A TEST OF SAWYER'S MODEL FOR THE COMPARISON OF PREDICTIVE METHODS FOR FIVE MANAGERIAL DIMENSIONS IN AN ASSESSMENT CENTRE

by Lyle D. Pederson

Thesis presented to the School of Graduate Studies of the University of Ottawa as partial fulfilment of the requirements for the degree of Master of Arts

Ottawa, Canada, 1974

ACKNOWLEDGEMENTS

This thesis was prepared under the supervision of Professor Charles E. McInnis, Ph.D., Assistant Professor in the Department of General-Experimental Psychology. The writer gratefully acknowledges his advice, direction, and interest in the research.

The writer is deeply indebted to Mr. L. W. Slivinski, Chief of the Managerial Assessment and Research Division, Canadian Public Service Commission, Ottawa, and wishes to thank him for his advice and assistance, in particular for making available the data and the computer facilities of the Public Service Commission for the research. The co-operation of his staff, in particular Mr. K. Grant and Dr. R. Bourgeois, was appreciated for providing additional advice and assistance.

Special gratitude is also extended to Mr. A. Finkelstein for his assistance in preparing computer programs and assisting in the preparation of the data for the computer.

CURRICULUM STUDIORUM

Lyle D. Pederson was born February 14, 1944, in Camrose, Alberta. He received two years of Philosophy and four years of Theology from Newman Theological College, Edmonton, Alberta, from 1962 to 1968. In 1971 he received a Bachelor of Arts degree with a concentration in Psychology from the University of Ottawa.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

Table page

Table page

Table page

Table page

Table page

LIST OF FIGURES

INTRODUCTION

Many investigations have shown that statistical methods of assessment are superior to clinical methods, especially in predicting managerial success. Many of these investigations, however, compare methods of prediction without cross-validating the results or using any statistical test for comparing the methods.

In an attempt to further clarify the clinical versus statistical prediction controversy, researchers have developed various models to classify the predictor methods. This has led to more refined comparisons, but few of the investigations have employed sound experimental designs.

The Assessment Centre approach has become a very popular means in recent years for assessing managerial potential. There are many investigations reported employing this approach, but few have attempted to compare the many methods of prediction which are used to assess the candidates.

The purpose of this research, therefore, is primarily to investigate various methods of predicting managerial performance using Sawyer's model as the basis for determining the types of predictive methods. A unique feature of the design for this investigation is that the same sample of subjects is employed for comparing the various methods of assessment. A review of the literature suggests that there is a need for research in this area since there are no

reported investigations using Sawyer's model for comparing a variety of predictive methods on the same sample of subjects in an assessment centre.

The investigator is interested in the general problem of prediction: are there methods of predicting managerial performance for specific assessment dimensions in an assessment centre which are statistically more effective than other methods? This study compares the Pure Clinical, Trait Ratings, Pure Statistical and Mechanical Composite Methods of assessment in Sawyer's model for five managerial assessment dimensions (communication, self-motivation, tolerance for stress, human relations and decision making). The investigator emphasizes that this is only a preliminary study which will hopefully lead to further research using larger samples.

Chapter I will review the models which have been proposed for classifying the methods of prediction and will report on the investigations which have compared methods of assessing managerial potential. The chapter will conclude with the general statement of the null hypotheses for this research. Chapter II will present the design in the context of the Behavioural Validation Model and will conclude with a statement of the specific null hypotheses. Chapter III will present and discuss the results and will conclude with a summary of the findings and suggestions for further research.

CHAPTER I

REVIEW OF THE LITERATURE

This survey of the literature focuses on models which have been proposed for the classification of predictive methods. A review of the literature suggests that if the researcher is concerned with comparing predictive methods he must first of all explore the predictive models used for classifying predictive methods. The importance of this consideration is emphasized since some authors argue that different predictor models give advantages to the predictor methods being compared. In view of this, the author will discuss each of the models proposed in the literature and point out the advantages or disadvantages of each model. The last section of the Review is related to research which compares various methods of managerial prediction in assessment center settings.

1. The Classification of Predictor Methods.

There are three models which have been proposed in the literature for the purpose of comparing predictive methods. The first model is formulated by Meehl and is the source of the clinical versus statistical prediction controversy. This controversy has given rise to two other

predictive models proposed by Sawyer and Holt. In addition to these models, Gough's research is discussed since he employed Meehl's model to investigate the comparison of predictive methods.

A. Meehl's Model

Meehl[1] in his book Clinical Versus Statistical Prediction emphasizes that the methods of prediction should be compared at the point where the data is combined for making the prediction. He distinguishes between two types of data combination (methods): mechanical and non-mechanical. The mechanical or statistical combining method refers to predictions which are arrived at by the application of regression equations or tables to the data. The defining aspect of this method is that no judging, inferring or weighting is done by a skilled clinician. The non-mechanical method refers to combinations of any other sort (judgmental, clinical, impressionistic, or subjective) and is defined by the fact that the judging or weighting is done by a skilled clinician.

Meehl reviewed the research literature and classified studies according to his two methods of combining data for predicting individual performance. Out of sixteen to twenty

---

1 P. Meehl, Clinical Versus Statistical Prediction, Minneapolis, Jones Press, 1963, p. 16-18.

studies investigated, he found that the mechanical (statistical) methods of prediction were equal or superior to the non-mechanical (clinical) methods in all but one case.[2] However, as he suggests, his findings cannot be clearly stated because little is known about the qualifications and skills of the clinicians and many of the studies were not cross-validated. In 1957 Meehl concluded that out of twenty-seven studies he found in the literature none indicate that the clinical method is superior to the statistical method.[3] Meehl extended his investigation in 1965 to include fifty relevant research articles but it is difficult to draw any conclusions since no bibliography is provided.[4]

B. Gough's Review of Predictive Methods

Gough published a lengthy summary of the clinical versus statistical controversy and reviewed several predictive studies.[5] Like Meehl he considers that the defining distinction between clinical and statistical methods is to be found

---

2 Ibid., p. 119.

3 P. Meehl, "When Shall We Use Our Heads Instead of the Formula?" Journal of Counseling Psychology, Vol. 4, No. 4, 1957, p. 272.

4 --------, "Seer over Sign: The First Good Example," Journal of Experimental Research in Personality, Vol. 1, 1965, p. 27-32.

5 H. Gough, "Clinical versus Statistical Prediction in Psychology," in L. Postman (ed.), Psychology in the Making, New York, Knopf, 1963, p. 526-584.

in how the data is combined for making the prediction. He concludes that the statistical methods have surpassed the clinical ones, but the accuracy of both methods is inadequate. He suggests that a thorough investigation of the various methods needs to be conducted since the adequacy of the clinician's forecasting skills has not been fully investigated.

In summary, both Meehl and Gough conducted studies which employed the same model for classifying predictors. Their emphasis is on how the data is combined to make the prediction. However, as will be suggested later in this chapter, this predictor classification schema appears to be inadequate because there is no allowance in this model for classifying how the data is collected. The following summary of Sawyer's model will suggest why a predictor classification model must account not only for how the data is combined, but also for how the data is collected.

C. Sawyer's Model

Sawyer's classification schema includes comparisons between clinical measurement and prediction.[6] Although Meehl made a distinction between psychometric and non-psychometric data, both he and Gough concentrated on the combination of data for the final phase of the predictive process, the actual

---

6 J. Sawyer, "Measurement and Prediction, Clinical and Statistical," Studies in Personnel Psychology, Vol. 2, 1970, p. 74-105; also in Psychological Bulletin, Vol. 66, 1966, p. 178-200.

predictions. In response to this Sawyer suggests that too much attention has been given to the clinician as an integrator of data rather than a measurer of his sensitivity as a defining instrument. Since both aspects affect the predictive process, they cannot be independently evaluated. If this insight is ignored, a confounding variable is introduced and at the same time many crucial procedures, even the clinical interview, are excluded since Meehl states that non-psychometric data can be collected mechanically or clinically.[7] Sawyer suggests that the problem which often arises when methods are being compared is that the comparisons exclude the use of observations which are outside of pre-established categories. In summary, both data collection and data combination must be included in a classification schema for comparing predictive methods.

Sawyer defines mechanical and non-mechanical modes of data collection and data combination as follows:

(a) Modes of Data Collection:

(i) mechanical (statistical): Rules are prespecified so that no clinical judgment is involved (psychometric tests, biographical data, personnel records);

(ii) non-mechanical (clinical): Only clinical judgment is involved (clinical interview or behavioural observations).

