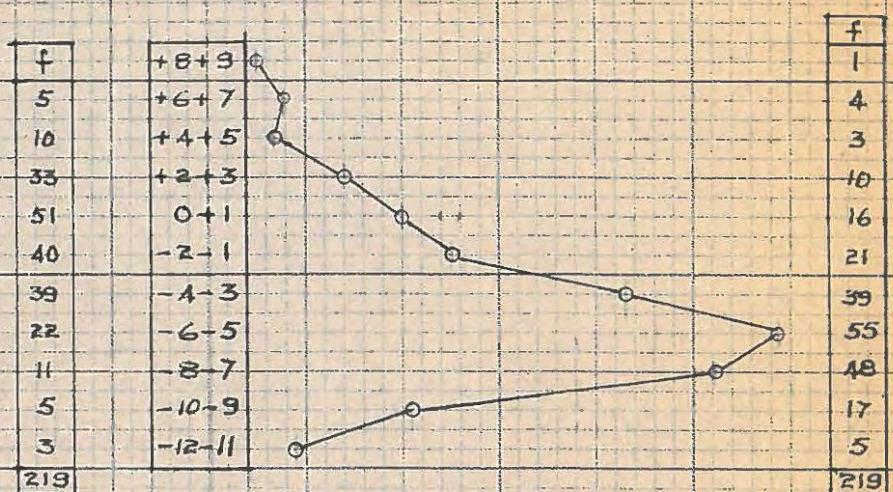


FIG. 3: WRITTEN PROJECT

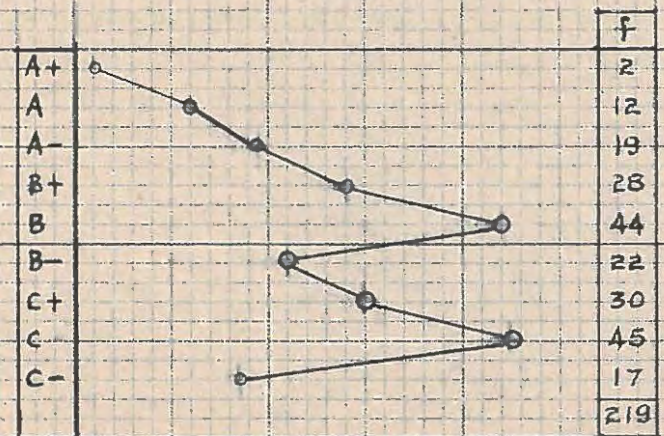


FIG. 4: OVERALL ASSESSMENT

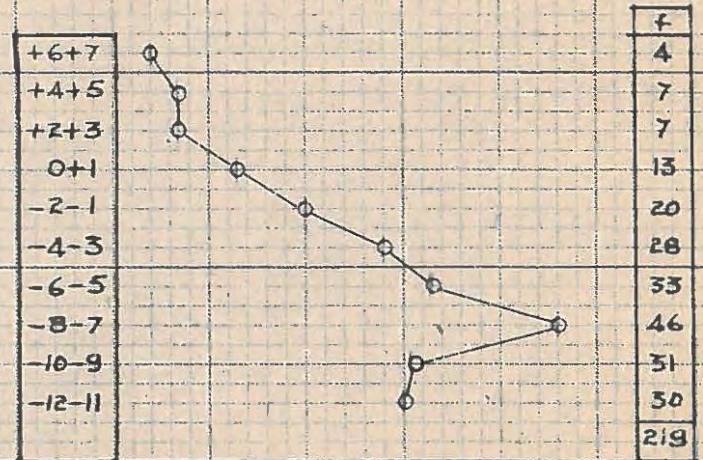
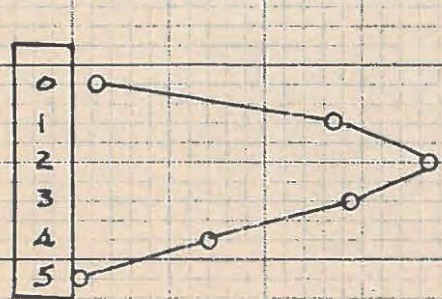


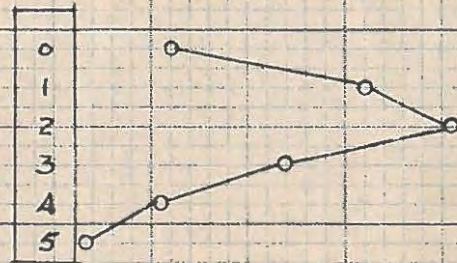
FIG. 5: FINAL RATING

-19-



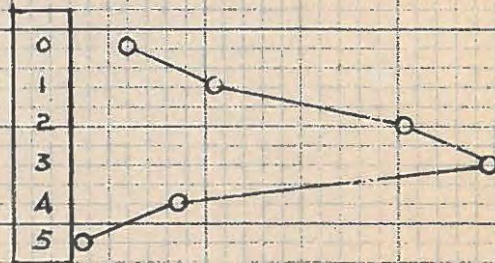
(i) SELF-CONFIDENCE

f
5
54
74
57
28
1
219



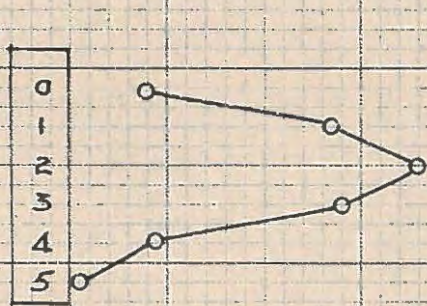
(ii) PRESENCE

f
20
60
77
43
17
2
219



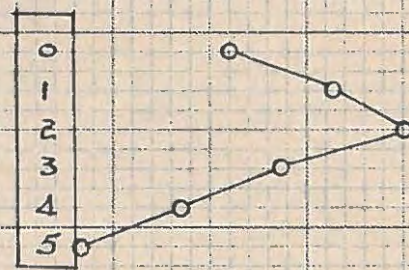
(iii) PLEASANTNESS OF MANNER

f
11
29
69
86
22
2
219



(iv) EASE OF ESTABLISHING SOCIAL CONTACT

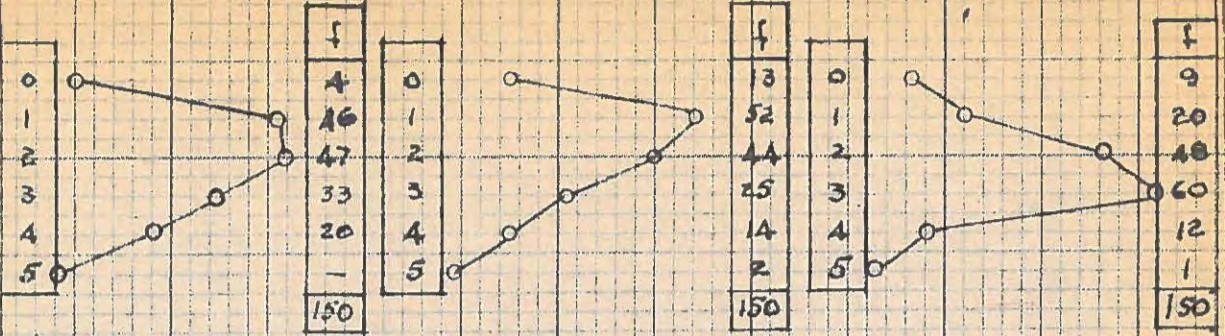
f
16
55
73
56
17
2
219



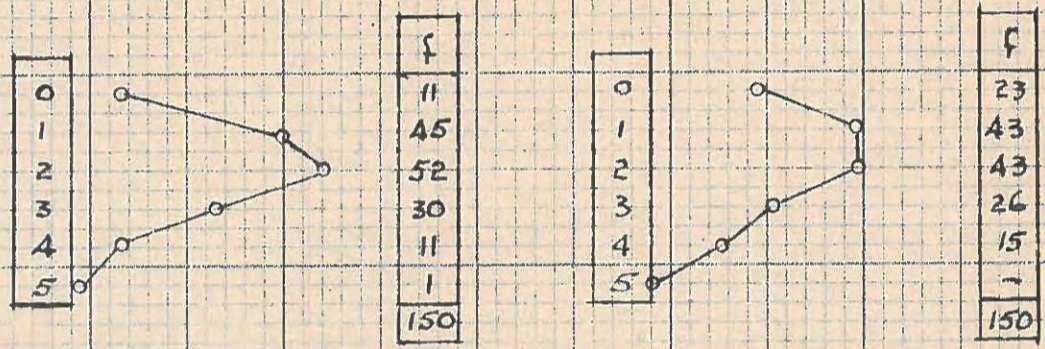
(v) IMPOSING PERSONALITY

f
32
53
68
43
22
1
219

-20-

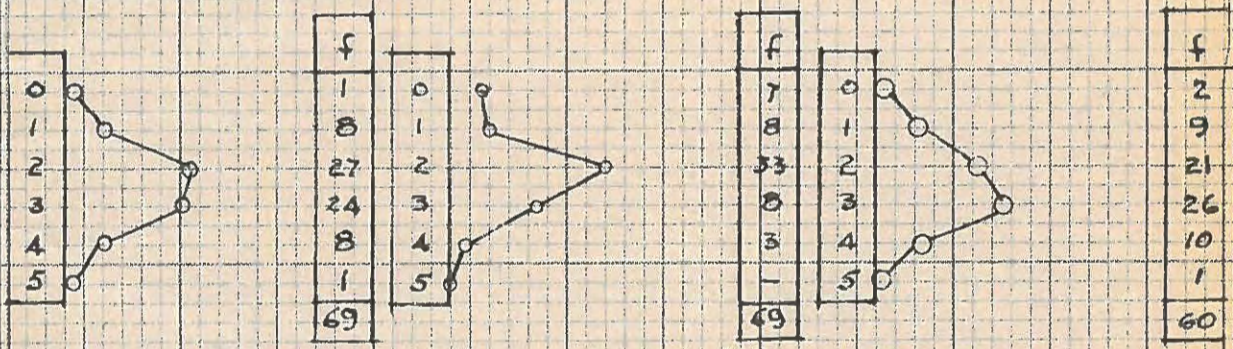


(i) SELF-CONFIDENCE (ii) PRESENCE (iii) PLEASANTNESS OF MANNER

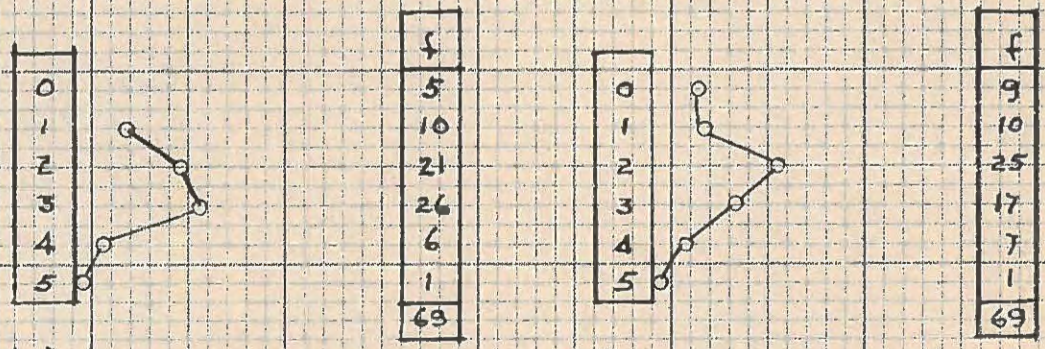


(iv) EASE OF ESTABLISHING SOCIAL CONTACT (v) ABILITY TO IMPOSE PERSONALITY

FIGS. 6B RATINGS BY MRS "B" PLUS REST OF KATERS

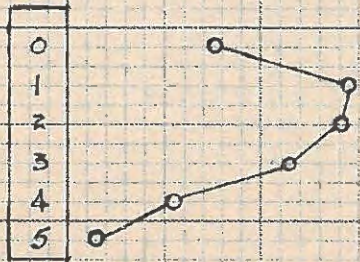


(i) SELF-CONFIDENCE (ii) PRESENCE (iii) PLEASANTNESS OF MANNER

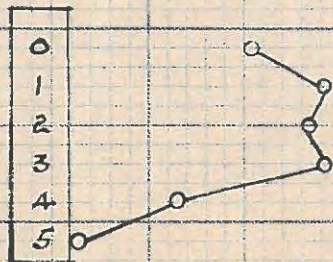


(iv) EASE OF ESTABLISHING SOCIAL CONTACT (v) ABILITY TO IMPOSE PERSONALITY

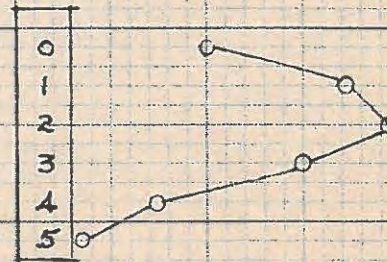
FIG 6c RATINGS BY MRS P



f
31
58
57
45
23
5
219



f
38
53
50
53
22
3
219

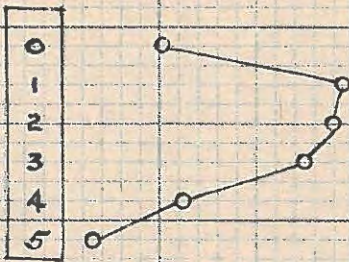


f
29
57
65
48
18
2
219

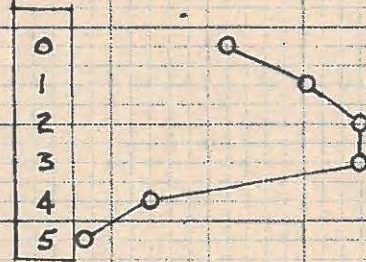
(vi) GETTING AT FACTS

(vii) USE OF QUESTIONS

(viii) USE OF DATA SUPPLIED



f
23
57
56
50
27
6
219



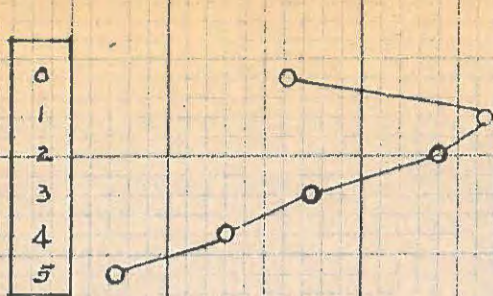
f
32
49
60
60
16
2
219

(ix) CONTROL OF SITUATION

(x) APPRECIATION OF REQUIREMENTS OF SITUATION

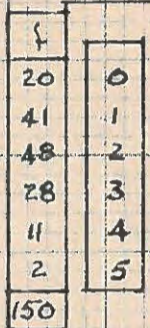
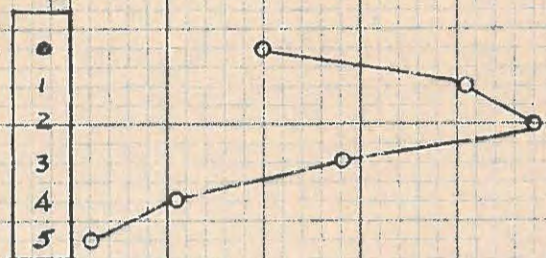
-22-

FIGS. 7A "ABILITY" ITEMS - RATINGS BY ALL ASSESSORS -



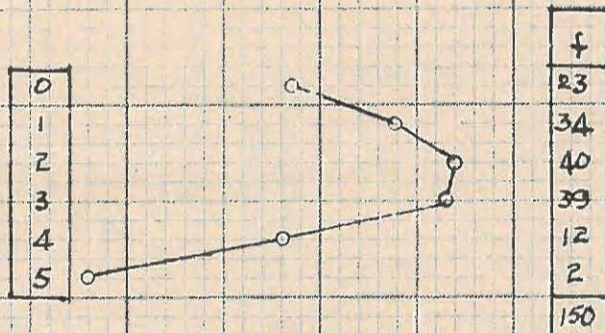
(vi) GETTING AT FACTS

(vii) USE OF QUESTIONS



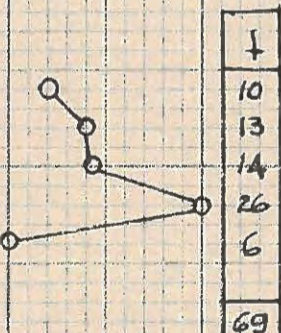
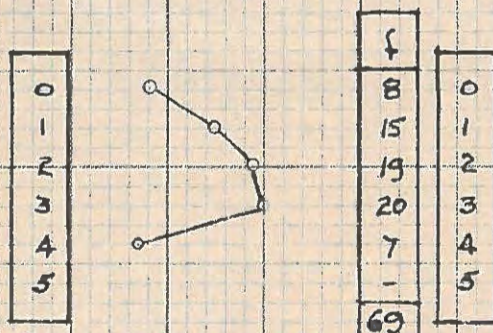
(viii) USE OF DATA SUPPLIED

(ix) CONTROL OF SITUATION



(x) APPRECIATION OF REQ. OF SITUATION

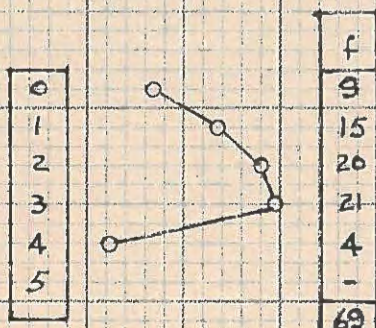
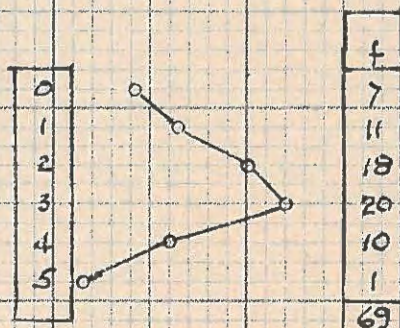
FIGS. 7b RATINGS BY MRS B & REST OF WATERS