There are also intermediate cases, such that a continuum actually exists between the pure clinical and mechanical modes.

---

7 Ibid., p. 78.

(b) Modes of Data Combination:

(i) mechanical (statistical): Predictors are combined mechanically according to rules (statistical equations or scoring systems) developed specifically for predicting specific types of behaviour.
(ii) non-mechanical (clinical): The defining property is that the inferring or weighting is done by the skilled clinician (for example, the clinician decides on the weights appropriate to the test information).

Sawyer's classification schema results in six distinct predictive methods as Table I depicts.

Using this model Sawyer surveyed the literature for research data which was relevant to his classification system. He abstracted the results of forty-five studies and his results are presented in Table II.[8] The data in Table II indicates that data which is combined mechanically is always equal or superior to data which is combined clinically and this is true for both modes of data collection (clinical and mechanical). Data which is combined clinically does not predict as well with data collected by both modes as it does with data which is collected only mechanically. Furthermore, data which is combined clinically but which includes a mechanical prediction does not predict as well as the mechanical composite. However, the possibility still exists that in collecting data the clinical method could have contributed to the predictive process by accounting for various aspects which did not enter the predictive process through mechanical collection.

---

8 Cf. Appendix 1 for Table II.

Table I.

Sawyer's Classification of Predictive Methods.

Mode of Data Collection	Mode of Data Combination: Clinical	Mode of Data Combination: Mechanical
Clinical	1. Pure Clinical	2. Trait Ratings
Mechanical	3. Profile Interpretation	4. Pure Statistical
Both	5. Clinical Composite	6. Mechanical Composite

D. Holt's Model

From yet another point of view, Holt suggests that all previous research has not given the clinician the same advantage that the statistician has received.[9] He points out that normally the statistical method has an advantage over the clinical method because the statistician is allowed to collect his test scores, test them against the criterion to determine the weights which are to be given to each test from a preliminary sample and then cross-validate them on another sample.

In order to give the clinician the same advantage as the statistician, Holt proposes that research should follow a design in which the psychologist would select meaningful criterion measures, intervening variables and techniques to measure them.[10] Then he would make his predictions on a preliminary sample, test these against a criterion measure and study his hits and misses. In this manner he would estimate the intervening variables and ask a statistician to correlate these with the criterion measure. This process would be repeated on a second sample and the results would be cross-

---

9 R. Holt, "Yet Another Look at Clinical and Statistical Prediction: Or Is Clinical Psychology Worthwhile?" American Psychologist, Vol. 25, 1970, p. 337-349.

10 Ibid., p. 341.

validated. At this point the correlations with the criterion can be compared for statistical and clinical methods.

On the basis of this argument Holt criticizes Meehl, Gough and Sawyer for their surveys. He rejects, as misleading, the classification of studies as clinical or statistical because the primary issue is the role of clinical judgment, which he suggests enters the predictive process at every level. Clinical judgment enters the predictive process in the analysis of the criterion, in discovering situational and intrapersonal intervening variables, in the choice and appropriateness of the measuring instruments, in the first empirical testing of the clinical predictors, in checking the data gathered (both clinical and statistical), and in the combination of the data.[11]

In addition to the above criticisms, Holt criticizes Sawyer's survey and lists the following deficiencies in his research: criterion contamination or un-cross-validated formulas; inadequate criterion measures; misleading classification of judges as "clinicians"; insufficient power to detect differences; and, use of quantitative data only. He also points out several biases which exist in favour of mechanical measurement and prediction in Sawyer's survey.

---

11 Ibid., p. 339-340.

Campbell et al.[12] have carefully studied Holt's proposal and they suggest that his clinical method is essentially the same as the statistical method. If the clinical method is to be a legitimate method the clinician must bear the burden of the method which he uses.

In summary, this investigator selects Sawyer's classification schema as the means for comparing predictive methods because it incorporates two of the most important aspects of the predictive process--data collection and data combination--and it allows for an adequate representation of all methods of prediction, including the clinical interview. Although Holt's argument favouring a different approach for predictor classification is a good one, it is essentially similar to the method which the statistician employs in making his predictions.

2. Classification of Predictor Methods in the Prediction of Managerial Success.

Many methods for predicting managerial success have been studied and reported in the literature. The majority of the studies, however, report only an overall global prediction of subsequent job performance, progress, or potential. An example of a review of this type is reported by Cohen, Moses

---

12 J. Campbell et al., Managerial Behavior, Performance and Effectiveness, New York, McGraw-Hill, 1970, p. 157.

and Byham.[13] Although this is a widely used and legitimate method of prediction, this review of the literature will focus only on studies utilizing a schema to classify and compare predictive methods.

A. Korman's Review of the Prediction of Managerial Performance

Korman surveyed the literature employing Meehl's distinction between the type of predictors and the kind of instruments utilized. Having reviewed forty attempts to predict managerial performance, he concludes that there is no basis for assuming superiority of the statistical method over the clinical method of prediction.[14] His survey suggests that there is evidence which contradicts the research of Meehl and Gough because judgmental predictions exemplified by executive assessment procedures and peer ratings emerge as relatively more useful procedures when compared to other methods. This type of contradictory statement clearly points to a problem often overlooked in many of the previously mentioned surveys of the literature. Each investigation of the literature can

---

13 B. Cohen, J. Moses, and W. Byham, "The Validity of Assessment Centres," an article which is to be published in Journal of Industrial and Organizational Psychology, 1-42 p.

14 A. Korman, "The Prediction of Managerial Performance: A Review," Personnel Psychology, Vol. 21, 1968, p. 316.

easily be biased depending upon each investigator's viewpoint and/or type of research which the investigator selects from the numerous research articles reported in the journals.

Korman divided the studies into the following classes; a summary of the results is included for each section (all samples involve first-line supervision to highest level management).

(a) Psychometric Prediction: Cognitive Ability Tests.- The twelve studies reviewed give positive results in only two cases and one of the two cases has a small number of candidates and the distinct possibility of criterion contamination.

(b) Psychometric Prediction: Objective Personality and Interest Inventories.- Korman observes that although there are not many research studies in this field, examination of reported studies shows that there is little hope for this method of prediction.

(c) Psychometric Prediction: "Leadership Ability" Tests.- Research in this area gives little evidence that managerial "leadership ability" tests and/or characteristics are predictive of managerial behaviour. However, some evidence is given for the Miner Sentence Completion Scale.

(d) Psychometric Prediction: Personal History Data.- The evidence indicates that Personal History Data does have some predictive value for first-line supervisors,

but little good evidence has been accumulated for higher-level managerial and officer groups.

(e) Judgmental Prediction: Executive Assessments.- The word "judgmental" is used rather than "clinical" because it is a more general term that includes judgments from test protocols, peer ratings and superior and faculty ratings. The seven research studies reported indicate that the results are fruitful and promising with correlation coefficients ranging from 0.10 to 0.68.

(f) Judgmental Prediction: Peer Ratings.- The results suggest that the peer rating paradigm is a more consistent and effective predictor of military officer behaviour. It is similarly effective in two other industrial situations: the peer ratings of insurance agents; and, peer ratings of middle-management training programs.

(g) Judgmental Prediction: Superior and Faculty Ratings.- Only superior-faculty ratings give some promise. However, more research is necessary for definitive conclusions.

B. Internal Validity of Assessment Procedures

In 1973 Huck conducted an extensive review of available research related to the historical development of assessment centers and, more specifically, he reviewed all research available on the external and internal validity of

assessment centers.[15] He defines the external validity of assessment centers as the utility of using multiple assessment procedures to predict managerial success. He concludes that this has been established. However, more relevant to this investigation is his concept of internal validity which he defines as the advantage of using assessment center techniques over traditional assessment methods. Although Huck makes no attempt to classify techniques according to the predictor classification schemas developed by Meehl, Sawyer or Holt, his study is relevant to this investigation because many of the studies he summarizes use assessment techniques employed in the present study.

The first study referred to was conducted by Glaser, Schwartz, and Flanagan. They showed that the addition of clinical and situational procedures to two basic ability tests (pure statistical procedures) increased the multiple correlations from .27 to .30 and .33.[16]

The best known and most extensive study of assessment center techniques refered to by Huck was conducted by Bray

---

15 J. Huck, "Assessment Centers: A Review of the External and Internal Validities," Personnel Psychology, Vol. 26, 1973, p. 200-203.

16 R. Glaser, P. Schwartz, and J. Flanagan, "The Contribution of the Interview and Situational Performance Procedures to the Selection of Supervisory Personnel," Journal of Applied Psychology, Vol. 42, 1958, p. 69-73.

and Grant with the American Telephone and Telegraph Company in 1966.[17] They showed that both situational techniques and paper-and-pencil tests must be included in making overall predictions of managerial success because of the additional information which these techniques add to the predictive process.

In another assessment procedure Wollowick and McNamara used the amount of increase in managerial responsibility as a criterion measure and classified the predictors in terms of tests, exercises and characteristics.[18] They found that the multiple correlations with the criterion measure were respectively .45, .41 and .39. When all three of the predictor types were included in the regression equation, the multiple correlation increased to .62. Even though the data was not cross-validated, they concluded that each predictor composite contributes uniquely and substantially to the prediction of managerial success. Furthermore, they found that when the overall assessment rating was derived by statistically combining the data, the multiple correlation

---

17 D. Bray and D. Grant, "The Assessment Center in the Measurement of Potential for Business Management," Psychological Monographs, Vol. 80, No. 17 (Whole No. 625), 1966, p. 1-27.

18 H. Wollowick and W. McNamara, "Relationship of the Components of an Assessment Center to Management Success," Journal of Applied Psychology, Vol. 53, 1969, p. 348-352.

was .62. This represented an increase of .25 over a judgmental combining method but the statistical combination was not cross-validated. They conclude that further research is required to assess the significance of this finding. Wollowick and McNamara's conclusion that further investigation is required in this area is confirmed because Moses reports obtaining a multiple correlation of .44 based on judgmental ratings with managerial progress, while a statistical combination of all data yielded a multiple correlation of .463.[19] It is impossible to make a full method comparison of the techniques since he calculates a multiple correlation of .461 for the behavioural assessment variables but does not calculate the multiple correlation for paper-and-pencil tests alone.

These studies clearly indicate that clinical or judgmental procedures, especially in situational exercises, increase the predictive validity of many managerial assessment programs. It is also clear that each type of predictor composite contributes a substantially unique element to the prediction of managerial success. Whether or not a statistical combination of data is superior to a clinical combination is an area for further research.

---

19 J. Moses, "Assessment Center Performance and Management Progress," a paper presented at the 79th annual meeting of the American Psychological Association, Washington, D.C., 1971.

3. Summary and Basic Hypothesis.

Three models to classify methods of prediction have been mentioned. Each model can be applied to the methods employed to predict managerial success in an assessment center approach.

Sawyer has advanced many sound arguments to demonstrate that Meehl's model does not account for the mode of data collection. If this insight is ignored a confounding variable is introduced and at the same time many crucial procedures are excluded.

Campbell et al. argue that Holt's objection to both Meehl's and Sawyer's model is a good one, but the type of procedures he describes to solve the difficulty demands the same type of group information and group statistics which are required for combining the data statistically. On the basis of the arguments presented by Sawyer and Campbell et al. this investigator accepts Sawyer's predictor classification schema as the most appropriate model for comparing predictive methods in an assessment center program.

Other studies were investigated which related to the prediction of managerial success and these studies indicated that the judgmental procedures were adding significant information to the predictions. Sawyer compared many methods of assessment in industrial settings but the samples employed in

the comparisons were independent of one another and came from a variety of managerial settings. Furthermore, all of the studies related to the prediction of managerial success used either a global prediction rating or a specific assessment dimension rating (Bray and Grant's study) with a global criterion rating of managerial performance. The present investigation correlates prediction ratings on several assessment dimensions with a criterion performance rating corresponding to each assessment dimension. Four methods of prediction (Pure Clinical, Trait Ratings, Pure Statistical and Mechanical Composite) are compared using Sawyer's model.

It is out of this general problem that the present investigation has grown. This writer seeks an answer to a specific question: Are there significant differences between the methods employed to predict managerial performance when the same sample of subjects is used and the data is cross-validated? To be more exact, he hopes to discover which method(s) of prediction is (are) significantly more valid for specified assessment dimensions when compared to other methods; he is concerned with this aspect of the total problem, and it is generally stated in the form of the null hypotheses:

A. There is no significant difference using Sawyer's model between the Pure Clinical and Trait Ratings Methods of Assessment for each **of the following assessment dimensions: communication; self-motivation; tolerance for stress; human relations; and, decision making.**

B. There is no significant difference using Sawyer's model between the Pure Clinical Method and the Pure Statistical Method of Assessment for each of the following assessment dimensions: communication; self-motivation; tolerance for stress; human relations; and, decision making.

C. There is no significant difference using Sawyer's model between the Pure Clinical amd Mechanical Composite Methods of Assessment for each of the following assessment dimensions: communication; self-motivation; tolerance for stress; human relations; and, decision making.

D. There is no significant difference using Sawyer's model between the Trait Ratings and Pure Statistical Methods of Assessment for each of the following assessment dimensions: communication; self-motivation; tolerance for stress; human relations; and, decision making.

E. There is no significant difference using Sawyer's model between the Trait Ratings and the Mechanical Composite Methods of Assessment for each of the following assessment dimensions: communication; self-motivation; tolerance for stress; human relations; and, decision making.

F. There is no significant difference using Sawyer's model between the Pure Statistical and Mechanical Composite Methods of Assessment for each of the following assessment dimensions: communication; self-motivation; tolerance for stress; human relations; and decision making.

CHAPTER II

EXPERIMENTAL DESIGN

This chapter presents the procedures involved in conducting the experiment to test the hypotheses proposed in the preceding chapter.

It begins with a description of each method of prediction from Sawyer's model which is the model selected to classify predictor methods. This is followed by a brief summary of the salient features of the Behavioural Validation Model which was used in this Assessment Centre Program. Within this context a description is given of the job analysis, the development of the dimensions measured in the predictors and the criteria, the predictors as they are employed in Sawyer's predictor classification model, the application of the Assessment Centre, the validation procedures and the managerial decision process for staffing purposes. After this the various statistical procedures, the sample, and the specific hypotheses are described.

1. Predictor Classification Model.

In order to test the general hypotheses stated in Chapter I, Sawyer's model was selected to classify methods of prediction. His model was employed since it reduces the possibility of introducing a confounding variable by

accounting for the mode of data collection. Although Holt's model[1] accounts for the objections he makes to Sawyer's model,[2] the argument presented by Campbell et al.[3] convinced the writer that Sawyer's model is the most appropriate to compare methods of predicting managerial performance.

All of the method comparisons which can be used in Sawyer's model are listed in Table I. The methods employed in this study are:

Pure Clinical;
Trait Ratings;
Pure Statistical; and
Mechanical Composite.

Each of the methods is described below.

A. Pure Clinical Method
(clinically collected data combined clinically)

The clinician predicts behaviour from an interview or other behavioural observations without using any test or other information. There may exist an intermediate step of constructing an explicit assessment of the personality characteristics and the prediction may be based on these assessments rather than directly from the observed behaviour.

---

1 R. Holt, "Yet Another Look at Clinical and Statistical Prediction: Or Is Clinical Psychology Worthwhile?" American Psychologist, Vol. 25, 1970, p. 337-349.

2 J. Sawyer, "Measurement and Prediction, Clinical and Statistical," Studies in Personnel Psychology, Vol. 2, 1970, p. 78-82.

3 J. Campbell et al., Managerial Behavior, Performance and Effectiveness, New York, McGraw-Hill, 1970, p. 157.

A written description of a candidate after he has been interviewed or observed in simulation exercises is an example of this strategy.

B. Trait Ratings Method
(clinically collected data combined mechanically)

The clinician, after he collects the data by interview or observation of the candidate's behaviour, explicitly assesses the candidate on certain pre-specified scaled dimensions in the form of ratings. Instead of allowing the assessor to predict future behaviour, the combination of his assessments into a prediction is performed mechanically (for example, by using a multiple regression equation or a pre-determined weighting system).

C. Pure Statistical Method
(mechanically collected data combined mechanically)

This method is best explained in terms of an example. A standard example of this strategy is the combination of biographical information and test scores in a multiple regression equation or pre-determined weighting system to predict managerial job performance.

D. Mechanical Composite Method
(both modes of collecting data combined mechanically)

This method uses pre-assigned rules (a multiple regression equation or weighting system) to determine the

final behavioural predictions from data which is collected both clinically and mechanically. For example, objective test scores, behavioural ratings and other clinical information are combined in a multiple regression equation.

2. The Application of the Behavioural Validation Prediction Model for the Purpose of Comparing Predictive Methods in an Assessment Centre.