(vi) GETTING AT FACTS

(vii) USE OF QUESTIONS

(viii) USE OF DATA SUPPLIED



(ix) CONTROL OF SITUATION

(x) APPRECIATION OF REQ. OF SITUATION

FIGS. 7c RATINGS BY MRS P

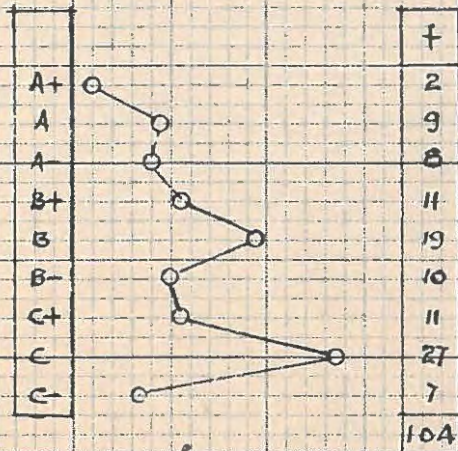


FIG. 8: MES B's RATINGS

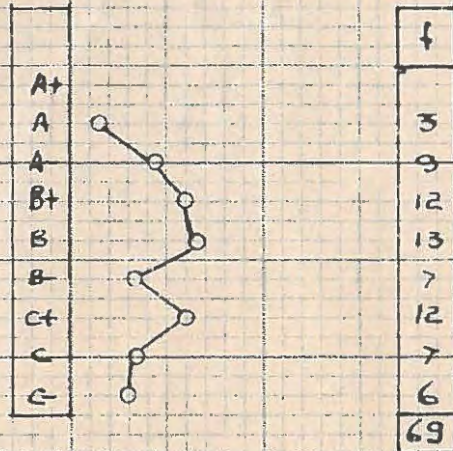


FIG. 9: MES P's RATINGS

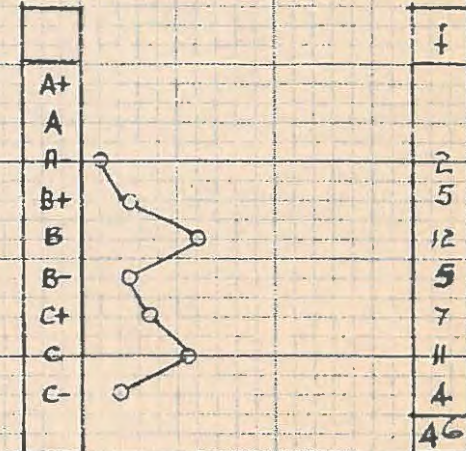


FIG. 10: RATINGS BY REST OF ASSESSORS

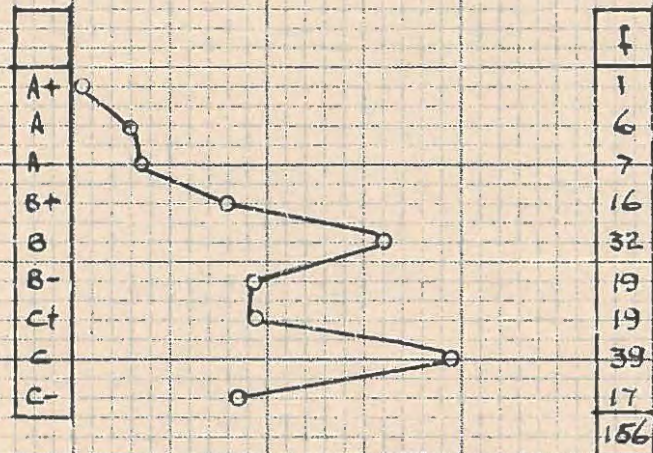


FIG. 11: INTERNAL CANDIDATES

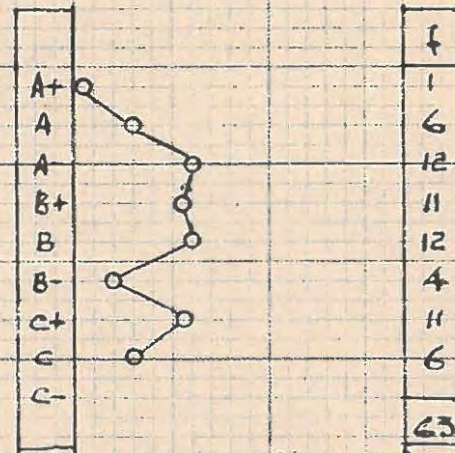


FIG. 12: EXTERNAL CANDIDATES

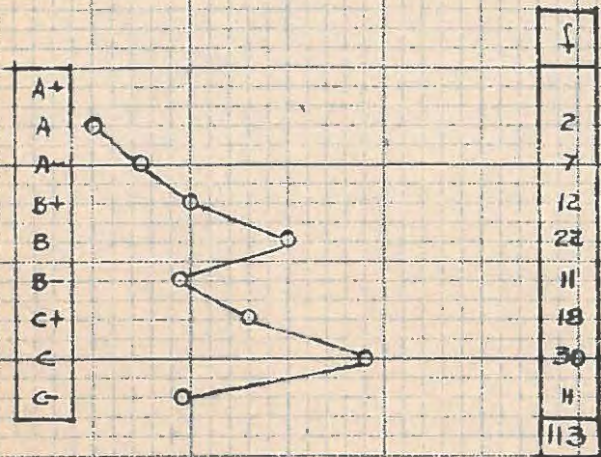


FIG. 13: AFRIKAANS SPEAKING CANDIDATES

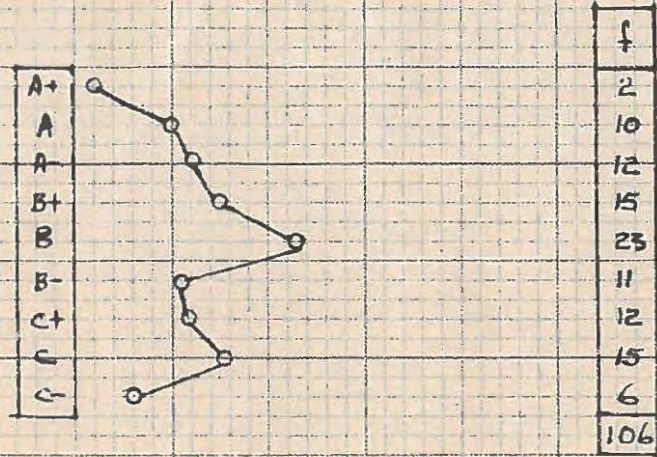


FIG. 14: ENGLISH SPEAKING CANDIDATES

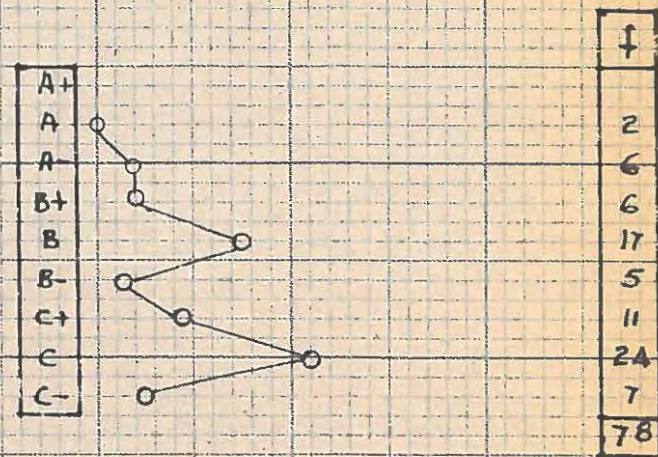


FIG. 15: AFRIKAANS SPEAKING CANDIDATES RATED BY MRS B + REST OF VOTERS

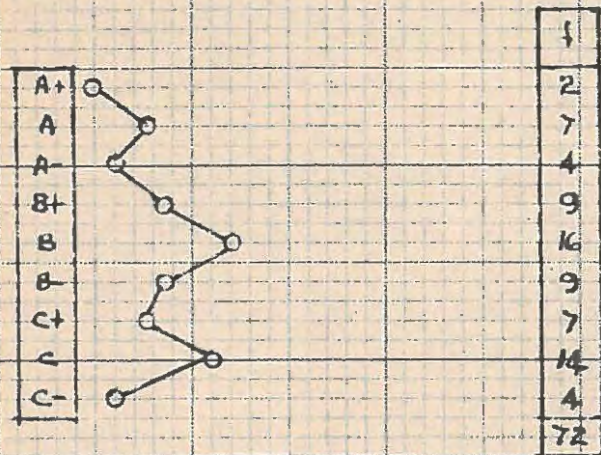


FIG. 16: ENGLISH SPEAKING CANDIDATES RATED BY MRS B + REST OF VOTERS

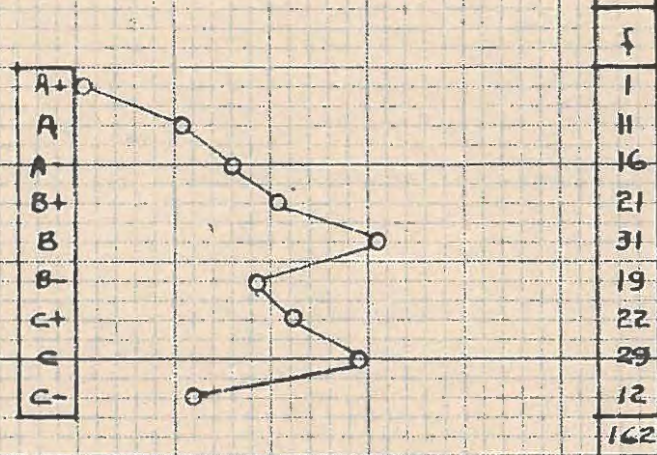


FIG. 17: 1948 CANDIDATES

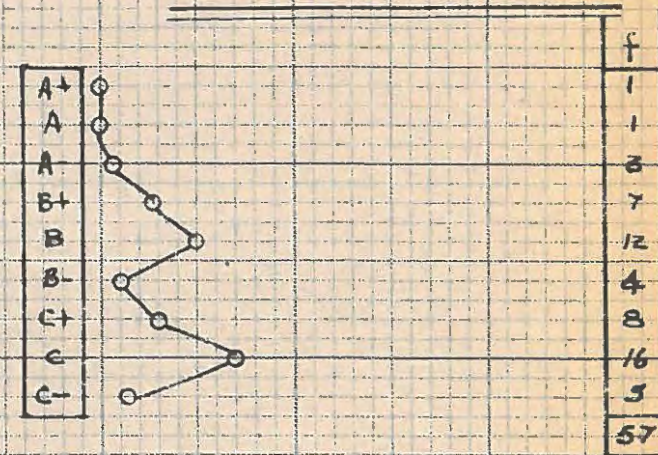


FIG. 18: 1949 CANDIDATES

-240-

fact that these fifteen distributions refer to the same 219 cases, distinct bimodality in only one of them is particularly striking.

However, there is no known method of testing for the significance of the bimodality of a distribution; besides, little is known of the effect of bimodality on reliability coefficients and prediction. Clearly a fuller appreciation of these problems will require advanced mathematical investigations. Though the need for a strict test is therefore realised, it was impossible to apply one in this case. Other means were accordingly used in an effort to determine the factor(s) causing the emergence of this bimodality.

2. Varying standards used by the Raters.

The first aspect investigated was the possibility that differing standards had been used by the raters resulting in bimodality when all the assessments were pooled to form one frequency distribution.

- (a) All Raters (Fig. 4)
- (b) Mrs. B (Fig. 8)
- (c) Mrs. P (Fig. 9)
- (d) Rest of Raters (Fig. 10).

The marked bimodality of each of these distributions dispels this first suggestion.

3. Non-homogeneity of the group.

Several explanations were investigated in view of the possible non-homogeneity of the group of applicants.

An almost equal number of candidates was drawn from the two official language groups of the Union, the numbers being 113 Afrikaans-speaking and 106 English-speaking. The candidates were further representative of two groups of employee material:-

- (i) employees of the sponsor organisation (internal

candidates) in whose case the qualification of a University degree or comparable diploma was not enforced.

(ii) external candidates, most of whom had University education while some had professional experience as well.

The regrouping of candidates was carried out as follows:-

- (a) Internal and external applicants (Figs. 11 and 12)
- (b) Afrikaans speaking and English speaking groups (Figs. 13 and 14).

The groups were broken down still further in terms of the Raters' assessments.

- (c) Mrs. B + Rest - assessments of Afrikaans-speaking (Fig. 15).
- (d) Mrs. B + Rest - assessments of English-speaking (Fig. 16).

Reference to these frequency distributions shows clearly that they all reveal the bimodal characteristics so definitely absent in the distributions of the battery's remaining test results. At this early stage of the investigation it therefore appears more probable that the reason for the distribution's bimodality will be found in an inherent psychological or structural feature of the test and a study of the following aspects becomes appropriate.

4. Administration of the testing procedures.

On account of the pending departure of the Director of the N.I.P.R. and the heavy programme necessary for testing over 200 candidates, it was decided to pass the apparently more promising applicants through the procedures first.

The letters of application formed the basis of this rough selection. It was therefore considered possible that the candidates tested in December, 1948 were superior to those tested in January, 1949 and the following distributions were made:-

- (a) All raters - 1948 (Fig. 17).
- (b) All raters - 1949 (Fig. 18).

The bimodality of these frequency distributions is as pronounced as before. An examination of the method of scoring test performances will follow in Section C and Section D at which stage the completed study of the rating of individual "Personality" and "Ability" items will allow the investigation to be more comprehensive and revealing.

5. Subjective influences such as, for example, sex bias have to be borne in mind. Could it be that a female assessor had rated a male interviewer to whom she felt consciously or sub-consciously attracted more leniently than his performance warranted? Clearly, if the answer is 'yes', the converse will also have operated thus introducing dissimilar standards of reference for the purposes of assessment. A confirmation of the operation of such a subjective influence would, it is suggested, provide an acceptable explanation of bimodality.

It is relevant at this juncture to refer to previous findings in a kindred field. (Ref. 5 pp 11 - 35). Men and women were asked to judge the character of persons of both sexes on the basis of an examination of their portraits. The summarised findings place the difference of judgement on various traits due to the following selective causes in the order

- 1st Liking/disliking-----acts most selectively.
- 2nd Age (over and below 30 years old).
- 3rd Sex.....acts least selectively.

"This is perhaps what might have been predicted, yet the comparatively small difference due to sex of the 'judges' is not one which would have been expected by everyone....."

While not losing sight of the evident importance of subjective influences a more thorough study of them is deferred to Section C below.

6. Considerations in terms of 1 to 5.

No amount of regrouping of the available results has removed or even reduced the bimodal characteristics of the overall assessment's frequency distribution. But it is evident that the explanation for it is not to be found in the non-homogeneous qualities of the group tested. The field of search is therefore being narrowed but, as yet, insufficient evidence has been collected to make it possible to pin-point the causal factor(s) with any degree of reliability.

The investigations so far carried out by the research workers concerned with the selection procedures nevertheless have led to the view that subjective inconsistency in some form (sex bias, liking or disliking, etc.) may have invalidated the assessments sufficiently to produce the two distinct modes in the frequency distribution. Consequently the Overall Assessment was considered to be suspect and was excluded from the battery. At this stage, however, a more detailed and searching analysis was planned for the purpose of obtaining a satisfactory explanation of bimodality in the present case. Such a finding would enable the introduction of changes and/or controls in the Trial Interview situation which would ^{reduce or} eliminate the ^{subjective inconsistency and the} bimodal characteristics of the distribution. It was anticipated that by doing so the test's contribution to the final battery score would become acceptable, a step not initially justified.

B. The assessment of individual "Personality" and "Ability" items.

1. The characteristics of the frequency distributions of "Personality" Total, "Ability" Total and

"Personality + Ability" Total.*

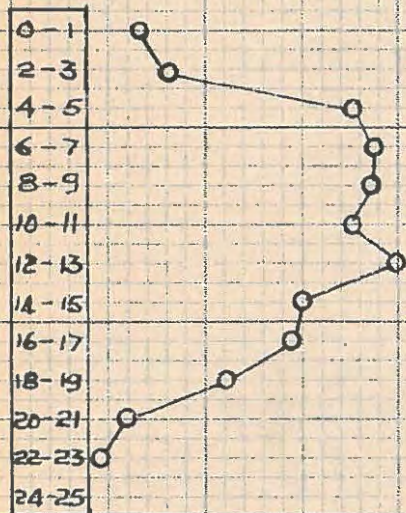
The following frequency distributions were examined:-

All Raters (N = 219)

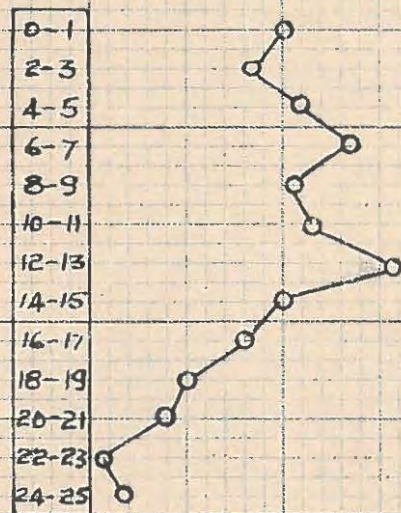
- (a) "Personality" Total (Fig. 19(i))
- (b) "Ability" Total (Fig. 19 (ii))
- (c) "Personality + Ability" Total (Fig. 19 (iii)).