The data for this study was collected in an Assessment Centre program which was specifically developed to meet the requirements of a Canadian federal government department for which managers were being selected. The validation model developed to meet the organization's needs was based on a behavioural validation model described by Slivinski and Grant.[4] They summarize the main points of the model as follows:

> 1. It does not use the traditional trait rating approach to measure occupational effectiveness but does a more careful job of studying actual job dimensions. Once these dimensions are obtained, the pertinent worker behaviours are identified.
> 2. The model dispenses with the hang-up of obtaining the ultimate criterion measure or the preoccupation of trying to encompass subcriteria into an overall global measure.
> 3. The construction of the predictor tests and the criteria are built upon the knowledge gleaned from the job, worker analysis, and organizational climate.
> 4. It allows for setting up typologies for classifying jobs, people, predictors, situations and criteria.

---

[4] L. Slivinski and K. Grant, Behavioural Validation Prediction Model, paper presented to the International Personnel Association, San Francisco, 1971, 1-39 p.

5. As a corollary it allows for the use of different statistical models, be they suppressor variables, moderators, profile analysis.
6. It allows the freedom to use clinical and mechanical data gathering techniques.
7. It goes beyond the paper-and-pencil approach and allows for simulated exercises.
8. Use of multiple assessors with multiple techniques with multiple situations for multiple variables ensures greater and more reliable information.
9. The model allows for a marriage between management and professional input.
10. It also allows for managerial decision making.[5]

The development of the dimensions measuring managerial potential and performance, the analysis of the climate of the organization, the description of the predictors, the training of the assessors, the organization's requirements for assessee eligibility, the validation process and the placement of the assessees in government positions are described in the context of the Behavioural Validation Prediction Model.[6] Sections A to G describe the actual steps used to develop this approach in this Assessment Centre program for a federal government department.

A. Job Analysis

The Executive Position Description Questionnaire (EPDQ) identified the important job dimensions and types of positions within the job. The results are summarized in

---

5 Ibid., p. 28-29.

6 Ibid., p. 25-28 for a description of each of the seven steps.

Identification of Middle Management Job Dimensions and Job Types for Use in an Assessment Centre.[7]

B. Managerial Dimensions and Criteria

The managerial dimensions were obtained by developing a Behavioural Anchored Rating Scale (BARS) which, in effect, resulted in the formulation of multiple criteria. The procedure for the development of the BARS followed a method recommended by Smith and Kendall.[8] The complete description of the formulation of the BARS and the dimensions that resulted from this study are contained in a paper by Grant and Slivinski.[9] A summary of the procedure follows.

A group of managers from various regional offices across Canada met to decide on the dimensions which were important to managerial performance, utilizing dimensions identified by EPDQ and adding anything which they considered relevant. Then a second group of managers assigned job

---

7 M. Leim, L. Slivinski, and K. Grant, Identification of Middle Management Job Dimensions and Job Types for Use in an Assessment Centre, Public Service Commission, Ottawa, 1972, 1-55 p.

8 P. Smith and L. Kendall, "Retranslation of Expectations: An Approach to the Construction of Unambiguous Anchors for Rating Scales," Journal of Applied Psychology, Vol. 47, 1963, p. 149-155.

9 K. Grant and L. Slivinski, Identification of Assessment Factors in an Assessment Centre, Public Service Commission, Ottawa, 1973, 1-100 p.

behaviours to each of the dimensions in order to define high, medium, and low examples. The 239 behavioural statements generated were retranslated by a third group of managers. This yielded the modal number of behavioural examples assigned to each dimension and indicated which of the managerial dimensions could not be differentiated in terms of specific on-the-job behaviour.

The final selection of dimensions was determined by establishing relevant cutoff measures. A final workshop of managers determined the weights (using a nine-point continuum) to be assigned to each of the 239 behavioural statements. The final format of the criteria measures and the definition for each of the five dimensions are presented in Figures 1 to 5.[10] The advantage of using a BARS approach is that it reveals the behavioural basis for managerial effectiveness.[11]

C. Organizational Climate Analysis

Although there was no study designed explicitly to assess the organizational climate, various aspects of the Assessment Centre program allowed for the incorporation of some aspects of the organizational climate. For example, the assessors utilized in the program were from middle

---

10 Cf. Appendix 2 for Figures 1 to 5.

11 J. Campbell et. al., Op. Cit., p. 123-124.

management positions in the Department and came from six Regional Departments across Canada. Each assessor attended a two-week training program prior to actual assessment procedures to acquaint himself with his role as an assessor and to integrate the purpose of the Assessment Centre with on-the-job functioning. In addition, the managers who developed the BARS were from the Department and knew the Department's needs.

D. Construction and Selection of Predictor Tests

The predictor tests were selected and developed from the knowledge obtained from the preceding steps. The techniques used to predict managerial performance are described under each of the four methods of Sawyer's model employed in this investigation for classifying methods of prediction. The predictor tests are described in terms of how they were employed for this investigation and only for the five dimensions utilized for the purpose of this research.

(a) Pure Clinical Method of Assessment (Integration Ratings)

The Integration Rating process[12] is a Pure Clinical Method of assessment since the data is collected and combined by clinical procedures. The assessors review the behavioural data which has been collected from all of the exercises for each

---

12 Cf. Appendix 3 for additional information on the Integration procedures. Appendix 4 gives a typical example of the behavioural ratings scales used in the assessment centre.

assessment dimension. This allows each assessor to exchange his observations of each assessee with the other assessors. To evaluate each assessee a chairman asks each assessor to read his report. While one assessor reads his report the other assessors make notes of the behaviours observed by the assessor. After all behavioural information is reviewed for each dimension in turn, each assessor independently rates the assessee. If ratings differ by more than one scale point, then a discussion follows to ascertain whether or not differences can be worked out by reviewing the evidence. Then a final rating is done which may or may not be unanimously accepted. If the final ratings are not unanimously accepted, an average of the assessors' final ratings is taken for each prediction on the following dimensions: communication; self-motivation; tolerance for stress; human relations; and, decision making.

The initial and final ratings for the integration process are based on a six-point continuum with the following rating classifications: no evidence (0); definitely inadequate (1); slightly less than adequate (2); adequate or satisfactory (3); more than adequate (4); and, outstanding (5).

(b) Trait Ratings Method of Assessment

The following exercises are Trait Ratings Method of assessment since the data is collected clinically through behavioural observation and is combined mechanically by

using a regression equation (the weights from one sample are applied to another sample and vice versa).

(i) The Assessment Interview.[13] The Assessment Interview gives the assessor an opportunity to meet and talk to each assessee in an informal one-to-one setting. The assessor reviews the assessee's work history, his attitude towards different kinds of working conditions, his achievements, his educational background, present interests and hobbies, what was satisfying to him in a job, and his future plans and career aspirations. The Assessment Interview provides the assessor with an opportunity to evaluate characteristics and qualities of the assessee which are not brought out in other structured exercises. Guidelines for structuring the interview, interpreting interview information in terms of managerial dimensions and reporting the interview are presented to assist assessors in achieving consistency.

From this information-seeking process, the assessors make judgments on each of the following assessment dimensions: communication; self-motivation; tolerance for stress; and, human relations. The ratings are based on a five-point scale ranging from one to five with each of the five points on the continuum having a summary statement of what would be expected from the assessee to achieve the rating.

---

13 Cf. Appendix 5 for more details on the Interviewing procedures.

(ii) Group Situations.- Situational or simulation testing is a relatively new approach for evaluating managerial potential. In general, the purpose is to place a group of assessees in a variety of situations related to the type of job in question. Rather than asking the assessee how he would perform the task (as in an interview or paper-and-pencil test) the emphasis becomes one of how he does the task. From the behaviour which the assessee manifests, conclusions are drawn. Besides a general contribution to the problem, the assessors observe specific behaviours relevant to certain managerial dimensions.

Group Situations: Task Force Exercise.[14] This is a problem-solving exercise in which the assessees have certain roles assigned to them. The assessees who assume the role of vice-presidents of a company have to decide on one of three possible courses of action for the company. The assessees are given a short time interval to identify with their various roles (marketing, finance, personnel, manufacturing, engineering, and administration). Competition is built into the exercises because each assessee receives a predetermined role and point of view in regard to the problem under discussion. Cooperation is also a part of the exercise since the assessees are instructed to arrive at a consensus by the end of the exercise. Behaviour relevant to two assessment dimensions, communication and human relations,

---

[14] Cf. Appendix 6 for more information on the Task Force Exercise.

is noted. These behaviours are rated on a six-point scale with relevant behavioural descriptions accompanying each point on the scale.

Group Situations: Personnel Problem.[15] - The Personnel Problem is a leaderless group discussion exercise in which the assessees are presented with some general concerns of a personnel nature. The assessees represent members of a special committee to formulate a plan of action regarding increasing numbers of customer complaints, low morale, and an increasing number of satisfactory performance appraisals. The group must not only reach a consensus on their approach to the problem, but must formulate the actual recommendations.

The exercise differs from the Task Force Exercise in that the roles are not predetermined for the assessees and no specific solutions or approaches to the problem are suggested to them. Neither competitiveness nor cooperativeness is built into the exercise.

Assessors observe individual approaches to solving the problem as well as noting behaviours relevant to the assessment dimensions of communication and human relations. A six point rating scale is used with each point on the scale having a description of the behaviour relevant to the rating.

(iii) Individual Situations.- This type of exercise refers to simulation situations which require the assessee to work alone.

---

15 Cf. Appendix 7 for a further description of the Personnel Problem Exercise.

Individual Situations: Individual Case Presentation Oral.[16] In this exercise the assessee has to analyze a statistical report on summer student employment and give a short oral presentation. The problem is given to the assessee the night before with the appropriate instructions. The assessee can work as little or as much as he desires. The oral presentation has to be ready for the following day. The focus is on the ability of the assessee to work alone, structure a rather complex mass of data and draw relevant conclusions and recommendations for the following year's employment program. Another important aspect under observation is the ability of the candidate to make an oral presentation.

In the Individual Case Presentation Oral exercise, the assessors evaluate the assessees' ability to communicate orally and tolerate stress. Each of these dimensions is assessed on a six-point scale with a relevant behavioural description accompanying each of the points on the scale.

Individual Situations: The In-basket.- The test which was used in this assessment is the Consolidated Fund In-basket Test developed by Educational Testing Service, Princeton, New Jersey. The In-basket Test is an exercise which places each assessee in a manager's position for approximately two and one half hours. The assessee is given an organization chart, a short description of his

---

16 Cf. Appendix 8 for more details on the Individual Case Presentation Oral.

staff and a brief history of the organization. The exercise takes the form of the preceding manager's In-basket. It is filled with letters, memos, organizational problems, pay sheets, etc. to which the assessee (the new manager) will have to respond. He (she) is free to write letters, solve problems, delegate problems, etc. in whatever fashion he (she) deems necessary.

Following a detailed procedure, the assessors evaluate (score) each assessee's In-basket Test and interview him regarding his manner of handling the In-basket material. Managerial dimensions under observation are communication and decision making. Each assessee is assessed on a five-point scale having behavioural statements attached to each point of the scale.[17]

(c) Pure Statistical Methods of Assessment

The following tests are pure statistical methods of assessment since the data is collected mechanically (test scores) and is combined mechanically by a regression equation (the weights from one sample are applied to another sample and vice versa).

(i) Personality Tests.- Three paper-and-pencil tests are employed to assess a variety of personality traits. The three paper-and-pencil tests are: the Gordon Personal Profile (GPP); the Gordon Personal Inventory (GPI); and, the Self Description Inventory (SDI). In all tests the individual describes himself.

---

17 Cf. Appendix 9 for a further description of the In-basket.

Personality Tests: Gordon Personal Profile. The Gordon Personal Profile measures four aspects of the normal personality. Two of these, namely, emotional stability and ascendancy, were selected as appropriate measures for the respective assessment dimensions of tolerance for stress and decision making.

High and low scores on the Scale of Ascendancy are interpreted in the following manner:

> Those individuals who are verbally ascendant, who adopt an active role in the group, who are self-assured and assertive in relationships with others, and who tend to make independent decisions, score high on this Scale. Those who play a passive role in the group, who listen rather than talk, who lack self-confidence, who let others take the lead, and who tend to be overly dependent on others for advice, normally tend to make low scores.[18]

High and low scores on the Emotional Stability Scale are interpreted as follows:

> High scores on this Scale are generally made by individuals who are well-balanced, emotionally stable, and relatively free from anxieties and nervous tension. Low scores are associated with excessive anxiety, hypersensitivity, nervousness, and low frustration tolerance. Generally, a very low score reflects poor emotional balance.[19]

Personality Tests: Gordon Personal Inventory.- Like the Gordon Personal Profile, the Gordon Personal Inventory measures four personality traits which determine the

---

18 L. Gordon, Gordon Personal Profile, New York, Harcourt, Brace & World, 1963, p. 3.

19 Ibid.

adjustment of normal individuals in an industrial situation. Two of these, namely, Cautiousness and Personal Relations, were used as measures for the respective assessment dimensions of decision making and human relations. The interpretations given to high and low scores of these scales are as follows:

> Cautiousness
> Individuals who are highly cautious, who consider matters very carefully before making decisions, and do not like to take chances or run risks, score high on this Scale. Those who are impulsive, act on the spur of the moment, make hurried or snap decisions, enjoy taking chances, and seek excitement, score low on this Scale.[20]

> Personal Relations
> High scores are made by those individuals who have great faith and trust in people, and are tolerant, patient, and understanding. Low scores reflect a lack of trust or confidence in people, and a tendency to be critical of others and to become annoyed or irritated by what others do.[21]

Personality Tests: Self Description Inventory.[22] The Self Description Inventory measures thirteen personality traits, two of which were used as Pure Statistical Methods of predicting managerial performance on-the-job. The two measures which were selected are achievement motivation and decisiveness **and were** utilized for the respective assessment dimensions of self-motivation and decision making. Each

---

20 L. Gordon, Gordon Personal Inventory, New York, Harcourt, Brace & World, 1963, p. 3.

21 Ibid.

22 Cf. Appendix 10 for more details on the Self Description Inventory.

scale in the Self Description Inventory was developed empirically.

Achievement motivation evaluates the need of some individuals to achieve appointments to high level occupations and the need of others to remain content with their present status and level of success. The criterion groups consisted of men and women from top management and professional jobs as well as men and women from semi-skilled and unskilled jobs. Higher scores are related to the challenge of executive and administrative duties and the need to progress to higher levels, whereas lower scores are related to the performance of prescribed tasks.

The decisiveness scale is intended to measure decision making approaches: those at the high end exhibiting a decisive behavioural approach; those at the low end exhibiting a cautious approach. The criterion groups for this scale consisted of people in top management and middle management positions.

(ii) Aptitude Test (Structure of Intellect Model).- An Expressional Fluency test developed by Christensen and Guilford measures the DMS factor (divergent-production semantic systems) in Guilford's Structure of Intellect Model.[23]

---

23 J. Guilford, The Nature of Human Intelligence, New York, McGraw-Hill, 1967, p. 150-152.

The test consists of constructing a variety of four-word sentences from four initial letters. This test is used as a measure for the assessment dimension of communication.

(d) Mechanical Composite Method of Assessment

The Mechanical Composite Method of assessment differs from the other methods in that all of the material obtained from the other methods is combined mechanically using a multiple regression equation and the weights from one sample are applied to another sample and vice versa. Like the other methods of assessment, all five managerial dimensions are measured.

E. Application of the Assessment Centre

As the previous paragraphs have outlined, behavioural data will be collected clinically, mechanically, or, both clinically and mechanically and will be combined clinically or mechanically according to Sawyer's model for each of the five managerial assessment dimensions.

The assessees are comprised of non-managerial workers within a Canadian federal government department. They were eligible for assessment providing they were self-nominated and met federal government standards. The assessees were required to meet government educational standards, be within certain job classifications, and be geographically mobile.

As previously mentioned the assessors received two weeks of training and were from the government department for which the assessees were being selected. Ratings were obtained from three assessors for all simulation exercises except the Assessment Interview and In-basket Test.

F. Validation

Supervisors' ratings were obtained for each assessment dimension utilized for this investigation on the Behavioural Anchored Rating Scale six months after the assessees were assessed at the Assessment Centre. All of the 108 assessees were placed in Canadian Public Service departmental positions after assessment.

G. Managerial Decisions for Staffing Purposes

All of the pertinent information was presented to management so that they could make the necessary decisions for the placement of each assessee.

3. Statistical Procedures.

The statistical procedures which were used to analyze the data for comparing the methods of assessment are presented in sections A to E.