The inspection leaves doubt as to whether the distributions are significantly bimodal or multimodal in the sense applied to the Overall Assessment's distribution. However, several modes are evident particularly in (b) and (c). At first this is particularly arresting because these distributions are each compounded of either 5 or 10 definitely unimodal distributions. The following tentative view therefore takes shape. When trained assessors genuinely striving at absolute objectivity direct their attention and observations at one segment of "Ability" or "Personality" it is possible to attain an apparent measure of consistency in so far as it is revealed by the expected unimodal distribution of scores. The totalling of all the applicants' scores on individual items results in a frequency distribution with more than one mode and, as explained below, may be indicative of a deep and not readily apparent in-

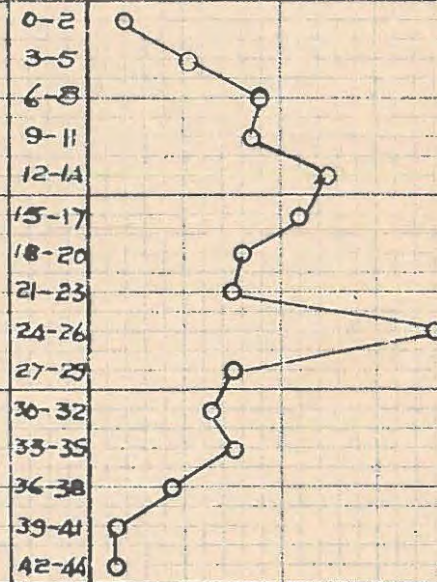
* On page 15 it was explained that each of 5 "Personality" and 5 "Ability" items was assessed on a 6 point scale. The actual raw scores obtained on the 5 "Personality" items were added to form the "Personality" Total; similarly for "Ability" Total. The "Personality + Ability" Total is obtained by adding the raw scores allotted to each of the 10 items. Fig. 19 (iii) shows, for example, that 4 candidates obtained a total score on all 10 items of 0 - 2; ten candidates obtained total scores ranging from 3 to 5.



f
5
8
27
29
29
27
32
22
21
14
4
1
-
219



f
5
8
27
29
29
27
32
22
21
14
4
1
-
219



f
4
10
18
17
24
22
16
15
36
15
13
15
8
3
3
219

1. TOTAL SCORE ON
"PERSONALITY" ITEMS

2. TOTAL SCORE ON
"ABILITY" ITEMS

3. TOTAL SCORE ON
"PERSONALITY + ABILITY" ITEMS

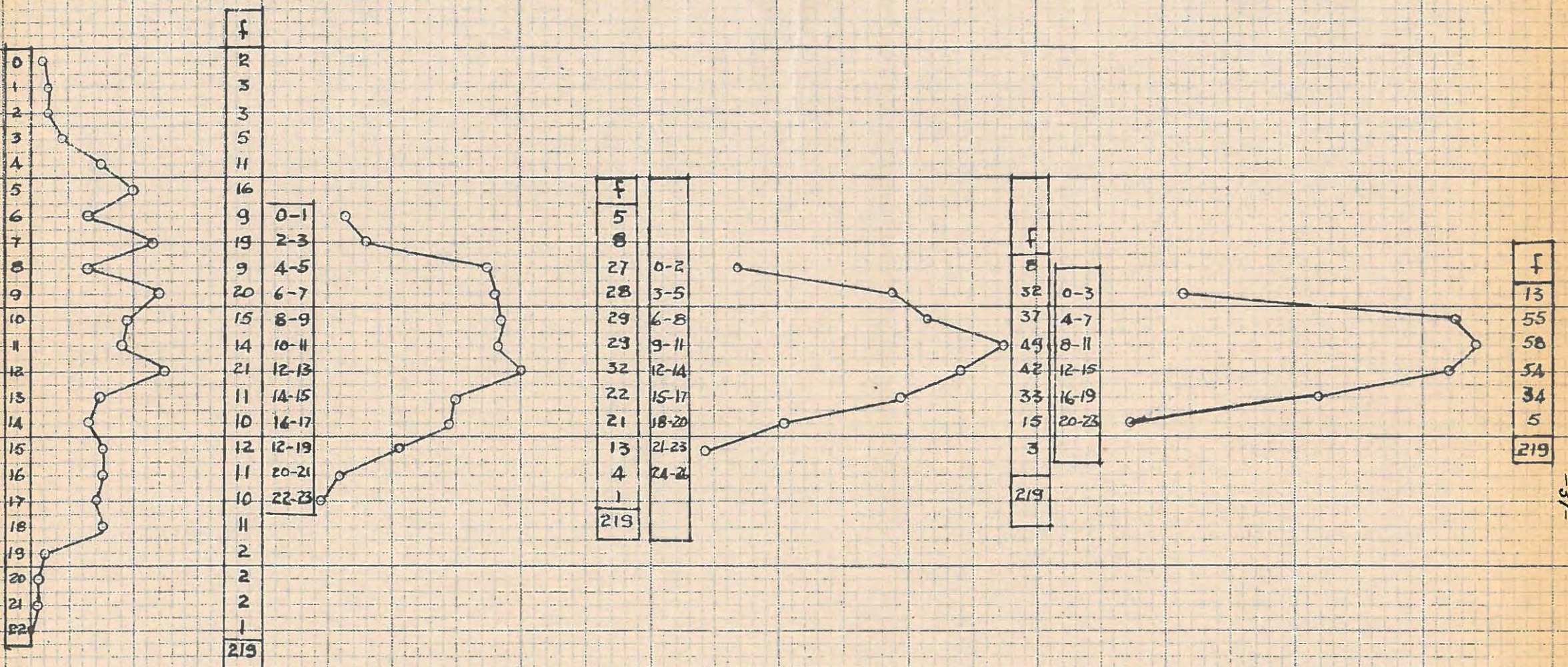


FIGURE 20 (c) "PERSONALITY" TOTAL

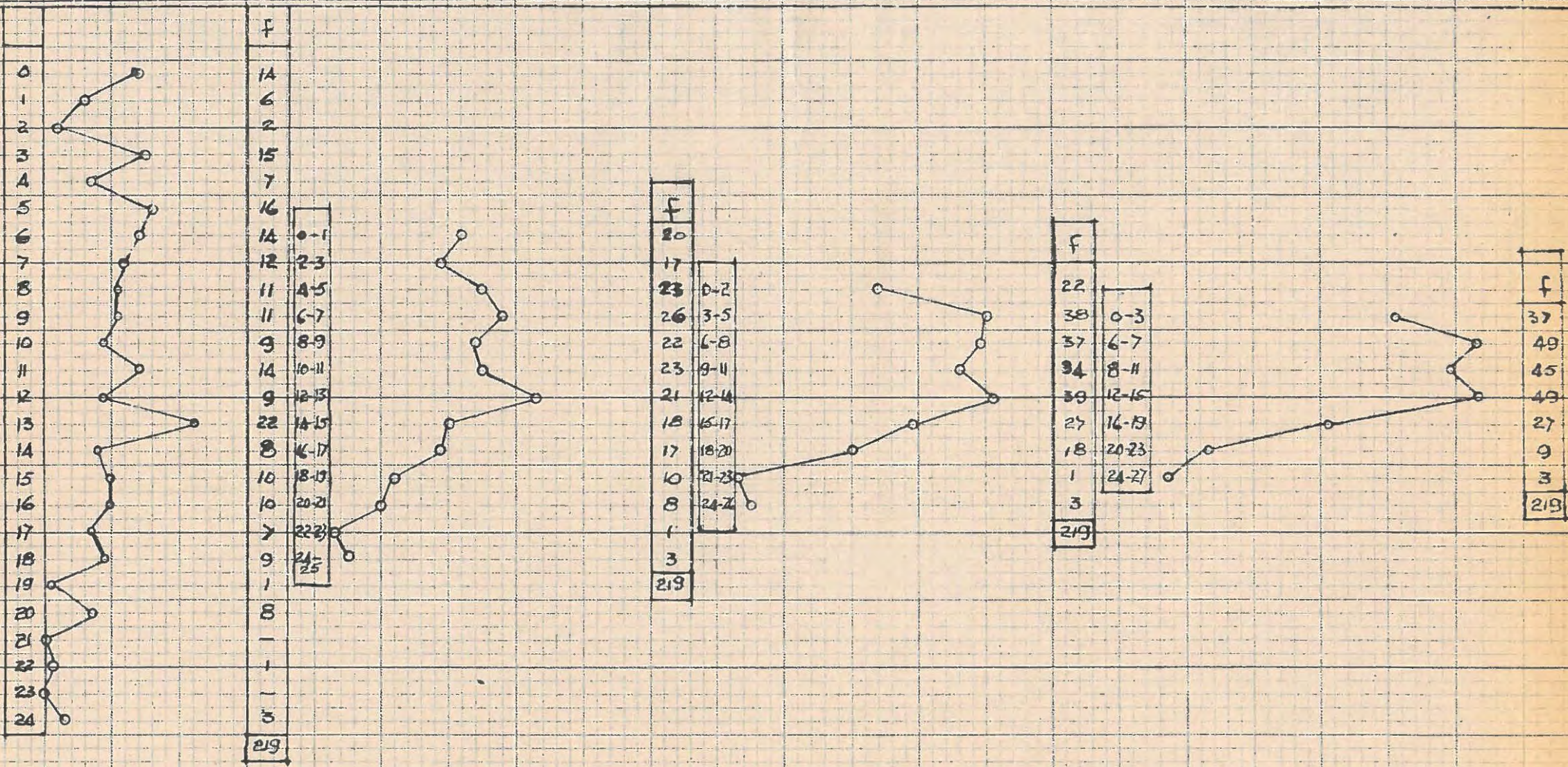
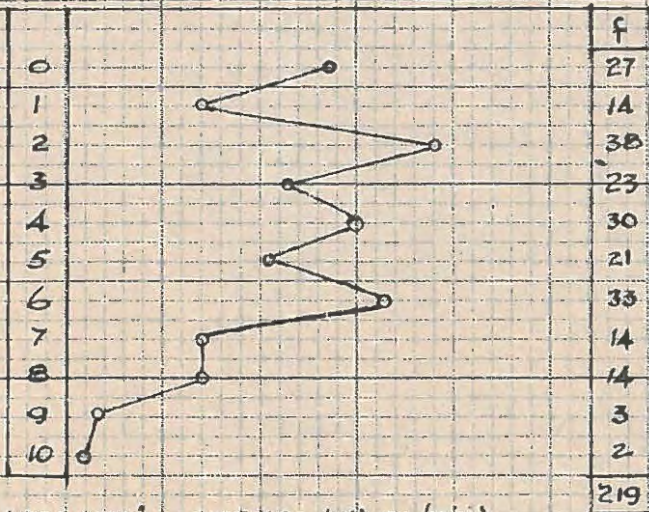
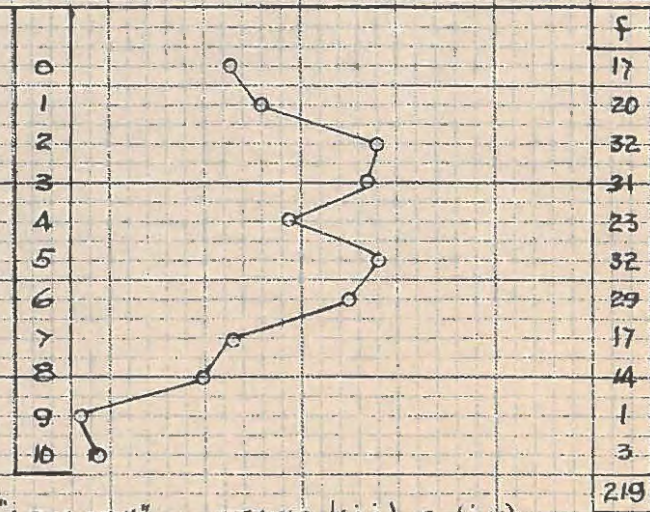


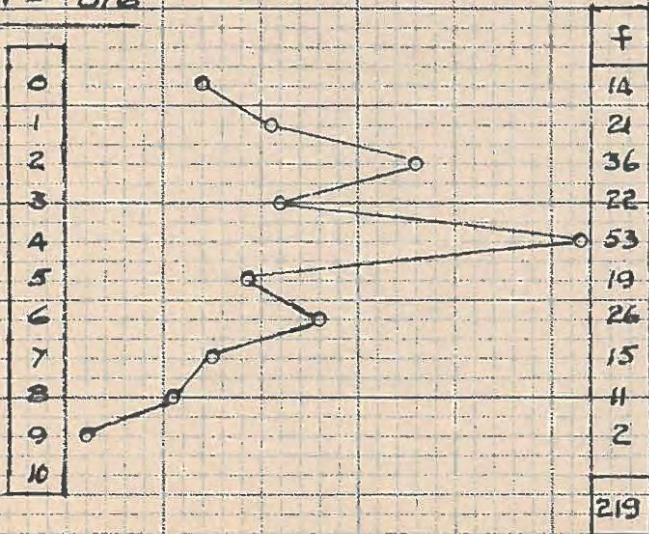
FIGURE 20 WU "ABILITY" TOTAL



"ABILITY" ITEMS (VI) & (VII)
 $r = .876$

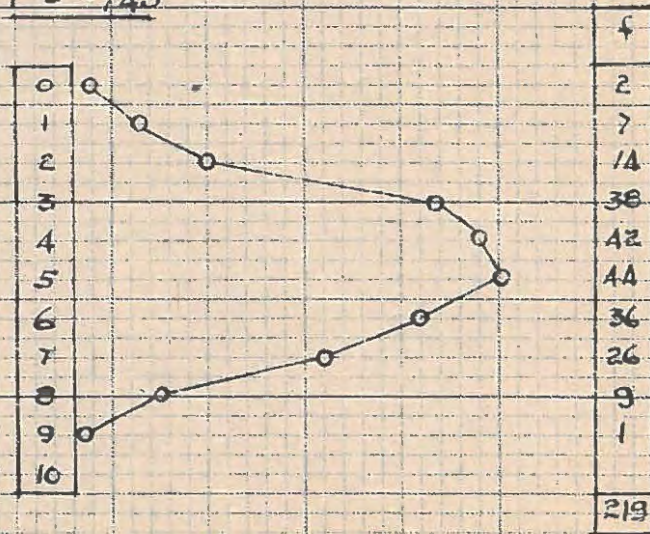


"ABILITY" ITEMS (VII) & (IX)
 $r = .743$



"PERSONALITY" ITEMS (II) & (V)
 $r = .812$

FIG. 21a.



"PERSONALITY" ITEMS (I) & (II)
 $r = .470$

FIG. 21b.

ability to maintain true objectivity. The writer believes that evidence of this weakness, if it does in fact operate, is more likely to be seen in overall assessments (Figs. 4, 8, 9 and 10) than in specific, directed ratings (Figs 6 and 7)

This initial view was amended as a result of the investigations which followed.

The rating scale for the individual items is 0 - 5 (6 class intervals) whereas the rating scales for the distributions of total scores consist of 13 (Figs. 19 (i) and (ii)) and 15 (Fig. 19 (iii)) class intervals respectively. The possibility was considered that the larger rating scale (more than double) revealed the assessors' inability to maintain true objectivity which was obscured in the more compact rating scale of 6 class intervals. The distributions (Figs. 20(i) and (ii)) with progressively condensed rating scales illustrate the basis of this supposition.

An examination of these distributions (Figs. 20 (i) and (ii)) shows that, at any stage of the condensation of the rating scales, the frequency distribution of the "Ability" Total has more modes than that of the "Personality" Total. This is contrary to expectations owing to the original tentative hypothesis (page 29) that lack of complete objectivity was, in some form or other, the root cause of the feature. If this view was reliable the converse tendency would have been observed in these distributions, the reasoning for this statement being as follows:-

- (a) the assessment of an item which allows direct observation to be made with assurance will be more reliable than the rating of an item which is subject to the influences of the assessor's opinion.

- (b) the individual items listed under "Ability" enable direct observation of fact to be made more readily than do the items listed under "Personality".
- (c) consequently ratings by the same assessor (irrespective of the measure of consistency or inconsistency maintained) of both sets of items will be more reliable in the case of the "Ability" items.

The assessments which are accordingly expected to be more reliable (the "Ability" items) are the ones which, when totalled and redistributed, reveal the more numerous modes. In view of the reasoning above this observation suggests that either (1) the effect of halo (in this case the influence of assessor's opinion), if present, contributes but little to the production of bimodality.

or (2) the logic of the above reasoning is at fault.

(Subsequent findings tend to confirm the acceptability of the first suggestion rather than of the second). For the present it is the writer's view that as evidence of inconsistency in the interviewees' assessments cannot be found it is unlikely to provide the main explanation of the several modes in the frequency distributions of the totalled individual items (before condensation of the rating scales). But further treatment of these assessments is necessary in order to confirm (or dispel) this view in a more convincing manner. This will be achieved by mustering evidence to substantiate the contention that multimodality can be caused (as in this case) by consistent statistical determinants.

2. A statistical explanation of multimodality.
Intercorrelations of the assessments of the ten

individual items were calculated as below (throughout this paper the Product Moment Method has been used. See annexure A):-

TABLE 1 Intercorrelations of "Personality" items.

	i	ii	iii	iv	v
i. Self confidence		.782	.470	.660	.730
ii. Presence			.644	.737	.812
iii. Pleasantness of manner				.766	.606
iv. Ability to impose personality					.754
v. Ease of establishing social contact.					

TABLE 2 Intercorrelations of "Ability" items.

	vi	vii	viii	ix	x
vi. Perseverance in getting at facts		.876	.838	.751	.845
vii. Use of questions			.874	.743	.831
viii. Use of data supplied				.758	.819
ix. Effective control of situation					.788
x. Appreciation of situation's requirements.					

In all the above cases the regression lines are linear and the critical ratios are significant. It is evident that (1) the intercorrelations of the "Ability" items are on the whole higher than those of the "Personality" items (mean r of "Ability" items is .812; for "Personality" items it is .696).

(2) The frequency distribution of the "Ability" Total has more modes than that of "Personality" Total (Figs. 20 (i) and (ii)).

In an attempt to determine the factors producing these

characteristics, the scores of the two "Personality" items (ease of establishing social contact, presence) and the two "Ability" items (getting at facts, use of questions) with the highest correlation are added and distributed. Frequency distributions of the totals compiled from the two "Personality" items (self-confidence, pleasantness of manner) and the two "Ability" items (effective use of questions, effectiveness of control of situation) with the lowest correlations are also studied (Figs. 21 (a) and (b)). It is found that the higher the r the more striking the multimodality of the distribution; also, in agreement with the theory presented below, when the r between two items is sufficiently high to result in multimodality, the modes are found corresponding to the 1st., 3rd., 5th., etc., class intervals. (See model below).