A. Hypothesis Concerning a Population Correlation

The Pearson product-moment correlation coefficients of the predictor with the criterion for each assessment dimension were tested using the hypothesis $\bar{r} = 0$. Fisher has developed a t ratio to test for the significance of a coefficient of correlation using the formula

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

where r = the Pearson product-moment correlation coefficient;

N = the number of cases; and,

N-2 = the number of degrees of freedom.[24]

For those methods of assessment having multiple predictors for each assessment dimension, a multiple correlation coefficient (stepwise regression model with forced entry of variables) indicates the correlation between the predictors and the criterion measure for each dimension. Reference to the tables in Guilford and Fruchter gives the value required for significance to test the hypothesis $\bar{R} = 0$.[25]

---

24 J. P. Guilford and B. Fruchter, Fundamental Statistics in Psychology and Education, New York, McGraw-Hill, 1973, p. 145.

25 Ibid., p. 367.

B. Cross-validation

The 108 subjects were divided into two randomized samples for cross-validating each method of assessment for each assessment dimension. The double cross-validation design suggested by Mosier allows for the application of the regression weights from the first sample (N = 54) to the second sample (N = 54) and vice versa.[26] The double cross-validation design was used for all methods of prediction.

The un-cross-validated and cross-validated correlation coefficients for samples A and B were compared by testing for the significance of the difference between two correlation coefficients for independent samples; that is, whether the two samples can be considered random samples from a common population. The following formula tested for significant differences between two correlation coefficients:

---

26 C. Mosier, "Symposium: The Need and Means of Cross-validation I. Problems and Designs of Cross-validation," Educational and Psychological Measurement, Vol. 11, 1951, p. 11.

$$z = \frac{z_{r_1} - z_{r_2}}{\sqrt{\frac{1}{(N_1-3)} + \frac{1}{(N_2-3)}}}$$

where z_{r_1} = Fisher's z transformation of the correlation coefficient for sample A;

z_{r_2} = Fisher's z transformation of the correlation coefficient for sample B;

N_1 = the number of subjects in sample A; and,

N_2 = the number of subjects in sample B.[27]

C. Reliability Estimates of Predictors and Criteria

The reliability estimates of the assessors' ratings for each assessment dimension on the following exercises were reported using Guilford's $\bar{r}_{cc}$ and r_{kk} statistic:

Integration Ratings;

Individual Case Presentation Oral;

Task Force Exercise; and,

Personnel Problem.

The $\bar{r}_{cc}$ statistic indicates the average of the intercorrelations of the ratings and is given by the formula

$$\bar{r}_{cc} = \frac{V_r - V_e}{V_r + (k-1)V_e}$$ (Intraclass correlation among k series)

where V_r = variance between rows, where each row stands for the person;

---

27 G. Ferguson, Statistical Analysis in Psychology and Education, New York, McGraw-Hill, 1971, p. 170-171.

V_e = variance for residuals (or error); and,

k = number of columns.[28]

This result can be interpreted as the typical reliability of a single rater's rating.

The r_{kk} indicates the reliability of a sum or mean of these three raters' ratings in this population and is given by the formula:

$$r_{kk} = \frac{V_r - V_e}{V_r} .$$

This indicates that if an average of the ratings for each rater could be correlated with a similar set of averages, the result would be r_{kk}.[29]

A test-retest design over a period of six weeks estimated the reliability of the supervisors' on-the-job performance ratings for each assessment dimension. The coefficient of stability (r_{tt}) was calculated using the Pearson product-moment correlation coefficient.

D. Linearity of Regression Measures

Guilford suggests that "it behooves the investigator who employs multiple regression procedures to assure himself

---

28 Guilford and Fruchter, op. cit., p. 263-264.

29 Ibid., p. 264.

that his regressions are linear."[30] All regressions were tested for linearity using Guilford's F test for linearity:

$$F(Lin) = \frac{(\eta^2-r^2)(N-k)}{(1-\eta^2)(k-2)}$$

where k = the number of columns or rows; and,

and $k-2$ / $N-k$ = the number of degrees of freedom.[31]

The correlation ratio for the regression of Y on X (η_{yx}) and X on Y (η_{xy}) was calculated to test for linearity of regression of the predictor on the criterion and the criterion on the predictor.

E. Statistical Formula to Test the Null Hypothesis

Hotelling has developed a t test that compares the coefficients of correlation when they are derived from correlated samples.[32] The Hotelling formula is:

$$t_{d_r} = (r_{12}-r_{13})\sqrt{\frac{(N-3)(1-r_{23})}{2(1-r_{22}^2-r_{12}^2-r_{13}^2-2r_{23}r_{12}r_{13})}}$$

where r_{12} = the correlation between the predictor and the criterion for a method of prediction;

---

30 J. Guilford, Psychometric Methods, New York, McGraw-Hill, 1954, p. 404.

31 Guilford and Fruchter, op. cit., p. 290-291.

32 Ibid., p. 167.

r_{13} = the correlation between the predictor and the criterion for a second method of prediction;

r_{23} = the intercorrelation of the two predictor methods; and

N-3 = the number of degrees of freedom.

A two-tail test was called for to test the hypothesis that there is no significant difference between two validity coefficients since directionality is unknown. Generalization was limited only to a subpopulation of all possible samples for which X and Y have exactly the same set of values as those in the observed sample.

4. A Description of the Sample.

The total group of candidates was divided into two randomized samples, each consisting of fifty-four subjects. All of the biographical data for the two samples was compared employing either a t test or a chi square test for independent samples. Out of the twenty-four variables compared between samples A and B, none of the variables were significantly different. This clearly indicates that the two samples originate from a common population.

A. Age and Salary (Table III)

The average present age for both samples was 40. The average age for completing high school was approximately 18

Table III.

Students' t Comparisons for the Independence of Samples A and B for Biographical Data Related to Age and Salary.

Variable	Group	N	Means	SD	Range	df	t	t .05 (two-tail test)
Age	A	54	40.44	11.44	21-58	106	0.237NS	1.98
	B	54	39.89	12.65	22-59			
Age when completed high school	A	50	18.06	3.56	15-39	99	0.62NS	1.98
	B	51	17.69	2.23	15-29			
Age when completed college or university	A	14	24.00	4.47	16-33	23	0.54NS	2.07
	B	11	23.09	3.33	19-32			
Present salary at time of assessment	A	54	9,195.11	888.07	7,138-11,000	106	-0.98NS	1.98
	B	54	9,612.78	2,962.81	7,100-30,000			

for both samples A and B. Sample A completed college or university at an average age of 24, while sample B's average age was 23. The salary at the time of assessment for sample A was approximately $9,195 and was about $9,600 for sample B.

B. Education (Table IV)

The educational level of both samples varied from some grade school to various educational degrees. Approximately 1 out of 3 candidates for each sample completed some high school and about 1 out of 5 graduated from high school. However, about 1 out of 5 candidates for sample B completed some college education and 1 out of 10 received a **Baccalaureate** degree, while 1 out of 6 for sample A obtained some college and the same proportion received a **Baccalaureate degree.** Approximately 1/2 of the subjects in both samples received their high school education in British Columbia, about 1/3 of sample A received their college education in British Columbia, while another 1/3 received their college education in Quebec. About 1/3 of sample B received their college education in Ontario.

C. Marks in School Courses (Table V)

Approximately 1 out of 4 subjects from both samples received 75% or higher during high school in history and

Table IV.

Chi Square Comparisons for the Independence of Samples A and B for Biographical Data Related to Education.

Variable	Group	Categories										N	df	χ^2	$\chi^2_{.05}$ table value
		a	b	c	d	e	f	g	h						
Educational level[1]	A	1	0	19	10	9	6	0	7			54	7	4.13NS	14.07
	B	1	2	18	9	12	6	1	5			54			
		a	b	c	d	e	f	g	h	i	j				
Province of high school education[2]	A	30	0	4	3	6	6	1	1	1	1	53	9	8.43NS	16.92
	B	27	2	6	1	7	6	0	0	0	4	53			
		a	b	c	d	e	f	g	j						
Province of college or university education[3]	A	6	0	1	2	2	5	1	0			17	7	7.09NS	14.07
	B	6	2	1	2	6	2	0	1			20			

1 Category a: Some grade school
Category b: Graduated grade school
Category c: Some high school
Category d: Graduated high school
Category e: Some college
Category f: Bachelor degree
Category g: Post-graduate work (M.A.)
Category h: Other degrees

No one stated that their highest educational level was either an Honors degree or a Ph.D.

2 Category a: British Columbia
Category b: Alberta
Category c: Saskatchewan
Category d: Manitoba
Category e: Ontario
Category f: Quebec
Category g: New Brunswick
Category h: Nova Scotia
Category i: Newfoundland
Category j: Outside Canada

3 Categories a to j are the same as in Province of High School Education.

Table V.

Chi Square Comparisons for the Independence of Samples A and B for Biographical Data Related to Marks in School Courses.

Variable	Group	Categories																		N	df	χ^2	$\chi^2_{.05}$ table value
Courses in which 75% or higher was obtained in high school[1]		a	b	c	d	e	f	g	h	i	j	k											
	A	4	0	8	1	11	8	6	3	0	1	1								43	10	12.11NS	18.31
	B	1	2	14	1	12	3	4	1	1	3	3								45			
Courses in which 75% or higher was obtained in college or university[2]		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q					
	A	1	0	0	1	1	0	1	0	1	0	2	1	0	1	0	2	3		14	16	21.96NS	26.30
	B	1	4	1	1	0	1	2	1	0	2	0	1	4	0	1	0	1		20			

1 Category a: Chemistry
Category b: Physics
Category c: Mathematics
Category d: Biology
Category e: History and Geography
Category f: Languages
Category g: Literature
Category h: Bookkeeping
Category i: Religion
Category j: Industrial Arts
Category k: Other

No one stated that they achieved marks of 75% or higher in zoology, economics or typing. The possibility exists that many did not have the opportunity to take these courses.

2 Category a: Accounting
Category b: Business Administration
Category c: Public Administration
Category d: Commerce
Category e: Economics
Category f: Industrial Relations
Category g: Mathematics
Category h: Chemistry
Category i: Finance
Category j: Geography
Category k: Sociology
Category l: Psychology
Category m: History
Category n: Philosophy
Category o: Pure and Applied Sciences
Category p: Languages
Category q: Other

No one stated that they achieved marks of 75% or higher in statistics, physics, computer science, finance, marketing, anthropology, political science, agriculture, or journalism. The possibility exists that many did not take these courses.

geography, while 1 out of 3 for sample A and 1 out of 5 for sample B obtained 75% or better in mathematics. In college or university about 1 out of 4 for sample B received 75% or higher in business administration, the same proportion received 75% or better in history, while 1 out of 5 for sample B reported receiving 75% or better in other courses.

D. Language (Table VI)

Each candidate had the opportunity to be assessed in either the French or English language. In both samples, 9 out of 10 candidates were assessed in English, while the remainder were assessed in French. About 4 out of 5 candidates stated that their mother tongue was English, with the remainder being French or other. About 9 out of 10 said that they were fluent in the English language. The same proportion said that their first working language was English.

E. Family Background (Table VII)

Most of the candidates were married: approximately 7 out of 10 in sample A; about 9 out of 10 in sample B. Approximately 10% of both samples stated they were the only child in their family, while the majority stated they had from 1 to 4 brothers and/or sisters. In sample A, 1/3 were born first in the family and 1/3 were born second, while

Table VI.

Chi Square Comparisons for the Independence of Samples A and B for Biographical Data Related to Language.

Variable	Group	Categories			N	df	χ^2	$\chi^2_{.05}$ table value
		English	French					
Language of assessment	A	48	6		54	1	0.19NS	3.84
	B	49	5		54			
		English	French	Other				
Mother tongue	A	45	7	2	54	2	2.18NS	5.99
	B	42	6	6	54			
		English	French	Both				
Language most fluent in[1]	A	47	6	1	54	2	1.49NS	5.99
	B	50	4	0	54			
		English	French					
First working language[2]	A	50	4		54	1	0.71NS	3.84
	B	52	2		54			

1 No one stated they were fluent in any other language.

2 No one stated that their first working language was a language other than French or English.

Table VII.

Chi Square Comparisons for the Independence of Samples A and B for Biographical Data Related to Family Background.

Variable	Group	Categories										N	df	χ^2	$\chi^2_{.05}$ table value
		Single	Married	Div/Sep	Widow(er)										
Marital status	A	13	39	1	1							54	3	4.40NS	7.82
	B	7	47	0	0							54			
		Male	Female												
Sex	A	54	0									54	1	2.04NS	3.84
	B	52	2									54			
		1	2	3	4	5	6	7	8	9	13				
		(number)													
Number of brothers and sisters including yourself	A	4	16	10	10	6	5	1	0	0	1	53	9	6.38NS	16.92
	B	6	10	11	8	6	5	3	1	2	2	54			
		1	2	3	4	5	7	9	13						
		(position)													
Position in family	A	17	18	9	5	3	0	0	1			54	7	9.74NS	14.07
	B	22	9	9	3	7	2	1	0			53			

2/5 of sample B were born first and 1/6 were born second. All the candidates in sample A were male, while 2 out of 54 were female in sample B.

F. Present Job and Publications (Table VIII)

The number of years of related experience prior to assuming the present job varied greatly for both samples. About 2 out of 5 for sample A and 1 out of 3 for sample B said they had from 0 to 3 years of related experience. A few candidates from both samples had as much as 24 to 27 years of related experience. Approximately 4 out of 5 candidates for both samples said that the first factor which influenced the choice of their career was personal interest. At the time of assessment 7 out of 10 candidates for each sample were working in British Columbia. Only 2 of the candidates for sample A and 1 for sample B had published any articles.

G. Areas Lived In (Table IX)

Approximately 1/2 **of the candidates lived in** cities from ages 5 to 13, 14 to 17, and 18 to the present, while the remainder lived in either towns or rural areas during these ages. About 1/2 of the candidates for both samples lived in 1 province; 1 out of 2 for sample B lived in one city; 2 out of 5 for both samples never lived in a

Table VIII.

Chi Square Comparisons for the Independence of Samples A and B for Biographical Data Related to Present Job and Publications.

Variable	Group	Categories									N	df	χ^2	$\chi^2_{.05}$ table value
Number of years related experience for present job		0-3	4-7	8-11	12-15	16-19	20-23	24-27						
	A	20	6	3	7	7	3	4			50	6	5.48NS	12.59
	B	19	11	6	10	3	1	4			54			
Factors influencing choice of job (first one only)[1]		Personal interest		Other people		Opportunities available		Income		Benefits of job				
	A	45		4		3		1		1	54	4	4.06NS	9.49
	B	39		2		9		1		1	52			
Present province of work[2]		1	2	3	4	5	6	7	8	9				
	A	36	0	1	1	6	7	1	1	1	54	8	4.75NS	15.51
	B	41	1	1	1	5	5	0	0	0	54			
Publications		Yes		No										
	A	2		52							54	1	0.34NS	3.84
	B	1		53							54			

1 No one stated that social status of the job, training, and development opportunities or other factors were the primary factors which influenced their choice of job or career.

2 Category 1: British Columbia
Category 2: Alberta
Category 3: Saskatchewan
Category 4: Manitoba
Category 5: Ontario
Category 6: Quebec
Category 7: New Brunswick
Category 8: Nova Scotia
Category 9: Yukon

Table IX.

Chi Square Comparisons for the Independence of Samples A and B for Biographical Data Related to Areas Lived In.

Variable	Group	Categories								N	df	χ^2	$\chi^2_{.05}$ table value
		Rural	Town	City									
Area lived in from age 5 to 13	A	15	15	24						54	2	1.20NS	5.99
	B	17	10	26						53			
		Rural	Town	City									
Area lived in from age 14 to 17	A	8	14	32						54	2	2.90NS	5.99
	B	15	12	26						53			
		Rural	Town	City									
Area lived in from age 18 to present	A	5	11	38						54	2	.004NS	5.99
	B	5	11	37						53			
		0	1	2	3	4	5	6					
		(number)											
Number of provinces lived in	A	0	24	16	9	5	0	0		54	6	6.50NS	12.59
	B	1	27	18	4	2	1	1		54			
		0	1	2	3	4	5	6	8				
		(number)											
Number of cities lived in	A	6	20	12	4	7	3	1	1	54	7	6.35NS	14.07
	B	8	27	11	4	2	2	0	0	54			
		0	1	2	3	4	5	6	9				
		(number)											
Number of towns lived in	A	21	17	7	5	1	1	1	1	54	7	3.20NS	12.02
	B	23	10	10	4	2	2	1	1	53			
		0	1	2	3	4	9						
		(number)											
Number of rural areas lived in	A	33	13	8	0	0	0			54	5	7.43NS	11.07
	B	27	15	5	2	2	2			53			

town; and 3 out of 5 for sample A and 1 out of 2 for sample B never lived in a rural area.