Accordingly, though it is impossible to establish the statistical significance of these multimodalities, the writer believes that they are caused by the high correlations between the assessments of the items which are added to form the "Ability" Total and "Personality" Total. The evidence (Figs. 21 (a) and (b)) supports this contention satisfactorily. It must be emphasised, however, that the comparative differences between r 's, as evidenced by the number of modes in the distributions, will be seen readily and will assume significance only if the differences are considerable. The last point becomes clear when considering an extreme case of high positive correlation as below

TABLE 3

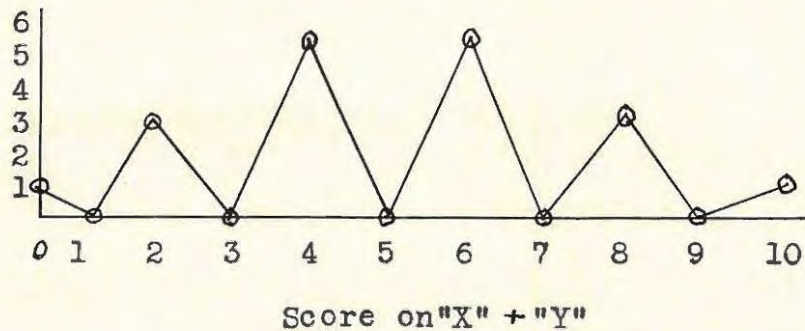
Candidates	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Score on "X"	0	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	5
Score on "Y"	0	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	5
Score on "X" + "Y"	0	2	2	2	4	4	4	4	4	6	6	6	6	6	8	8	8	10

TABLE 4

Bivariate Distribution

		Score on "Y"					
		0	1	2	3	4	5
Score on "X"	0	1					
	1		3				
	2			5			
	3				5		
	4					3	
	5						1

A frequency distribution of the scores of "X" + "Y" will have several modes.



Confirmation of the views presented above is given by J.S. Maritz's (Ref. 6 pp 1 et seq.) explanation of the theoretical approach to this phenomenon. He shows that if sets of correlated scores which, when distributed form typical Poisson distributions are added and redistributed without condensing the rating scale, marked multimodality usually emerges in the new distribution.

Recapitulation at this stage will facilitate further progress.

(a) It was suggested (page 29) that some form of bias caused the multimodalities of the "Ability" Total and "Personality" Total frequency distributions; it was also reasoned (page 34) that the "Personality" items were more susceptible to inconsistency of this nature.

(b) Contrary to expectations, therefore, the several modes were eliminated more readily in the "Personality" Total frequency distribution as a result of the condensation of the rating scales.

(c) However, it was demonstrated and explained (pages 36 and 37) that the addition and redistribution of the values of items which correlate highly results in frequency distributions with more than one mode.

(d) It has been seen that the intercorrelations of "Ability" items are generally higher than those of "Personality" items (page 36).

(e) It is therefore suggested that the reason for the multimodality in the frequency distributions of "Personality" Total and particularly "Ability" Total is statistical rather than one involving the influence on the assessments of halo or assessor's opinion.

3. The reason for the high correlation between individual items within the "Personality" and "Ability" fields.

(i) The existence of true correlation between those aspects of "Personality" and "Ability" which were assessed in the Trial Interview (This refers not only to items within each field but also to the correlation between the two fields. The r of "Personality" Total and "Ability" Total ($N = 219$) is .709 (see page 42)).

(ii) Interaction between the two fields. Though accepting the existence of true correlation it is not suggested that this is absolutely rigid and constant. Rather, a situation is envisaged where interaction between these fields allows for correlation as well as variability in the influences of each upon the other. Support for this view is given in the report by S. Biesheuvel on the findings

of a similar study (Ref. 9 pp. 18, 19). He says:-

".....whereas there is a good deal of independence in the personality assessments the latter is nevertheless weighted by poor skill performances. This can only be due to the fact that a man who has difficulty with a test is more likely to reveal his weakness than one to whom the test is plain sailing. This finding therefore points to "true" interaction rather than to halo effect as the more dominant influence."

(iii) Halo effect is the tendency, assisted in this situation by insufficiently differentiated items, to assign a value to an item according to the value of the general impression. Despite the assessors' training and undoubted efforts to avoid this tendency, it is suggested that it operated considerably owing to (iv) below.

(iv) Selection and definition of rated items.

Due to the characteristics of the interview situation and the role of the assessor (i.e. the interviewee who had no direct control on the satisfactory conduction of the interview), it was difficult to include more dissimilar items and thus obtain greater coverage of the interviewer's "Personality" and "Ability" attributes. Notwithstanding careful and thorough briefing, it is impossible to accept that the assessors were able to view the listed items as individual and separately assessable items. (This weakness is pointed out while remaining fully aware of the dynamic organisation of personality). As a result considerable overlapping of rated items surely exists and the correlations have received artificial boost (page 36).

The above points attempt to explain why the correlations are high. But it is not implied that their correction would reduce the correlation to zero. There is certainly a measure of correlation between the items but it

is impossible to ascertain its extent absolutely in view of the existence of influencing factors. If a revised list of items, freed from the above two main sources of possible interference, were rated, correlation would still exist and an addition of the scores of the individual items would still result in multimodal frequency distributions, though no doubt of a less pronounced nature.

At this point progress can be stated briefly as follows. A statistical explanation has been offered for the multimodality of the frequency distributions of "Personality" Total, "Ability" Total and "Personality + Ability" Total. On the other hand, whatever regrouping has been attempted, it has not been possible to remove (or explain) the bimodality of the Overall Assessment's frequency distribution. The next step is therefore a comparative examination of these assessments.

C. A comparative study of the Overall Assessments and the assessments of the "Personality" and "Ability" items.

The assessments of specific items when distributed, produce the observed unimodal distributions (Figs. 6 and 7). The multimodalities evident particularly in "Ability" Total and "Personality + Ability" Total have been explained statistically. On the other hand the distribution of the scores of the Overall Assessments results in a frequency distribution with pronounced bimodality, the modes being separated by two class intervals. It is anticipated that a review of these essential differences may be of assistance in identifying the causal factor(s) responsible for them.

Treatment of the ratings in Chapter III Section B 1 has strengthened the view that the approach of the assessors was consistent. Further light may therefore be thrown on the characteristics of the Overall Assessments by a comparative examination of the two sets of ratings for the

total group (N = 219).

TABLE 5

	Overall Assessment	"Ability" Total	Test A(F)
"Personality + Ability" Total	.944	--	.370
"Personality" Total	.837	.709	.342
"Ability" Total	.908	--	.348

The regression lines are linear and the critical ratios are significant.

The following points draw attention:-

(i) Test A(F) measures general mental ability.

The correlation coefficients of the scores of this test and , firstly, "Ability" Total and, secondly, "Personality" Total differ only in the third decimal figure; there is therefore virtually no difference in the amount of correlation. This is unexpected despite the accepted correlation between the "Personality" and "Ability" fields (page 39). The listed "Ability" items are associated with the existence, development and training of general mental ability. This leads to the view that the correlation of "Ability" items and Test A(F) should be higher than that of "Personality" items and Test A(F). The implication of the finding that this is not so must be examined fully:-

(a) Though the r's of both "Personality" Total and "Ability" Total with A(F) are comparatively low they are significant (critical ratios are 5.712 and 5.848 respectively).

(b) "Personality" Total and "Ability" Total correlate to the extent of .709. This is contributed to partly by actual correlation between the two sets of items (page 38) and partly by halo effect. It is difficult to make a very precise statement as to the relative contribution of each to the resulting $r = .709$. The partial correlation between these variables has not been calculated

- because (1) it is impossible to hold constant each of the variables experimentally.
- (2) it is difficult to make a satisfactory statement as regards the consistency of the correlations among themselves.

McNemar mentions these difficulties in connection with the use of partial correlations. (Ref. 8 pp. 141, 142).

Garrett also advises caution in the use of partial correlations. (Ref. 7 pp. 455, 456). He says:-

"When we "partial out" the influence of clear-cut and relatively objective factors such as age, height, etc., we have a reasonably clear notion of what the "partials" mean."

He adds that "the result is by no means so unequivocal" when there is overlap between the scores (as there surely is in the present case).

However, $r = .709$ is too high to represent the true extent of the correlation between the items; it is suggested that the influence of halo must be accepted.

(c) Reasons have been given why it is expected that the correlation coefficient of "Ability" Total vs. Test A(F) should be higher than that of "Personality" Total vs. Test A(F). The fact that the anticipated difference is not found suggests that either (1) the above reasoning is at fault or (2) the extent of the influence which one set of items had on the other was sufficiently large to obscure the true relationship of each with Test A(F).

(d) In deciding which alternative to favour the following aspects exert considerable weight.

(i) the acceptance of halo as a contributing factor (page 40).

(ii) the view that "Ability" items allow for more consistently reliable assessments than "Personality" items (page 35).

(e) As a result of a full consideration of all these factors, alternative C (2) is put forward as the more valid. Though it therefore enjoys preference over alternative C (1), attention must be drawn to the fact of interaction between the two fields as stated in paragraph (ii) page 39. The influences have certainly not been one-sided; it is suggested, however, that the influence of the "Ability" field on the "Personality" field may have been far greater than that of the "Personality" field on the "Ability" field.

(ii) The high correlation coefficients found when correlating the Overall Assessment with "Personality + Ability" Total (.944), "Personality" Total (.837) and "Ability" Total (.908) indicate extremely good agreement between the assessments of the "Personality" and "Ability" Totals and the Overall Assessments. An explanation for this is partly supplied by the influences tabulated on pages 39 and 40. The particularly high agreement in the first instance (.944) leaves but little room for improvement and warrants a closer examination.

The high r of .944 is obtained by correlating the Overall Assessment with the "Personality + Ability" Total. It has been seen that no definite evidence of lack of consistency while obtaining the latter can be established. Moreover, it has been demonstrated that the unimodal distributions are replaced by multimodal distributions when the ratings of highly correlated items are totalled and distributed anew (before condensing the rating scales). It has been suggested that this statistical fact satisfactorily explains the form of the frequency distributions of "Personality"

Total, "Ability" Total and "Personality + Ability" Total (Figs. 17 (i)(ii) and (iii)). But as yet no explanation can be given for the bimodality of the Overall Assessments' frequency distribution.

An examination of the relative frequency distributions Figs. 4 and 20 (iii) (each with nine class intervals) shows an important difference unexpected initially owing to the high correlation. The similarity evident in the half of the scale representing the higher values is not maintained in the class intervals representing the lower values where the second mode is found. In the Overall Assessments' distribution the two distinct modes are separated by two class intervals while the distribution of the "Personality + Ability" Total is unimodal. (The rating scale has been condensed to nine class intervals to minimise the effect of statistical influences on the frequency distribution).

In the light of the above evidence it appears reasonable to believe that the process of rating the Overall Assessments may have been unavoidably hampered by weaknesses inherent in the test or test situation; whatever these weaknesses they did not influence the assessment of individual items to the same evident extent. However, in view of the high r of .944 it is clear that the rank order remains largely unchanged in both sets of assessments.

These views suggest a closer examination of the Trial Interview's assessments and the raters' comments. The writer hoped that this would reveal factors causally related to bimodality which up to this point have escaped detection.

D. The inconsistency of assessments.

Firstly, the single Overall Assessment is obtained mainly by considering the existing values in the "Personality" and "Ability" Fields. The imposition of one set of values on

another (without the actual addition of the ratings) may result in bimodality of the frequency distribution. An examination of the characteristics of the distributions (Figs. 20 (i)(ii) and (iii) with nine class intervals) does not strengthen the view that this is the possible explanation. It must also be remembered that additional aspects, beyond the confines of the "Personality" and "Ability" spheres, may have exerted influence on the Overall Assessment.

Secondly, two sets of assessments have been made in the present test situation.

(1) When the assessor's attention is focussed directly on one item only an acceptable standard of consistency has been attained as demonstrated. No unusual characteristic is apparent when the rating scales have been condensed to reduce the effect of the statistical influences on the forms of the frequency distributions of "Personality" Total and "Ability" Total.

(2) When the assessors consider the interviewer and his attributes with a view to making an all-embracing overall assessment they are less able to direct and control the consistency of their evaluations. This tendency, phrased in slightly more specific terms, is assignable to the effect on the ratings which is exercised by the initial or principal impression of the interviewer on the interviewee. The result is that the assessments are not all made with reference to a constant yardstick and an unreliable set of values is superimposed on the reliable ones. This can be expressed in a different way. Bimodality would emerge if "liked" interviewers tended to obtain higher-than-warranted assessments and "disliked" interviewers tended to obtain lower-than-warranted assessments. It was considered that evidence indicating the operation of subjective influences consistent with these views would further the interests of the investigation.

Accordingly a detailed study of each of the interviewees' reports (Chapter II section C) was undertaken in order to obtain, if possible, indications which would support this contention or yield additional material necessary for a new approach.

E. A study of the interviewees' reports.

(1) Inadequate interviews.

The interviewee, on account of the characteristics of the test situation, had little direct control over either the manner of handling or the duration of the interview. In several cases the interview had lasted such a short time that mention was made of the fact in such terms:-

"Short, incomplete interview."

"Asked two questions and terminated interview."

"Terminated interview after a few minutes." etc.

In other instances, though the interview may have lasted a reasonable length of time, comments of another nature indicated that the interview as such had been unsatisfactory. A sample of these comments will illustrate what is signified by the term unsatisfactory.

"The interview was entirely farcical, the candidate being fully aware of the interviewee's real identity."

"Didn't ask one question. Went on droning in a monotonous voice reading from the assignment."

"Had prepared a speech in advance which he delivered to Miss van Jaarsveld in 15 minutes; didn't seem to expect anything from her except a good hearing." etc.

Both short and unsatisfactory interview are obviously inadequate performances in the test situation. The instructions to the interviewers were clear and unambiguous. In the above cases these instructions were not carried out;

in other words, the test has not been done. All the interviews which are either short or unsatisfactory have therefore been called "Inadequate."

(2) Method of separation.

Although the effect of inadequate interviews is not the same for all the assessments (see (3) below), it was decided to separate all the inadequate interviews from the adequate interviews. All the interviewees' reports were studied carefully by three independent judges who effected a dichotomous separation in terms of adequacy and inadequacy. The conception of inadequacy embraces both short and unsatisfactory interviews (page 47). Before labelling an interview as either adequate or inadequate an answer had to be given to the question

"Did the interviewer carry out his instructions and perform the test?"

Throughout the process of separation care was taken to differentiate between a negative answer to this question and the case where a poor interview was conducted on account of poor understanding of and ability in the situation. In the latter case the assessment must be low; in the former case, though the assessment is also low it is not necessarily representative of the interviewer's overall "Personality" and "Ability" qualifications.

It must be emphasised that the instructions to the judges were to maintain a rigid standard of inadequacy; if any doubt existed an interview was not to be considered as inadequate. Only the interviews which were considered by all three judges to be inadequate (N = 53) were extracted from the total group. Accordingly it is conceivable that some inadequate interviews remained with the adequate ones.

But, in the interests of the scientific treatment of the results, the purity of the group of inadequate interviews is of primary importance.

Contingency coefficients have been calculated for the separations of each two of the three judges and the results are (annexure B):-

TABLE 6

	χ^2	P
A - B	108.476	<.001
B - C	92.370	<.001
A - C	118.207	<.001

In each case, therefore, there is overwhelming evidence that a definite selective process was in operation.

It is impossible to draw too much attention to the vital fact that two of the three judges had not seen the bimodal distribution before selecting the inadequate interviews. They therefore had no idea of the implications of extracting interviews with any particular assessment. In fact, the interviewees' reports formed the sole basis of the process of separation absolutely irrespective of any of the assessments.

(3) The assessments of candidates whose interviews are inadequate.

Having established that a considerable number of the interviews were inadequate, it is necessary to consider to what extent this will have influenced the assessments made immediately the interview is ended. The situation as presented by an inadequate interview is certainly not conducive to accurate assessments, but the effect will not be the same for both sets of assessments.

(1) Assessments of "Personality" and "Ability" Items.

An inadequate interview is labelled as such

on account of the apparent failure of the candidate to carry out his instructions to interview Miss van Jaarsveld. He has therefore made it extremely difficult for the interviewee to evaluate his abilities in considering and dealing with human material; she is unable to pass reliable judgment on whether he accepts facts at their face value or whether he is an individual who has a sense of proportion when faced with the task of employee placement. However, it has nevertheless been possible for an assessor to appraise certain more specific aspects of the "Personality" and "Ability" fields; his self-confidence, pleasantness of manner, use of data supplied, etc., may be measurable with varying degrees of accuracy. Two tendencies have to be mentioned to guard against their being overlooked on account of generalisations; the writer believes that

- (a) The degree of reliability of assessments will usually be less for short interviews than for unsatisfactory interviews (refer to explanations on page 47).
- (b) The degree of reliability will usually be greater for assessments of "Ability" items than for those of "Personality" items (page 35).

In Section C it was found that the correlation coefficient between "Personality" items and "Ability" items is .709 ($N=219$). However, the views presented above have indicated that the effect of halo (page 35) on the assessments would tend to be emphasised by inadequate interviews.

A further correlation coefficient was therefore calculated between the "Personality" and "Ability" items of adequate interviews and it was found to be .644. The reduction in r from .709 to .644 is small and may represent the partly removed influence of halo present par-

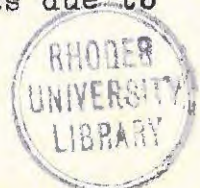
ticularly when the assessments of inadequate interviews are made. On the other hand the small change in r may be attributable to a changed sample and therefore no definite conclusion can be drawn from it.

Despite the halo effect which .644 still suggests, it is considered that inadequate interviews do not invalidate the assessments of individual items. At most the assessment of some of the items in some of the cases are less reliable than if the interview had been adequate (^{pp.} 49⁻⁵⁰ and above). Consider, for example, the case of an inadequate interview. Seven out of the ten individual items may have been sufficiently observed to enable reliable assessments to be made. The next inadequate interview may also enable the reliable assessment of seven items to be made but they will not necessarily be the same seven items referred to in the first example. And so on; eventually, though 53 of the 219 interviews are inadequate, the number of inconsistent assessments of each individual item will definitely be less than 53 (this refers to inconsistency due to inadequacy of the interview). The high proportion of inadequate to adequate interviews (53 : 166) will thus be considerably reduced. Consistent with these views, the overall effect of inadequacy on the specific items' assessments for the 219 cases has been insufficient to result in an unusual form of the frequency distribution such as was seen in the Overall Assessments' bimodality.

(2) The Overall Assessments.

On account of the broader, less specific view which this assessment necessitates, a particularly brief or unsatisfactory interview is liable to be influenced unduly by (a) a strikingly favourable or unfavourable quality or characteristic.

(b) the unavoidable (largely) influence of bias due to



the formation of impressions of liking or disliking. The operation of these influences is enhanced by the fact that the interviewee's observations have been seriously curtailed and she has less founded judgement on which to base her assessments. The implication of this deficiency assumes significance when remembering the extent to which thorough observation was meant to be possible during the Trial Interview situation. The view is held, therefore, that the Overall Assessments of inadequate interviews are suspect.

(4) The frequency distributions of adequate and inadequate interviews.

(a) Adequate interviews.

A re-distribution of the Overall Assessments of the adequate interviews (Fig. 22 (a)) produces a frequency distribution whose characteristics show little trace of the original bimodality. Only one class interval separates the two modes which are no longer as pronounced as in the original frequency distribution with $N = 219$. This new distribution (Fig. 22 (a)) has lost the significance of the first distribution's unusual form (Fig. 4 and page 18). The assessments excluded, therefore, appear to be those whose objectivity and reliability had been questioned; but it is nevertheless probable that in the case of some of the longer and apparently adequate interviews hardly any additional observation was possible.

(b) Inadequate interviews.

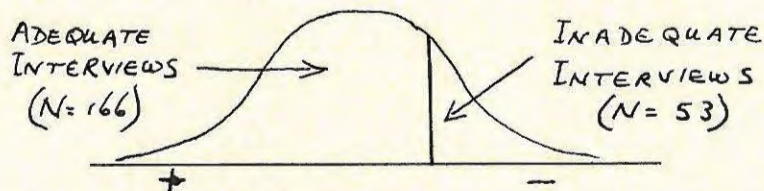
The frequency distribution of the Overall Assessments of inadequate interviews (Fig. 22 (b)) shows a definitely skewed pattern. The scatter of ratings covers most of the scale, viz., B+ to C-. The class interval with the highest frequency of cases is C, which corresponds with the second mode of the original bimodal fre-

quency distribution. It would seem that, following short and unproductive interviews, the assessors' reactions are obedient to certain apparently similar processes of reasoning. The fact of having had a short and unsatisfactory interview is not indicative of thorough preparation, method or ability; moreover little has been observed. Consequently the rating must be a low one, but, unless a particularly poor performance earns the lowest assessment, there appears to be a tendency to give an insufficiently observed candidate the second lowest assessment possible.

(5) The correlation of Trial Interview assessments with the scores of the remainder of the battery's tests.

It is necessary at this stage to examine the extent to which the correlation coefficients of the Trial Interview's assessments with the battery's tests have been affected by the inadequacy of almost 25% of the interviews.

(a) Before examining the coefficients in Table 7, it is pertinent to consider whether the dichotomous separation of interviews has resulted in the truncation of the group. In other words, can the group of candidates who conducted the inadequate interviews be closely identified with the group of candidates whose test scores and assessments were low? Truncation can be explained diagrammatically as follows



An examination shows that the frequency distributions of both adequate and inadequate interviews (Fig. 22 (a), 27, 28, 29, 30 and 22 (b), 23, 24, 25, 26) are similar in form and the possibility that truncation may have affected the situation is not borne out. The most striking dissimilarity is evident when the Overall Assessment's frequency

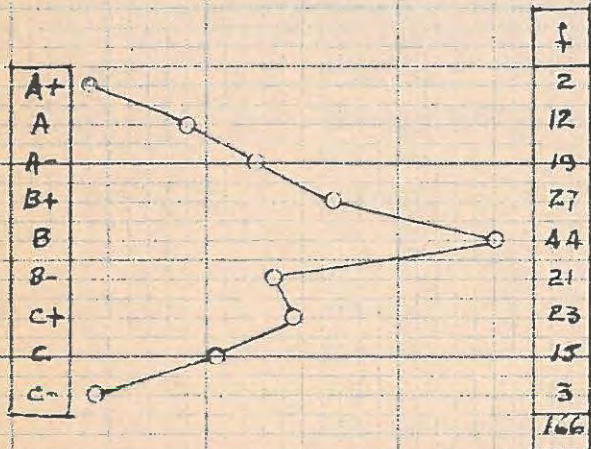


FIG. 22a OVERALL ASSESSMENT

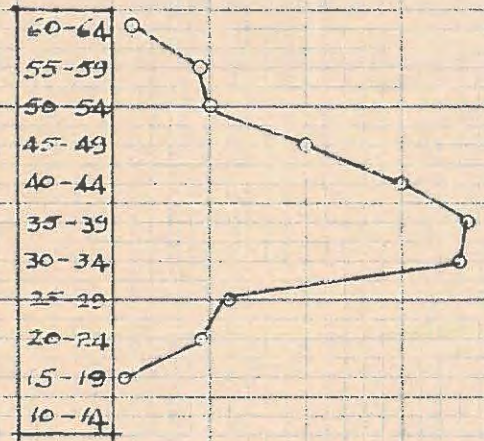


FIG. 27 TEST A(1)

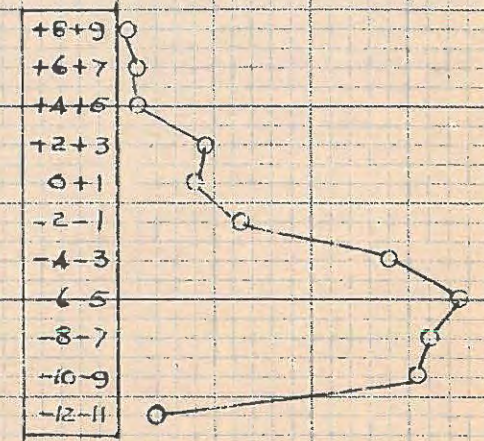


FIG. 28 WRITTEN PROJECT

FREQUENCY DISTRIBUTION OF ADEQUATE INTERVIEWS

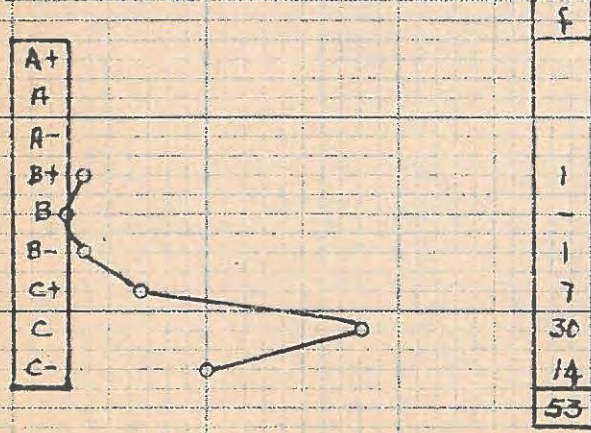


FIG. 22b OVERALL ASSESSMENT

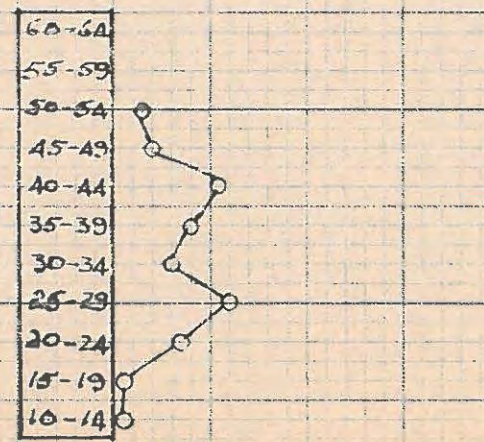


FIG. 23 TEST A(1)

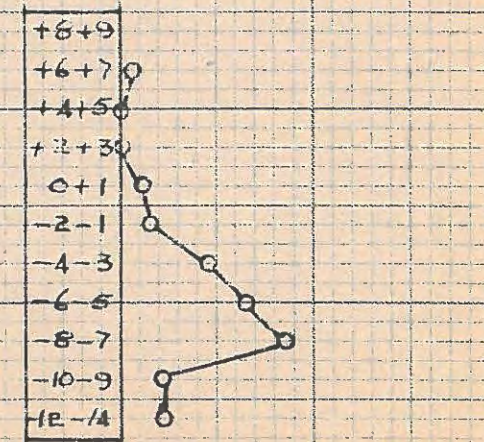


FIG. 24 WRITTEN PROJECT

FREQUENCY DISTRIBUTIONS OF INADEQUATE INTERVIEWS

-53a-

40-42
37-39
34-36
31-33
28-30
25-27
22-24
19-21
16-18
13-15
10-12
7-9

FIG. 29. TEST - M

FREQUENCY DISTRIBUTIONS OF ADEQUATE INTERVIEWS

f
5
11
25
38
37
26
20
6
2
2
166

+6+7
+4-5
+2+3
0+1
-2-1
-4-3
-6-5
-8-7
-10-9
-12-11

FIG. 30. FINAL RATING

f
4
6
7
11
19
25
26
33
21
14
166

40-42
37-39
34-36
31-33
28-30
25-27
22-24
19-21
16-18
13-15
10-12
7-9

FIG. 25. TEST M.

f
1
6
13
9
13
2
5
3
1
53

+6+7
+4+5
+2+3
0+1
-2-1
-4-3
-6-5
-8-7
-10-9
-12-11

FIG. 26. FINAL RATINGS

f
-
1
-
2
1
3
7
13
10
16
53

FREQUENCY DISTRIBUTIONS OF INADEQUATE INTERVIEWS

-536-

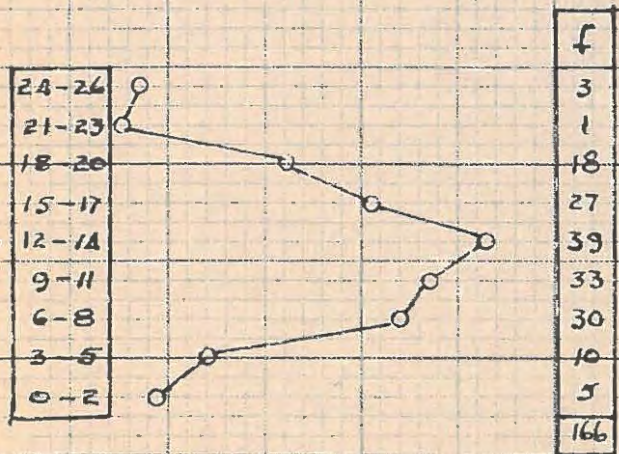


FIG. 31 "ABILITY" TOTAL FOR ADEQUATE INTERVIEWS

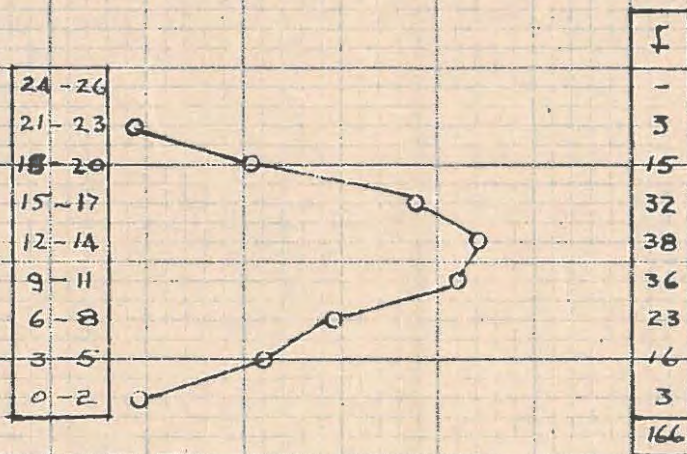


FIG. 33 "PERSONALITY" TOTAL FOR ADEQUATE INTERVIEWS

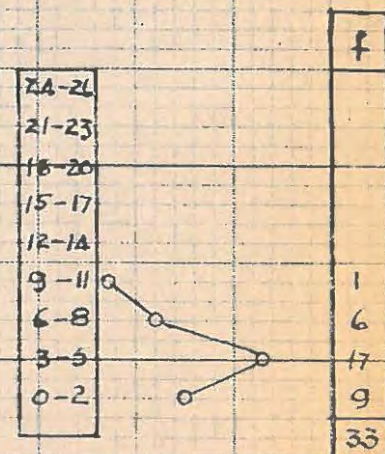


FIG. 35 "ABILITY" TOTAL FOR INADEQUATE INTERVIEWS

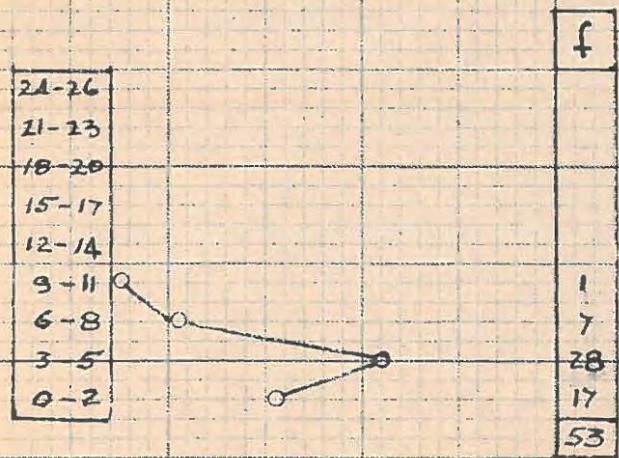


FIG. 32 "ABILITY" TOTAL FOR INADEQUATE INTERVIEWS

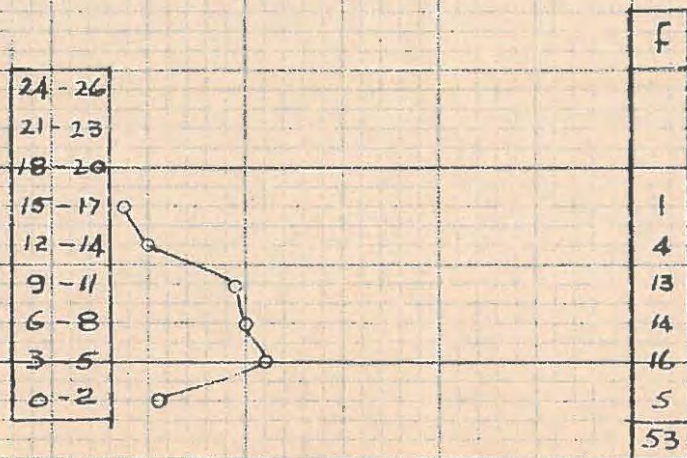


FIG. 34 "PERSONALITY" TOTAL FOR INADEQUATE INTERVIEWS

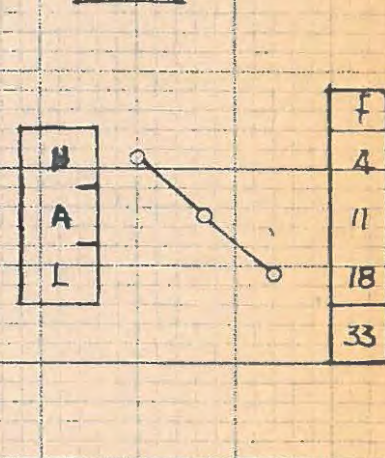


FIG. 36 C5+6 FOR INADEQUATE INTERVIEWS

-53C-

distributions are examined; but this is not unexpected in view of the suspicion of inconsistency concerning these assessments on account of the inadequacy of such a high proportion of the interviews.

The distributions of "Ability" Total and "Personality" Total in groups of adequate and inadequate interviews were examined (Figs. 31, 32, 33 and 34). In the case of the "Ability" Total especially it appears that the inadequate interviews' assessments are clustered around the end of the rating scale representing the lower values. At first this may be disturbing, but two observations have to be made:-

- (i) On page 16 it was stated that the assessments of individual items indicated the extent to which each particular attribute had been observed during the Trial Interview situation. As the inadequate interviews allowed only limited observations to be made, the assessments are expected to be low.
- (ii) The r's of the "Ability" Total (Table 7) with the remainder of the battery's test scores, whether $N = 166$ or 53 , are very similar (none of the t values shows a significant difference). The cluster of assessments evident when $N = 53$ has therefore not unduly disturbed the relationships with the other tests.

Having seen that the candidates who conducted adequate ($N = 166$) and inadequate interviews ($N = 53$) are, according to the scores of the battery's tests, randomly selected samples from the whole group ($N = 219$), the following r's were calculated. The purpose was to study the differences, if any, caused by the division of the interviews into groups of adequacy and inadequacy.

TABLE 7

	A(F)	M	W.P.	F.R.
Overall Assessment (N=166)	.33808	.30628	.19847 ^x	.49129
ditto (N = 53)	.13991 ^x	.10110 ^x	.19909 ^x	.04556 ^x
"Ability" Total (N=166)	.30689	.26806	.13303 ^x	.38860
ditto (N = 53)	.19639 ^x	.20100 ^x	.10346 ^x	.32557
"Personality" Total (N=166)	.30299	.30312	.15903 ^x	.50947
ditto (N = 53)	.17965 ^x	.23122 ^x	.06662 ^x	.25548 ^x

The regression lines are linear. The r's marked ^x are not significant, the critical ratios ranging from .483 to 1.965.

The significance of difference test was calculated for each pair of the above correlation coefficients with the following results: (Annexure C).

TABLE 8

Values of t

	Test A(F)	Test M	Written Project	Final Rating
Overall Assessment (N = 166)	1.306	1.330	.004	3.045
ditto (N = 53)				
"Ability" Total (N = 166)	.731	.439	.186	.447
ditto (N = 53)				
"Personality" Total (N = 166)	.812	.477	.579	1.861
ditto (N = 53)				

The only significant difference in the correlation coefficients is found when the Final Rating is correlated with the Overall Assessment in groups consisting of adequate interviews and inadequate interviews. This is of vital interest.

It will be remembered that the Final Rating assessment is obtained from the ratings of a team of assessors who have closely observed each candidate in two group discussions lasting practically three hours. Apart from

the thorough observation thus rendered possible, the harmful effect of faulty judgement is cushioned by the fact that the Final Rating reflects the views of the whole team and it is extremely unlikely that all the assessors would have made a seriously inaccurate assessment in the case of the same candidate. There is therefore justification for the view that the Final Rating is a valuable measure of a candidate's potential in the administrative field particularly in so far as his personality qualities are concerned (page 7). The importance of the above analysis lies in the fact that the Final Rating and the Overall Assessment, as designed, are comparable assessments (Chapter II, section C (b)) and yet the correlation of these two assessments in respect of independent groups (adequacy and inadequacy of interviews) produce coefficients which differ significantly. The division of the total group (N=219) into adequate and inadequate categories has not resulted in any other significant difference between the correlation coefficients presented in Table 7 on page 55. The detrimental effect of an inadequate interview on assessments is therefore detectable statistically only in assessments which, having laid emphasis on personality attributes, endeavour to quantify potential in the administrative field (Chapter II, section C 3(b)). This view is further supported by the critical ratio in the case of the correlation of "Personality" Total and Final Rating which is the second highest in Table 8 on page 55.

These facts lend weight, if it is necessary, to the consideration of the interviews in terms of adequacy and inadequacy. Further, in view of the inability to extract all the inadequate interviews from the total group (see (2) above; p.48), it also strengthens the contention

that the Overall Assessments are suspect.

(b) The evidence obtained in (a) indicates that the assessments of individual items are less influenced by the inadequacy of interviews than are Overall Assessments. It was consequently decided to construct a correlation table using various combinations of 2 or 3 "Personality" or "Ability" items instead of the Overall Assessments, "Personality" Total and "Ability" Total as above. The frequency distributions of the individual items formed the rough basis for the selection of the items grouped together. (Figs. 6 (c) and 7(c)). Where the distributions were similar in the case of Mrs. B+Rest on the one hand and Mrs. P on the other, it was considered probable that the raters had a comparable standard in mind while making the assessments; it was also considered that in such a case the item being rated presented additional individuality which would facilitate the reduction of the effect of inconsistency. The correlation table differs from the previous one owing to the following adjustments:-

- (i) The contribution which a test makes to the predicting efficiency of the battery can be calculated only when its correlation with the criterion of success (the validity of the test) is available. Criterion values could only be obtained for 123 cases (Ref. 1 pp. 55 - 65). As the ultimate purpose underlying this study is to ascertain which assessment made in the Trial Interview situation contributes most to the final prediction, it was decided to calculate the correlation coefficients for which criterion values (C5 + 6)^{*} have been obtained.
- (ii) Because the assessments of individual items are not influenced by inadequate interviews to an extent determinable statistically, the sample

used represents the total 123 cases irrespective of adequacy or inadequacy.

- (iii) The table of correlation coefficients has been extended to include "Ability" Total and "Personality" Total on the same basis of N=123 for the purpose of comparison.

TABLE 9

	C5 + 6	Test A(F)	Test M	Written Project	Final Rating
Items (iii) + (ix)	.32602	.36913	.40201	.28868	.49635
Items (iii) + (x)	.33569	.35778	.36195	.28398	.49121
Items (iii)+(viii)+(x)	.34088	.33498	.33417	.29682	.49920
"Ability" Total	.37757	.28453	.29147	.28721	.47438
"Personality" Total	.26999	.28667	.34506	.25464	.50741

All the critical ratios are significant. An examination of these correlation coefficients reveals that the various groupings of items do not show any marked tendency to reduce or increase the amount of correlation with the criterion and the tests of the battery. The value of this table will become evident, however, when attention is drawn to the main features which it was hoped to identify in these coefficients.

Firstly, a battery consisting of several tests

* During the year following immediately on the conclusion of the testing procedures considerable work was done to establish a reliable criterion against which the predicting efficiency of each individual test and the battery as a whole could be measured (Ref. 1 pp. 30 - 40). It was found that a combination of criterion 5 and criterion 6 (the value denoted as C5 + 6) produced a final criterion whose correlation coefficient with the final battery score is .67.

the function of each is to provide a measure of an ability or attribute of a particular nature. Each, therefore, has individuality. There will naturally be correlation between the various tests but if each test's score is a true measure of a specific sphere or area, these intercorrelations will tend to remain low.

Secondly, each test must correlate significantly and highly (generally) with the criterion of success.

The above correlation coefficients have therefore been calculated in an endeavour to find an assessment obtained during the Trial Interview, which, while having low correlation coefficients with the tests of the battery, correlates significantly and highly (comparatively) with the criterion of success. An examination of Tables 7 and 9 shows that the "Ability" Total score is the one which best satisfies these requirements.

This finding leads to the next stage in the investigations.

C H A P T E R IV

A STUDY OF THE WEIGHTS OF THE BATTERY'S TESTS.

A. Before purification.

The battery score, which was found to have an $r = .67$ with the criterion (see page 58) is obtained using tests A(F), M, Written Project (W.P) and Final Rating (F.R.). Using the Pivotal Condensation method (Annexure D) the weights of these tests were calculated and found to be:-

<u>Test</u>	<u>Weight</u>	(N = 123)
A(F)	.231	
M	.032	
W.P.	.236	
F.R.	.322	

The calculation of weights was repeated using the Overall Assessment (O.A.) as a 5th test in the battery with the following results:-

<u>Test</u>	<u>Weight</u>	(N = 123)
A(F)	.232	
M	.031	
W.P.	.235	
F.R.	.318	
O.A.	.008	

Clearly the weighted contribution of the Overall Assessment is negligible. For this reason alone its inclusion in the battery cannot be justified; moreover the time and effort spent in administering it is not commensurate with its contribution towards the end for which the test was designed. However, it is important to note that the interviewee's report does provide useful information for the clinical interviewer.

The above weights of the tests were calculated with the whole group (N = 123) for which C5 + 6 values had been

obtained. What, if any, would be the effect on the weights when the inadequate interviews are removed reducing the group to N = 90?

B. After purification.

The new set of weights with the size of the sample reduced by approximately 25% is as follows:-

<u>Test</u>	<u>Weight</u>
A(F)	.296
M	-.001
W.P.	.224
F.R.	.353
O.A.	-.091

The changes are small. Though the numerical contribution of the Overall Assessment has deteriorated it is impossible to attribute the cause to any specific factor. In fact, the variation may be due to the use of data relevant to a considerably altered sample of cases and no conclusions can therefore be drawn from it.

C. The weights of various sets of assessments made during the Trial Interview.

On pages 57 and 58 the reasons were given for attempting to correlate the tests of the battery with various sets of the individual assessments made during the Trial Interview situation. Though the correlation coefficients obtained were not of much use as such, the weights of the various sets of assessments used were calculated because a scientific approach to the analysis of data necessitates an exhaustive study of all possibilities.

The following table was compiled (N = 123).

TABLE 10

Additional Test	Additional Test	Weights				Multiple R
		A(F)	M	W.P.	F.R.	
Overall Assessment	.008	.232	.031	.235	.318	.671
"Personality" Total	-.069	.220	.046	.235	.359	.673
"Ability" Total	.110	.237	.019	.235	.273	.678
"Personality + Ability" Total	.048	.236	.024	.235	.299	.672
Items (iii) + (ix)	-.001	.231	.032	.236	.322	.671
Items (iii) + (x)	.022	.231	.029	.236	.313	.671
Items (iii)+(viii)+(x)	.030	.232	.028	.235	.310	.671

The numerical weighted value of the "Ability" Total in the battery is approximately fourteen times larger than the weight of the Overall Assessment; it is three times as large as the mean weight of all the other new tests examined in the table. To what extent is this apparent difference actually significant?

D. The significance of the weight of "Ability" Total

There is no test for the significance of the weight of a set of assessments. In the first place it will therefore be necessary to examine the available evidence and from it to infer the extent of the significance of the weight of .110. Having done so, the statistical selection of tests for inclusion in a battery will be used to corroborate the view formulated above.

(a) It was explained on pages 48 and 49 that the usefulness of a test in a battery is dependent on two general requirements being satisfied:-

- (i) high correlation between the test and the criterion.
- (ii) low correlation between the test and the remainder of the test in the battery.

Bearing these requirements in mind, Table 11 was

completed as follows. The r's listed under A(F) (N=123) were examined. The lowest r received a score of 1; the second lowest r received a score of 2, etc. This was done for each set of r's listed under the various tests of the battery. The validity coefficients tabulated under C5+ 6 (N=123) were next examined. This time the highest r was given a score of 1; the second highest a score of 2, etc. The whole procedure was repeated for the coefficients calculated with N=90. When this had been completed, the allocated scores were totalled across the page thus producing a grand total score for each new test or set of assessments.

TABLE 11

Test	N	A(F)	M	W.P.	F.R.	C5+6	Grand Total Score
"Ability" Total	123	.28453 ²	.29147 ¹	.28721 ²	.47438 ¹	.37757 ¹	7
Overall Assessment Items (iii) + (ix)	123	.28295 ¹	.30562 ²	.35081 ⁵	.51896 ⁴	.33921 ³	15
"Personality" Total	123	.36913 ⁵	.40201 ⁵	.28868 ³	.49635 ²	.32602 ⁴	19
"Personality + Ability" Total	123	.28667 ³	.34506 ⁴	.25464 ¹	.50741 ³	.26999 ⁵	16
"Ability" Total	123	.29755 ⁴	.32790 ³	.30734 ⁴	.51937 ⁵	.35370 ²	18
"Ability" Total	90	.30561 ¹	.26701 ¹	.23074 ¹	.42175 ¹	.26886 ²	6
Overall Assessment Items (iii) + (ix)	90	.32928 ³	.33759 ²	.33045 ³	.51885 ³	.26353 ⁴	15
"Personality" Total	90	.40037 ⁴	.40567 ⁴	.53118 ⁴	.50685 ²	.30350 ¹	15
"Personality" Total	90	.31746 ²	.35460 ³	.26046 ²	.51912 ⁴	.26771 ³	14

^xThis r is not significant (critical ratio = 2.299).

It will be evident that the test with the lowest grand total score is the one which conforms more satisfactorily to the above requirements. Furthermore an inspection shows clearly that it is the "Ability" Total which best satisfies these general requirements whether the inadequate interviews are removed from the group or not.

Ofcourse this finding does not immediately endow the weight of .110 with adequate significance. Rather, it merely serves to confirm the original view that the "Ability" Total, less affected by inadequate interviews, will be a more meaningful assessment than the others made in a situation which has in-

sufficient safeguards to ensure a higher proportion of adequate interviews.

It is necessary, however, to examine more closely the view that the assessments of "Ability" items are unique among the assessments made during the present Trial interview and the following table is drawn up:-

TABLE 12

	C5 + 6		
	N = 123	N = 90	N = 33
"Ability+Personality" Total	.35370	.28284	.18023 ^x
"Personality" Total	.26273	.26771	-.16992 ^x
"Ability" Total	.37758	.26866	.54284
Overall Assessment	.33921	.26353	.15123 ^x

^xThese r's are not significant.

When the group consists largely or wholly of inadequate interviews (N = 123 or N = 90) there is at the most only a small difference between the validity coefficients of the assessments. However, the validity coefficient of "Ability" Total is the highest in the table compiled for N = 123 (the sample used for the calculations of weights on page 60); though the difference is not large the "Ability" Total is unique not by virtue of this fact alone but, as explained on pages 58, 59 and 62, in respect of the low intercorrelations with the battery's other tests as well (Table 11).

The point in Table 12 which attracts particular attention is that the "Ability" Total's validity coefficient is .26866 when the interviews are adequate (N = 90) whereas for inadequate interviews (N = 33) it is .54284. Although the difference between these two r's is not statistically significant ($t = 1.56$), the coefficient is higher for the inadequate interviews than it is for the adequate inter-

views. The other results shown in Table 12 are not the same. The writer believes that the following reasoning may explain this finding.

(i) The criterion value (page 58) was obtained by assessments and systems of showing preference for one employee over another used by the supervisors with whom the 123 employee/candidates worked daily. It is accepted, therefore, that both as individuals (personality make-up) and as workers (ability) they were well known to the supervisors.

(ii) It is suggested furthermore that an individual's approach to a work situation is similar, basically, to his approach to a test situation. In other words, if an individual's approach is poor and negative in the one case it will very probably be similarly poor and/or negative in the other case. This being so, it should not be difficult to assess such a subject's merits and (a) the interviewee will make a low assessment in respect of the subject's performance in the Trial Interview; an inadequate interview can leave but little doubt in the assessor's mind as to the candidate's ability as revealed in that situation (Fig. 35).

(b) the supervisor will also have made a low assessment in respect of the subject's work ability for the purposes of the criterion value (Fig. 36).

(iii) If a candidate's approach either to his work or to the test situation is a positive and considered one (adequate interviews), it is possible to find larger differences of judgement, interpretation and appreciation reflected in the assessments of the interviewee and the supervisor.

(iv) In view of (i), (ii) and (iii) it is suggested that the "Ability" assessments of inadequate interviews

will have sufficient comparable content in the criterion values to ensure a relatively high correlation coefficient between the two sets of assessments.

(v) The suggestion made in (iv) will not necessarily apply when "Personality" assessments are made because, during the inadequate interviews, the interviewees have had no more than limited opportunities for making consistent and valid assessments of these attributes - these items must be thoroughly observed to enable consistent and reliable assessments to be made of them. It is therefore to be expected that these assessments will not correlate highly with the criterion which includes the supervisors' considered opinions (accepted in (i) as being reliable) of the employees' personality qualities.

While considering the above reasoning in respect of inadequate interviews it is important to remember that the assessments of each of the individual items represents the extent to which the particular attribute was manifested during the Trial Interview situation. As a result of an inadequate interview (the judges used a rigid interpretation - see page 48) it is practically impossible to consider that an interviewer has earned more than a very low assessment of the listed "Ability" items. However, these assessments are considered to be reliable; the most inexperienced assessor who makes comments such as "interview lasted one minute" ; "did not ask one question" ; etc., will be able to rate reliably such items as perseverance in getting at facts, use of data supplied, effective use of questions, etc.

The "Personality" items, however, are liable to be misjudged more easily (pages 34 and 35). It has been suggested that the assessments made following an interview which lasted just one minute, for example, will be low

and reliable for "Ability" items - but it is submitted that the assessment of self-confidence, ease of establishing social contact, etc., will be less reliable. The point need hardly be laboured that even experienced assessors require adequate observation of a candidate in order to assess personality attributes satisfactorily. It is also evident that the Overall Assessments, designed to indicate potential in the administrative field, can scarcely be expected to have a high validity coefficient when the interviews have been inadequate.

The evidence examined up to this point indicates that the "Ability" Total is the assessment which does not appear to be affected detrimentally by inadequate interviews. An inspection of the r's in Table 11 shows that the "Ability" Total is the assessment which promises to contribute most to the predicting efficiency of the battery. Viewed with this knowledge the increased weight (.110) of the "Ability" Total becomes not only interesting but also meaningful.

(b) The battery used in the selection of the Administrative Trainees consisted of four tests (A(F), M, W.P. and F.R.). The Overall Assessment was not included for the reasons given on pages 27 and 28. The study of the available data has shown that the assessments of individual items, particularly the "Ability" items, appear to be more reliable measures than the Overall Assessments in situations which allow 1 out of 4 of the Trial Interviews to be inadequate.

It was accordingly decided to use the Wherry - Doolittle Test Selection method (Ref. 7 pages 435 - 451) in three stages in an effort to obtain evidence that

(i) the "Ability" Total does serve a purpose in the battery.

(ii) the weight of .110 has some significance.

Stage 1

The five tests from which the first battery is to be selected according to the Wherry - Doolittle method are the four used in the first instance (A(F), M, W.P. and F.R.) together with the Overall Assessment. The method can only be applied to a group of assessments for which criterion values have been obtained. In the present case therefore the investigation is limited to either N = 123, 90 or 33. The largest sample (N = 123) was selected after a study of Table 12 which shows that this is the group for which the validity coefficients are the highest.

TABLE 13

Correlation Coefficients (N = 123)

	1	2	3	4	5
	A(F)	M	W.P.	F.R.	O.A.
C5 + 6	.525	.414	.516	.599	.339
1		.696	.386	.561	.283
2			.308	.459	.306
3				.563	.351
4					.519

As a result of the calculations which follow a ratio $\frac{V_m^2}{Z_m}$ is obtained for each test. The test selected for inclusion in the battery at each successive step (see column a Table 14) is the one having the highest $\frac{V_m^2}{Z_m}$ value. Once the first test is selected the Wherry shrinkage formula corrects the multiple R for chance error. The whole process is then repeated to determine which test is the next to be included. The battery is finally in its most effective form when the multiple R (column g Table 14) ceases to increase with the inclusion of additional tests.

TABLE 14

a	b	c	d	e	f	g	h
	$\frac{V_m^2}{Z_m}$	K^2	$\frac{N-1}{N-m}$	$\overline{K^2}$	$\overline{R^2}$	\overline{R}	Test
0		1.000	N=123				
1	.3588	.6412	1.000	.6412	.3588	.5995	4
2	.0521	.5891	1.008	.5938	.4062	.6373	1
3	.0379	.5512	1.017	.5606	.4394	.6629	3
4	.0006	.5506	1.025	.5644	.4356	.6600	2

An examination of these results allows the following observations to be made:

- (a) The most efficient battery consists of tests 4, 1 and 3.
- (b) The tests contribute to the predicting efficiency of the battery in the order in which they were selected, viz., firstly the Final Rating, secondly Test A(F) and thirdly the Written Project.
- (c) Test M contributes little or nothing to the battery's validity - the multiple R shows a small decrease.
- (d) The Overall Assessment should not be included in the battery at all. In fact, were it to be included as a fifth test, the multiple R would be reduced still further.

Stage 2

In the following calculations the "Personality" Total assessment replaces the Overall Assessment as the fifth possible test in the battery.

TABLE 15
Correlation Coefficients (N = 123)

	1 A(F)	2 M	3 W.P.	4 F.R.	5 O.A.
C5 + 6	.525	.414	.516	.599	.270
1		.696	.386	.561	.287
2			.308	.459	.345
3				.563	.255
4					.507

The final table, with the correction for chance error is as follows:

TABLE 16

a	b	c	d	e	f	g	h
	$\frac{V_m^2}{Z_m}$	$\frac{K^2}{N-m}$	$\frac{N-1}{N-m}$	$\frac{\bar{K}^2}{K^2}$	$\frac{\bar{R}^2}{R^2}$	$\frac{-}{R}$	Test
0		1.000	N=123				
1	.3588	.6412	1.000	.6412	.3588	.5995	4
2	.0521	.5891	1.008	.5938	.4062	.6373	1
3	.0379	.5512	1.017	.5606	.4394	.6629	3
4	.0011	.5501	1.025	.5638	.4362	.6605	5

It is noted that -

- (a) The battery consists of test 4, 1 and 3 selected in the same order found in Stage 1.
- (b) The fourth test is the "Personality" Total which, owing to the preferential selection it enjoys over Test M, may be placed higher in order of merit (in terms of selection for the battery) than the Overall Assessment. Test M in turn was available for selection in preference to the Overall Assessment (Stage 1).
- (c) Though the "Personality" Total is therefore a more suitable measure than either Test M or the Overall

Assessment, its inclusion in the battery would result in a reduction of the Multiple R.

Stage 3.

This time the fifth possible test made available for inclusion in the battery is the "Ability" Total.

TABLE 17

Correlation Coefficients (N = 123)

	1 A(F)	2 M	3 W.P.	4 F.R.	5 "A" T.
C5 + 6	.525	.414	.516	.599	.378
1		.696	.386	.561	.284
2			.308	.459	.291
3				.563	.287
4					.474

TABLE 18

The table including the correction for chance error is as follows :-

a	b	c	d	e	f	g	h
	$\frac{V_m^2}{Z_m}$	K^2	$\frac{N-1}{N-m}$	$\overline{K^2}$	$\overline{R^2}$	\overline{R}	Test
0		1.000	N=123				
1	.3588	.6412	1.000	.6412	.3588	.5995	4
2	.0521	.5891	1.008	.5938	.4062	.6373	1
3	.0379	.5512	1.017	.5606	.4394	.6629	3
4	.0093	.5419	1.025	.5554	.4446	.6668	5
5	.0002	.5417	1.034	.5601	.4399	.6633	2

- (a) The results indicate clearly that the "Ability" Total is selected as a fourth test in the battery.
- (b) The first three tests are selected in the undisturbed order found in Stages 1 and 2.

- (c) The inclusion of the fourth test raises the multiple R from .66 to .67. Though the increase is not considerable, it is the tendency which is important in respect of the objectives outlined on pages 67:68. The values tabled in column b are measures of the amount which the respective tests contribute to the squared multiple correlation coefficient. Though not particularly large, the contribution of the "Ability" Total is definitely greater than that of the Overall Assessment, Test M or the "Personality" Total.
- (d) As in the first two Stages, Test M's inclusion in the battery results in a reduced multiple R. However, it is of interest to note that this reduced multiple R is nevertheless numerically higher than it was after the addition of Test 3 (W.P.) to the battery (step 3 in column a).

Following on the selection of tests by the Wherry - Doolittle method, the calculation of the weights of the battery's tests is straightforward. This was done for the recommended new battery as follows:-

TABLE 19

	4	1	3	5	-C
C ₁	-1.000	-.561	-.563	-.474	.599
C ₂		-1.000	-.102	-.026	.276
C ₃			-1.000	-.027	.237
C ₄				-1.000	.110

The C values are obtained while calculating the selection of tests above. Each line is equated as below:-

$$-1.000\beta_4 - .561\beta_1 - .563\beta_3 - .474\beta_5 + .599 = 0 \quad (1)$$

$$-1.000\beta_1 - .102\beta_3 - .026\beta_5 + .276 = 0 \quad (2)$$

$$-1.000\beta_3 - .027\beta_5 + .237 = 0 \quad (3)$$

$$-1.000\beta_5 + .110 = 0 \quad (4)$$

The weights of the respective tests in the battery are obtained by solving these equations.

$$\begin{aligned}\beta_4 &= .275 \\ \beta_1 &= .249 \\ \beta_3 &= .234 \\ \beta_5 &= .110\end{aligned}$$

At this juncture it is interesting to study a comparative table showing the two sets of weights obtained by different methods.

TABLE 20

	Weights calculated by the Wherry-Doolittle method.	Weights calculated by the Pivotal Condensation method.
Final Rating	.275	.277
Test A(F)	.249	.249
Written Project	.234	.233
"Ability" Total	.110	.108

Both these methods of calculating weights are sound and the similarity of the results is therefore not unnatural. However, the following observations are relevant. The Pivotal Condensation method is simply a solution of equations; no assumptions are involved. On the other hand the Wherry - Doolittle method includes the application of a shrinkage formula on the grounds that chance errors related to the selection of a sample from the whole population have to be corrected. The close agreement of the results therefore gains interest.

The Test Selection method has singled out the "Ability" Total for inclusion in the battery (besides the original three tests) from the four tests "Ability" Total, "Personality" Total, Overall Assessment and Test M. A further indication of the significance of this can be obtained as follows:-

$$\begin{aligned}
 R^2 &= \beta_1 r_1 + \beta_3 r_3 + \beta_4 r_4 + \beta_5 r_5 \\
 &= (.249)(.525) + (.234)(.516) + (.275)(.599) + (.110)(.378) \\
 &= .1307 + .1207 + .1647 + .0415 \\
 &= .4576 \\
 R &= .6765
 \end{aligned}$$

The correction for chance errors reduces this multiple R to .6663

The above formula, besides enabling the multiple R to be obtained, indicates (Ref. pp. 425 - 426) that 46% represents "the proportion of the variance of the criterion measure attributable to the joint action of the variables" i.e., the battery's four tests. Moreover each test's quantitative contribution is seen to be

- F.R. — 16%
- A(F) — 13%
- W.P. — 12%
- "A".T — 4%

Comparatively, the contribution of "Ability" Total is small but it is definite nevertheless; the relative significance of the various test's contributions is important.

In order to complete this stage of the review, the weights and percentage contributions of the four tests originally used in the battery were calculated.

TABLE 21

	Recommended New Battery		Old Battery	
	Weight	Percentage Contribution	Weight	Percentage Contribution
Final Rating	.275	16%	.322	19%
Test A(F)	.249	13%	.231	12%
Written Project	.234	12%	.236	12%
"Ability" Total	.110	4%	-	-
Test M	-	-	.032	0%
Multiple R	.6663		.6600	

The problem has been to obtain, if possible, evidence of the significance of the "Ability" Total's weight (.110) in order to justify its inclusion as a fourth test in the battery. The considerations presented on pages 62 - 74 have served this end; moreover the comparative statistical claims of the "Ability" Total and Test M as seen in the above table are adequate justification for the inclusion of the former in preference to the latter.

However, one final word of caution must be sounded. It may appear that the recommended battery consists of tests to measure general mental ability, generally high potential with emphasis on personality attributes as well as general ability qualities (the new test is "Ability" Total). Initially the interviewees were instructed to assess two sets of individual items which have been called "Personality" Total and "Ability" Total. (Refer to page 14). As a result of the findings of this study it becomes pertinent to consider whether the ratings assigned to the "Ability" items are not in actual fact more truly representative of an overall estimate incorporating evidence from the "Personality", "Ability" and, perhaps, additional fields. The "Ability" Total may not be a highly reliable measure of ability as such; in other words, it would probably be more correct to designate the value obtained from the assessment of "Ability" items as an Overall Assessment.

C H A P T E R V

A. SUMMARISED REVIEW OF THE STUDY OF THE TRIAL
INTERVIEW'S ASSESSMENTS.

A. The frequency distributions of the results obtained for all the tests in the battery produced the expected unimodal distributions with the striking exception of the Trial Interview's Overall Assessment which featured two distinct modes. No immediate explanation was available and the assessments were accordingly treated in an effort to establish the causal factor(s); the aspects studied initially were:-

- (1) varying standards used by the assessors.
- (2) non-homogeneity of the group in respect of
 - (a) the two well represented language groups of the Union.
 - (b) the two distinct sources from which the candidates were recruited, viz., university graduates (or the equivalent) and employees of the sponsor organisation for whom advanced educational qualifications were not enforced.
- (3) The method of administering the testing procedures.

None of these approaches reduced the bimodal characteristics of the distributions; the explanation was therefore not to be found in the non-homogeneous qualities of the group. Subjective influences such as the formation of liking or disliking or sex bias could not be dismissed summarily at that stage but were examined later.

B. The frequency distributions of the individual "Personality" and "Ability" items (assessments obtained during the Trial Interview) were found to be unimodal thus

underlining the unusual nature of the bimodality of the Overall Assessment's distribution (assessments also obtained during the Trial Interview). The addition of the individual items, however, produced distributions of "Personality" Total and "Ability" Total which were multimodal. The implications of this were studied more closely.

The rating of individual items was directed and specific in contrast with the wide, comprehensive Overall Assessment. Could it have been that the nature of the latter allowed the intrusion of subjective influences? This possibility, considered tentatively, led to an examination of the effect on the multimodal distributions of "Personality" Total and "Ability" Total of condensing the rating scale to a point where it consisted of nine class intervals (equal to that used for the Overall Assessment). The result was that the "Personality" Total distributions lost bimodality more readily than the "Ability" Total distributions. In view of the suggested influence of subjectivity this was unexpected because of the greater degree of consistency which the assessments of "Ability" items enjoyed.

The influence of subjectivity therefore did not make itself strikingly evident and a still closer examination of the data was carried out.

The inter-item correlation coefficients were found to be high. A consideration of the available material led to the conclusion that the addition of values assigned to highly correlated items produced frequency distributions whose multimodality was revealed in relation to the extent of the correlation. It was submitted that this provided the main explanation of bimodality in the frequency distributions of "Personality" Total and "Ability" Total. At the same time insufficient evidence had been obtained to

allow the possible influences of subjectivity to be discounted conclusively.

At this stage of the investigations the following statements could be made.

- (1) The frequency distributions of individual "Personality" and "Ability" items were unimodal.
- (2) The several modes of the distributions of "Personality" Total and "Ability" Total were explicable in terms of the high correlations between the individual items.
- (3) No acceptable regrouping of the Overall Assessments had in any way reduced the unexpected bimodality of the frequency distributions of these assessments.

C. A comparative study of the Overall Assessment and the "Personality" Total, "Ability" Total and "Personality + Ability" Total showed very high agreements (r ranged from .837 to .944). This was not entirely unexpected but the correlation coefficient of .944 between the Overall Assessment and the "Personality Ability" Total left little room for improvement and warranted careful attention. The frequency distributions of these two assessments were found to be very similar except in the portion of the rating scale representing the lower values. Now the assessments of "Personality + Ability" Total had been demonstrated to be satisfactorily consistent - no evidence of the assessors' inconsistency had emerged. It seemed apparent, therefore, that the process of rating the Overall Assessments had been sufficiently hampered by weakness in the test or test situation to produce the observed bimodality.

D. The Overall Assessment, on account of the broad, all embracing view which it necessitated, was considered to be the point most susceptible to the detrimental influence

of inconsistency. The intrusion of subjective factors or bias was facilitated owing to the lack of specificity of an Overall Assessment. Consequently the interviewees' reports were studied closely in an effort to obtain evidence likely to support this contention.

E. It was found that approximately 25% of the interviews had been either so short or so one-sided (the interviewer reading from the job description supplied instead of interviewing) that the performances were absolutely unsatisfactory in terms of the test's requirements; in other words, the test had not been done. These inadequate interviews were extracted from the total group by three independent judges.

The frequency distribution of the Overall Assessments of the adequate interviews was found to be practically unimodal. The frequency distribution of the inadequate interviews' Overall Assessments covered most of the rating scale (B+ to C-) but it was strikingly skewed, the class interval with the highest frequency corresponding to the class interval in which the second mode of the Overall Assessment's distribution (N = 219) was situated.

These findings explained satisfactorily why it was principally the Overall Assessments which were affected detrimentally. An inadequate interview, though curtailing the information necessary to make an Overall Assessment, nevertheless enabled the observation of individual items (self-confidence, effective use of data supplied, etc.) to be reasonably accurate and reliable; the frequency distributions of the individual "Personality" and "Ability" items were found to be unimodal.

An effort was then made to determine the extent to which adequacy/inadequacy introduced variations in the correlations between the tests of the battery. The Over-

all Assessment, "Ability" Total and "Personality" Total were each correlated with the four tests of the battery in respect of the two groups of adequate and inadequate interviews. Each set of coefficients was tested for the significance of the difference. The Overall Assessment's correlations with the Final Rating were the only ones for which a statistically significant difference for adequate and inadequate interviews was found.

The detrimental effect of an inadequate interview was therefore detectable statistically only in Overall Assessments which have taken special account of personality attributes. This is of great interest because the Final Rating and the Overall Assessment were designed to be comparable assessments. These facts confirmed both the finding that inadequate interviews contributed to the production of bimodality and the view that the Overall Assessments should be treated with suspicion.

An examination of the correlation tables of the battery's four tests with Overall Assessments, "Ability" Total, "Personality" Total and various sets of "Personality" and "Ability" items showed conclusively that, from the assessments examined, the "Ability" Total best satisfied the demands made of a test to be included in a battery, i.e., high correlation with the criterion and low correlations with the remaining tests.

F. The next stage consisted of studying the weights of each test in relation to the predicting efficiency of the battery as a whole. It was found that the weighted contribution of the Overall Assessment was negligible numerically. This was also the case for all other possible sets of assessments considered as the battery's fifth test with the notable exception of the "Ability" Total. The weights for the various tests were ± 0 ; in the case of the

"Ability" Total it was .110. Was this increase in weight significant?

A study of the correlation Tables 9 and 11 showed that, irrespective of the exclusion of inadequate interviews, the "Ability" Total had constantly the highest correlation with the criterion and the lowest with the original tests in the battery. This indicated that the "Ability" Total was the assessment which would contribute most to the predicting efficiency of the battery. The Wherry - Doolittle method of selecting tests for a battery was then applied and it was found that the "Ability" Total was selected in preference to either the Overall Assessment, Test M or "Personality" Total. It was concluded that, under the circumstances in which the test was administered, the "Ability" Total was less affected adversely by the inadequacy of interviews and was a more reliable and valid assessment than the Overall Assessment.

G. As a result of the above investigation a new situation has been devised which endeavours to remove the weaknesses responsible for the unreliability of the Overall Assessments. These are briefly (1) a considerable transfer of the control of the interview from the candidate to the assessor.

(2) the presence of more than one assessor during the administration of the test.

The new test situation will be discussed in the next chapter.

C H A P T E R VI

SUGGESTED IMPROVEMENTS IN THE TEST SITUATION.

The analysis and study of the assessments made in the Trial Interview situation has focussed attention on the main weaknesses of this test. These are:-

(1) The desirability of making the artificial situation resemble a true and natural interview as closely as possible necessarily limits the number of trained assessors present to one. There is therefore no method available to reduce the frequency and severity of faulty evaluations as is the case when the assessments of the raters in the group discussions are averaged to obtain the Final Rating. After the Leaderless and Assigned Leadership Group discussions a review of the events during the situations may bring to the notice of a tester an incident or characteristic which he has overlooked or failed to appreciate; this additional information may influence the tester's estimate sufficiently to warrant a revision of the assessment. In such a case the undesirable effect which a faulty assessment would have had on the Final Rating is reduced. Though assessments once made are seldom amended (the incidence of widely varying assessments for one candidate is low) the advantage of using one score obtained from several independent ratings is not available in the Trial Interview.

(2) The nature of the interview places the interviewee (assessor) in a position from which it is not possible to exercise more than indirect control on the course which the interview follows. This resulted in practically 25% of the interviews being classified as in-

adequate performances in the given situation. It is therefore reasonable to conclude that, had the assessors' control on the interviews been more effective, the number of inadequate interviews would certainly have been reduced; hence the number of unreliable Overall Assessments would have been smaller and the bimodality of the frequency distribution of these assessments would have been less striking (at least).

A new situation has therefore been devised in an endeavour to remove the above two main sources of weakness. The candidate is prepared as follows:-

"You are the accountant of a large industrial concern. Two female employees on your staff have made an appointment with you in order to obtain your ruling in respect of a matter of procedure about which they have had a difference of opinion. The employees are engaged on the same assignment in connection with the execution of which each has adopted a definite system. Employee A commends method 1 and employee B commends method 2 (details of each method are supplied in Annexure E). You are to study the details and implications of each method and decide which is the more suitable for use in your department. Having done so, interview the employees explaining your views and attempt to reconcile the opinion of the employee whose method you have discarded with your own."

The situation as envisaged will be revealing. The method which appears to be the more desirable one has, in actual fact, disadvantages which are not readily apparent; consequently the "Accountant's" selection of a method will indicate understanding and insight if he supports his choice with the arguments necessary to underline the weaknesses of the discarded method. A study of the procedure

which he uses as well as his manner when in a position of authority (as in the Trial Interview) will be of interest. He will have to approve of one of two methods each commended by an independent "employee" prepared to go to the head of the department to obtain support for her contentions. The process of expressing his thoughts will be embarrassed owing to the opposing views of the "employee" whose method is rejected.

The "accountant's" position could become extremely unenviable in the presence of two particularly determined "employees". To obviate such a deadlock and to enable the "accountant" to make headway during the interview, the "employees" are instructed as follows. At the outset they must be equally convinced of the merits of their respective methods. As the interview progresses the "accountant" signifies his acceptance of one of the methods. Such an indication must not weaken the conviction of the supporter of the rejected method. The weakening process has to follow three definite stages.

(a) One sound, considered reason for discarding a method (and adopting the other method by argument or implication) must not result in any noticeable change of opinion.

(b) A second sound reason results in evident wavering though by itself does not produce the desired effect.

(c) A third sound reason is followed by the growing uncertainty of the "employee" concerned and, eventually, a recognition of the advantages of the recommended method.

An interview consisting of a series of unsound and inconclusive arguments will, after the lapse of sufficient time

for the purpose of assessment, be terminated by the "employees" who may either accept the views put forward such as they are or ask for additional time to ponder over the whole problem in the light of the discussion just concluded. On the other hand, should an interviewer lead the proceedings swiftly through the above three stages, the employees can prolong the interview in order to allow thorough observations to be made.

The advantages of the new situation are as follows:-

(a) Two assessors are present and two sets of assessments will be available.

(b) The situation makes it possible for the assessors to keep the discussion alive (if it shows signs of doing otherwise) until such time as the candidate has been adequately observed and all doubts in the assessors' minds have been dispelled.

These two benefits alone are very important as it is precisely the failure to operate these features in the Trial Interview which led to the lack of reliability of the Overall Assessments.

Additional innovations are also incorporated. In the Trial Interview the interviewer was faced with one main question - acceptance or rejection of Miss van Jaarsveld's application. The new situation resembles this in so far as the "accountant" must select method 1 or method 2. But over and above this principal decision his evaluation and understanding of human material and human relationships will have to be more refined, discriminating and subtle. A difference of opinion has to be resolved - to succeed in the situation as constructed requires an effort of a high standard. The nature of the role of the "employees" allows them to lower this high standard for a candidate who is obviously failing to cope in order that he may attain some

semblance of success; information is thus obtained which would not otherwise have been available. The wider range of performances which such a system ensures will yield more meaningful results.

In view of the findings of the present study certain amendments to the assessors' instructions and the rating methods themselves are recommended.

(i) The assessment of a revised set of individual items will be undertaken. Owing to the improvements in the test situation (reducing the possibility of onesidedness and /or brevity of the interview) these items will be more specific than was the case in the Trial Interview. It is considered that this will reduce the effects of halo and interaction (pages 39 - 40) as well as increase the validity of the ratings.

(ii) Overall Assessments are also to be made but the instructions to the assessors will be to consider all that has actually transpired during the interview for the purposes of forming a Gestalt impression of the "personality" and "ability" qualities. (The Trial Interview's Overall Assessment endeavoured, according to the interviewees' instructions, to indicate the interviewer's potential level in the administrative field with particular reference to the personality qualities revealed.)

The changes incorporated in the new situation are such that the possibility of a candidate failing to allow himself to be observed in an inadequate interview has been reduced appreciably. It is considered that the additional benefits will enable the Overall Assessment to become a statistically acceptable measure. This may be established,

eventually, by the application of the new test to a large group of candidates applying for admission to a similar training scheme and a subsequent study of the new set of assessments.

ANNEXURE A.

Product Moment Correlation Coefficient.

(ii) Presence

(i) Self-confidence

	0	1	2	3	4	5				fd	fd ²
5					1		1	4	} 3.48	5	25
4			2	13	11	2	28	97		112	448
3		1	27	24	5		57	147	2.58	171	513
2	3	28	38	5			74	119	1.61	148	296
1	13	30	10	1			54	53	} .92	54	54
0	4	1					5	1		0	0
	20	60	77	43	17	2	219			490	1336
	19	89	175	135	64	8		1140			
	.95	1.48	2.27	3.14	3.79						
fd	0	60	154	129	68	10	421				
fd ²	0	60	308	387	272	50	1077				

(i) Self-confidence.

$$m_1 = \frac{490}{219} = 2.23744$$

$$m_2 = \frac{1336}{219} = 6.10046$$

$$S.D. = \sqrt{6.10046 - 5.00614}$$

$$= \sqrt{1.09432}$$

$$= \underline{1.04610}$$

(ii) Presence.

$$m_1 = \frac{421}{219} = 1.92237$$

$$m_2 = \frac{1077}{219} = 4.91781$$

$$S.D. = \sqrt{4.91781 - 3.69551}$$

$$= \sqrt{1.22230}$$

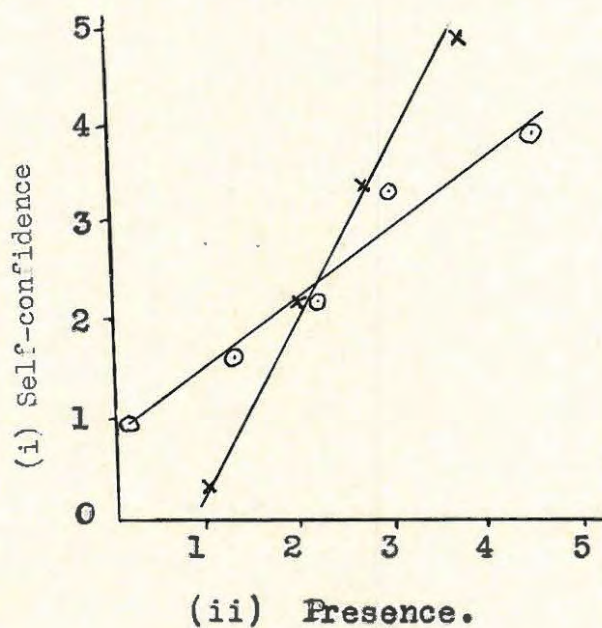
$$= \underline{1.10558}$$

$$\begin{aligned} r &= \frac{1140}{219} - \frac{(2.23744 \times 1.92237)}{1.04610 \times 1.10558} \\ &= \frac{5.20548 - 4.30119}{1.15655} \\ &= \frac{.90429}{1.15655} \\ &= \underline{.78188} \end{aligned}$$

$$\begin{aligned} \sqrt{r} &= \frac{1 - r^2}{\sqrt{N - 1}} \\ &= \frac{1 - .61134}{14.76483} \\ &= \frac{.38866}{14.76483} \\ &= \underline{.02632} \end{aligned}$$

$$\begin{aligned} \text{Critical Ratio} &= \frac{.78188}{.02632} \\ &= \underline{29.707} \end{aligned}$$

Regression Lines.



ANNEXURE B.

Contingency Coefficient.

Judge B.

		<u>Judge B.</u>		
		Adequate	Inadequate	
Judge A.	Adequate	A = N ₁ = 123	B = N ₂ = 11	A + B = 134
	Inadequate	C = N ₃ = 25	D = N ₄ = 60	C + D = 85
		A + C = 148	B + D = 71	N = 219

$$\begin{aligned}
 X^2 &= \frac{N(AD - BC)^2}{(A+B)(C+D)(A+C)(B+D)} \\
 &= \frac{219(7380 - 275)^2}{134 \times 85 \times 148 \times 71} \\
 &= \frac{219(7105)^2}{11390 \times 10508} \\
 &= \frac{11,055,344,475}{119,686,120} \\
 &= 92.36956
 \end{aligned}$$

Therefore P is less than .001

ANNEXURE C.

The significance of the difference of the correlation coefficients between Overall Assessments and Final Ratings with N=166 and N=53.

$$\begin{aligned} z_1^1 &= \frac{1}{2} \log e \frac{1+r_1}{1-r_1} \\ &= \frac{1}{2} \log e \left(\frac{1.49129}{.50871} \right) \\ &= \frac{1}{2} \log e (2.93151) \\ &= .5 \left(\frac{.4670914}{.43429448} \right) \\ &= \underline{.53776} \end{aligned}$$

$$\begin{aligned} z_2^1 &= \frac{1}{2} \log e \frac{1+r_2}{1-r_2} \\ &= \frac{1}{2} \log e \left(\frac{1.04556}{.95444} \right) \\ &= \frac{1}{2} \log e (1.09547) \\ &= .5 \left(\frac{.0396005}{.43429448} \right) \\ &= \underline{.04559} \end{aligned}$$

$$\begin{aligned} X &= \frac{z_1^1 - z_2^1}{\sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}} \\ &= \frac{.53776 - .04559}{.16164} \\ &= \frac{.49217}{.16164} \\ &= \underline{3.045} \end{aligned}$$

	Z1 A(F)	Z2 M	Z3 W.P.	Z4 F.R.					
A(F)	1.0000	.6962	.3857	.5606	-1.0000				1.6425
M	.6962	1.0000	.3080	.4593		-1.0000			1.4635
W.P.	.3857	.3080	1.0000	.5629			-1.0000		1.2566
F.R.	.5606	.4593	.5629	1.0000				-1.0000	1.5828
C5 + 6	.5254	.4140	.5163	.5995					2.0552
		(.51530)	.03948	.06901	.69620	-1.00000			.31999
		1.00000	.07661	.13392	1.35103	-1.94058			.62098
		.03948	.85124	.34668	.38570		-1.00000		.62309
		.06901	.34668	.68573	.56060			-1.00000	.66202
		.04822	.31365	.30496	.52540				1.19223
			(.84822)	.34139	.33236	.07661	-1.00000		.59857
			1.00000	.40248	.39183	.09032	-1.17894		.70568
			.34139	.67649	.46737	.13392			.61917
			.30996	.29850	.46025	.09357			1.16229
				(.53909)	.33360	.10309	.40248		.37826
				1.00000	.61882	.19123	.74659	-1.85498	.70166
				.17375	.33880	.06557	.36542		.94356
					.23128	.05234	.23570	.32230	.82165

$$R^2 = .44981$$

$$R = \underline{.67068}$$

ANNEXURE D
PIVOTAL CONDENSATION.

ANNEXURE E

I. Information supplied to both candidates and testers.

A. A manufacturing concern in the Western Province employs Cape Coloured cutters and machinists. These workers form a comparatively stable labour force, due in part, to the concern in their welfare which the management has shown. Some of the features which are known to be a source of satisfaction are:-

- (i) Canteen facilities.
- (ii) Weekly pay (each Friday).
- (iii) The issue of overalls for use in the factory.

B. The organisation's Housing Company is responsible for collecting the weekly rental from the tenants as well as keeping an up-to-date record of the accounts. This is done by employees A and B each of whom keeps the receipt books and accounts for the tenants in her collecting zone. The different accounting systems used by each employee are as follows.

Notes on Method 1 (used by employee A).

1. A receipt book is kept.
2. Rents are collected on Fridays.
3. The Arrears and Advance columns (Page 1 below) reflect the position on 1.9.51.
4. The pencilled entries in the Advance column on the right hand side of the page are designed to show at any time the advances for the last two weeks. Each week, therefore, the date and entries in one column are erased to make way for the new and more recent entries.
5. The "Arrear Schedule" is a permanent record.
6. The two summaries as compiled enable the accounts to be balanced each week.
7. Cash payments, Unlets and Rentals are carried forward each week in the summaries.

Method 1. (Used by employee A)

PAGE 1.

Address	Tenant	Weekly Rent	Arrears	Advance	September, 1951				Advance	
					7	14	21	28	28	21
Firgrove Avenue										
4	Retief	10. 2.			10. 2.	15. 2.	10.2.	10.2.	5/-	5/-
6	Segal	10. 2.	10.2.		1. 0. 4.	4. 2.	10.2.	10.2.		
8	Rosen	7. 6.		10/-	5. 0.		7.6.	7. 6.		
10	Smit	7. 6.			7. 6.	5. 0.	7.6.	7. 6.		
Paul Kruger Street										
1	McIntyre	7. 6.	7.6.			10. 0.	10.0.	10. 0.		
3	Forsberg	10. 6.		10. 6.	10. 6.		10.6.	12. 6.	2/-	
5	Esprey	12. 6.		12. 6.		12. 6.	12.6.	12. 6.		
7	Williams	12. 6.			12. 6.	15. 0.	12.6.	10. 0.		2/6
T O T A L		£3.18. 4.	17.8.	£1.13. 0.	£3. 6. 0.	£3.1. 0.	£4.0.10.	£4. 0.4.	7/-	7/6

Page 2.

Arrear Schedule

Address	September, 1951			
	7	14	21	28
Firgrove Ave.				
4				
6		6/-	6/-	6/-
8				
10		2/6.	2/6.	2/6.
Paul Kruger Street				
1	15/-	12/6	10/-	7/6
3				
5				
7				
Total	15/-	£1.1.0.	18/6	16/-

Summaries

	September, 1951			
	7	14	21	28
Arrears	15. 0.	£1. 1. 0.	18. 6.	16.0.
Cash payments	£3. 6. 0	£6. 7.10.	£10. 8. 8.	£14. 19.0.
Unlets				
Advances B/F	£1.13. 0.	£1.13. 0.	£1.13. 0.	£1. 13.0.
Total	£5.14. 0.	£9. 1.10.	£13. 0. 2.	£16. 18.0.
Arrears B/F	17. 8.	17. 8.	17. 8.	17.8.
Rental	£3.18. 4.	£7.16. 8.	£11.15. 0.	£15. 13.4.
Advance	18. 0.	7. 6.	7. 6.	7.0.
Total	£5.14. 0.	£9. 1.10.	£13. 0. 2.	£16. 18.0.

Method 2. (Used by employee B)

Address	General Remarks	Tenant	Weekly Rent	Balance 31.8.51	1	8	15	22	29	Rent Charge for Sept, 51.	Balance 30.9.51.
Firgrove Avenue											
3		de Bruin	10.2.		10. 2.	10. 2.	5. 2.	10. 2.	10.2.	£2.10.10.	5. 0.
5		Venter	10.2.	10.2.	10. 2.	10. 2.		10. 2.	10.2.	£2.10.10.	£1.0. 4.
7		Goosen	7.6.	10.0.	5. 0.	10. 0.	7. 6.	7. 6.		£1.17. 6.	2. 6.
9		Fourie	7.6.		7. 6.	7. 6.	7. 6.	5. 0.	7.6.	£1.17. 6.	2. 6.
Paul Kruger Street											
2		Smith	7.6.	7.6.	5. 0.	7. 6.	7. 6.	7. 6.	5.0.	£1.17. 6.	12. 6.
4		Brown	10.6.		10. 6.	10. 6.	10. 6.	7. 6.	10.6.	£2.12. 6.	3. 0.
6		Jones	12.6.		10. 0.	10. 0.	12. 6.	12. 6.	10.0.	£3. 2. 6.	7. 6.
8		Whittal	12.6.	12.6.	12. 6.	10. 0.	12.6.	10. 0.	12.6.	£3. 2. 6.	17. 6.
			£3.18.4.	£1. 0.2.	£3.10.10.	£3.15.10.	£3. 2. 3.	£3.10. 4.	£3.5.10.	19.11. 8.	£3.5.10.

Notes on Method 2.

1. A receipt book is kept.
2. Rents are collected on Saturdays.
3. Advance payments are entered in red.
4. The position as regards Advances and Arrears is shown monthly. These balances are a permanent record.
5. The accounts are balanced each month as follows:

Balance at 31.8.51	£1. 0. 2.	Balance at 30.9.51	£3. 5. 10.
Rent charge for September, 1951	£19.11. 8.	Rents collected 1 to 29 September	£17. 6. 0.
	<hr/>		<hr/>
	£20.11.10.		£20.11. 10.
	<hr/>		<hr/>

II. Information supplied to testers only.

Method 1

A. Advantages

1. The position of Advances or Arrears for each tenant is immediately evident.
2. A permanent record of arrears is maintained - it is important to know which tenants are bad payers.
3. The accounts are balanced each week.

B. Disadvantages.

1. It involves slightly more work each week than method 2.
2. Two schedules have to be kept.
2. No column for General Remarks is provided, though the need for this is partly discounted by A 1 and 2 above.
3. The total monthly rent due for each tenant is not available without computation as the system is maintained on a weekly basis.

C. General

Employee A considers that the two reasons principally responsible for a fairly steady and regular payment of

rents are

1. Tenants believe that it is a blot on their record to have an entry on "Page Two", i.e., the Arrear Schedule.
2. Rents are collected on pay day.

Method 2.

A. Advantages.

1. Slightly less work is involved each week than in method 1.
2. Each tenant's monthly rental due is recorded.
3. A column for General Remarks is provided.
4. The system is brief and concise.

B. Disadvantages.

1. If the current accounts are to be balanced at any intermediate stage during the month, special computations have to be made.
2. It is necessary to study each week's entries during the course of the month in order to establish which tenants, if any, are in arrears.

C. General.

Employee B supports method 2 rather than method 1 because she believes that

1. Maintaining the records up-to-date and accurate is easy and involves a minimum of work.
2. The position of Arrears and Advances, though not reflected on a weekly basis, is shown as a permanent record at the end of each month. She considers that this is quite adequate as any additional comments are noted in the special column provided for the purpose.
3. The large number of cases of tenants failing to pay the full week's rent during September, 1951 is explained by Employee B as "an unusual occurrence."

III. Information given by the Employees to the Accountant at the beginning of the interview.

One of Employee B's tenants has complained to her that, whereas he has paid rent five times during the course of the month, his neighbour across the street has only paid rent four times. This has resulted in an investigation of the methods used by the two rent collectors. As a result neither of them is prepared to change her system and, in view of the tenant's complaint, they have come to the Accountant for a decision.

IV. The Accountant's Action.

The following aspects of the situation placed in true perspective are the ones which should guide the accountant to the correct decision.

The listed advantages and disadvantages are all comparatively minor and do not on their own provide a completely adequate basis for accepting or rejecting either of the methods. However, a study of the entries in the two given examples reveals the following significant facts.

1. Employee A's record shows that during the one month under review she has only two cases of tenants either increasing or incurring entries in the arrear schedule.
2. Employee B's record shows 11 similar entries.
(There is no evidence to support her view expressed above in method 2 Section C 3).

The reason for this divergence is not to be found in either method used ^{except} in so far as Section C 1 method 1 contributes to this end. The principal reason for the difference which is offered is that Employee A collects rents on pay day (2 cases of arrears) whereas Employee B collects rents one day after pay day (11 cases of arrears).

In view of this evidence as well as the necessity for uniformity, the Accountant is expected to take the following action:-

- (a) instruct Employee B to collect rentals on Fridays instead of Saturdays.
- (b) recommend the use of method 1.

BIBLIOGRAPHY.

1. A.G.Arbus, Technical Report on Research Project
H/49/171 (1950).
 2. S.Biesheuvel, "N.I.P.R. Bulletin" Vol.1 No.4,
(Oct.1949).
 3. Tedgold, "A Textbook of Mental Deficiency." (1947).
 4. S.Biesheuvel, "The diagnosis of temperament."
(Aptitude Test Section, S.A.A.F.,
unpublished.)
 5. "The Journal of General Psychology." Vol.40
(January 1949.)
 6. J.S.Maritz, "Psychological Bulletin." Vol.47,
No.5 (1950).
 7. H.E.Garrett, "Statistics in Psychology and
Education." (1949).
 8. Q.McNemar, "Psychological Statistics."
 9. S.Biesheuvel, "The Reliability and Validity of
Temperament and Personality
Assessments." (based on research
material available as at May 1945.)
-