H. Interests (Table X)

The first present interests of both samples were very similar. About 4 out of 5 said their first interest was reading. Approximately 80% of the first personal talents which the candidates listed were about evenly distributed among musical talent, artistic talent and crafts. About 2 out of 5 candidates for sample A and 1 out of 2 for sample B belonged to athletic organizations.

5. Specific Null Hypotheses

Each of the problems which the writer has identified are stated in terms of the null hypotheses as follows:

1. There is no significant difference using Sawyer's model for the assessment dimension of communication between:

    a. the Pure Clinical and Trait Ratings Methods;
    b. the Pure Clinical and Pure Statistical Methods;
    c. the Pure Clinical and Mechanical Composite Methods;
    d. the Trait Ratings and Pure Statistical Methods;
    e. the Trait Ratings and Mechanical Composite Methods; and,
    f. the Pure Statistical and Mechanical Composite Methods of Assessment.

Table X.

Chi Square Comparisons for the Independence of Samples A and B for Biographical Data Related to Interests.

Variable	Group	Categories						N	df	χ^2	$\chi^{2}_{.05}$ table value
		Reading	Sports (participation)	Sports (observer)	Movies	Other					
Present interest (first choice)[1]	A	45	5	2	0	1		53	4	2.34NS	9.49
	B	44	7	2	1	0		54			
		Musical Talent	Artistic Talent	Professional Sportsmanship	Crafts (building,etc.)	Other					
Personal talents not associated with work (first choice)	A	15	13	4	12	4		48	4	0.26NS	9.49
	B	13	10	4	12	4		43			
		Athletic	Hobby	Fraternal	Civic (community)	Business (professional)	Other				
Organizations belonged to	A	17	5	11	2	1	6	42	5	3.59NS	11.07
	B	18	5	6	5	1	3	38			

1 No one said that their first interest or activity was either television, a hobby, a musical performance or a music listener.

2. There is no significant difference using Sawyer's model for the assessment dimension of self-motivation between:

    a. the Pure Clinical and Trait Ratings Methods;
    b. the Pure Clinical and Pure Statistical Methods;
    c. the Pure Clinical and Mechanical Composite Methods;
    d. the Trait Ratings and Pure Statistical Methods;
    e. the Trait Ratings and Mechanical Composite Methods; and,
    f. the Pure Statistical and Mechanical Composite Methods of Assessment.

3. There is no significant difference using Sawyer's model for the assessment dimension of tolerance for stress between:

    a. the Pure Clinical and Trait Ratings Methods;
    b. the Pure Clinical and Pure Statistical Methods;
    c. the Pure Clinical and Mechanical Composite Methods;
    d. the Trait Ratings and Pure Statistical Methods;
    e. the Trait Ratings and Mechanical Composite Methods; and,
    f. the Pure Statistical and Mechanical Composite Methods of Assessment.

4. There is no significant difference using Sawyer's model for the assessment dimension of human relations between:

    a. the Pure Clinical and Trait Ratings Methods;
    b. the Pure Clinical and Pure Statistical Methods;
    c. the Pure Clinical and Mechanical Composite Methods;
    d. the Trait Ratings and Pure Statistical Methods;
    e. the Trait Ratings and Mechanical Composite Methods; and,
    f. the Pure Statistical and Mechanical Composite Methods of Assessment.

5. There is no significant difference using Sawyer's model for the assessment dimension of **decision making** between:

    a. the Pure Clinical and Trait Ratings Methods;
    b. the Pure Clinical and Pure Statistical Methods;
    c. the Pure Clinical and Mechanical Composite Methods;
    d. the Trait Ratings and Pure Statistical Methods;
    e. the Trait Ratings and Mechanical Composite Methods; and,
    f. the Pure Statistical and Mechanical Composite Methods of Assessment.

CHAPTER III

PRESENTATION OF RESULTS

The results presented and interpreted in this chapter test the null hypotheses presented in Chapter II. Chapter III begins with a presentation of the reliability estimates of the predictors and the criteria measures. This is followed by a presentation and interpretation of the results obtained for each assessment dimension. Within each of these sections, the conclusions are expressed in terms of maintaining or rejecting the null hypotheses for each assessment dimension.

1. Reliability.

The reliability of a predictor or criterion measure can place limits on validity, and the crucial question thus becomes whether reliability is high enough to allow satisfactory validity. Thus the reliability estimates were calculated for most of the predictors and for each criterion measure.

A. Reliability of the Predictors

In the simulation exercises and integration procedure three assessors rated each assessee on five assessment dimensions. To accommodate this type of data two types of

reliability coefficients were calculated: the $\bar{r}_{cc}$ gives the intrinsic agreement of the three assessors, in other words, the average of the intercorrelation of the ratings which is interpreted as the typical reliability of a single assessor's rating; and, the r_{kk} gives the reliability of a mean of the three assessors' ratings in this population.

Table XI presents the results. The $\bar{r}_{cc}$'s range from .46 to .78 for the assessment dimension of communication. The F-ratio for the Personnel Problem indicates that there were significant differences ($p<.01$) between the assessors' ratings. The r_{kk}'s vary from .72 to .92.

The $\bar{r}_{cc}$ for the Integration Ratings measuring self-motivation was .84 and r_{kk} was .94.

The exercises measuring tolerance for stress have $\bar{r}_{cc}$ estimates of .59 and .75 while the r_{kk} estimates are .81 and .90.

The lower variabilities (standard deviations) of the scores for the exercises measuring human relations probably account for the lower $\bar{r}_{cc}$ and r_{kk} estimates. For this reason the standard error of measurement (SEM) was calculated since such indices are independent of the variability of the group on which the reliability coefficient is computed. The $\bar{r}_{cc}$ estimates range from .35 to .68, while the r_{kk} estimates range from .61 to .87. The F-ratio

Table XI.

Intraclass Correlation Coefficients for Integration Ratings (Pure Clinical Method of Assessment) and Exercise Ratings (Trait Ratings Method of Assessment).

(No. of Assessors: 3; No. of Assessees: 108)

Exercise by Dimension	Mean of Ratings	SD	F-Ratio	$\bar{r}_{cc}$	SEM_{cc}	r_{kk}	SEM_{kk}
1. Communication							
A. Personnel Problem	3.45	1.07	5.21**	0.46	.79	0.72	.57
B. Task Force Exercise	3.33	1.07	2.62	0.47	.78	0.73	.56
C. Individual Case Presentation Oral	3.19	1.11	2.72	0.60	.70	0.82	.47
D. Integration Ratings	3.00	1.03	0.06	0.78	.48	0.92	.29
2. Self-motivation Integration Ratings	2.80	1.15	0.09	0.84	.42	0.94	.28
3. Tolerance for Stress							
A. Individual Case Presentation Oral	2.89	1.10	1.97	0.59	.70	0.81	.48
B. Integration Ratings	2.73	1.06	0.09	0.75	.53	0.90	.34
4. Human Relations							
A. Personnel Problem	3.05	0.94	7.91**	0.43	.71	0.69	.52
B. Task Force Exercise	2.98	0.92	1.44	0.35	.74	0.61	.57
C. Integration Ratings	2.94	0.84	1.40	0.68	.48	0.87	.30
5. Decision Making Integration Ratings	2.69	0.93	1.52	0.73	.48	0.89	.31

*Significance at $p < 0.05$ requires an $F \geq 3.04$.
**Significance at $p < 0.01$ requires an $F \geq 4.71$.

is significant at $p<.01$ for the Personnel Problem exercise.

The estimate of $\bar{r}_{cc}$ for the Integration Ratings exercise measuring decision making is .73 and r_{kk} is .89. Even though the variability of the scores is low, the reliability estimates are acceptable.

For the purpose of this investigation the mean of the three assessors' ratings was used for predicting managerial performance. Thus the r_{kk} statistic is probably a more relevant reliability estimate than the $\bar{r}_{cc}$ since it gives the reliability of the mean of the three assessors' ratings. Except for the Task Force and Personnel Problem exercises measuring human relations, all of the r_{kk} estimates are acceptable.

The reliabilities of the following tests and exercises were not estimated because of the candidates' tightly packed schedule: Assessment Interview; Gordon Personal Profile; Gordon Personal Inventory; Self Description Inventory; Structure of Intellect Model - Expressional Fluency; and the In-basket.

B. Reliability Estimates of Criteria Measures

Coefficients of stability over a period of six weeks were calculated for the criteria measures. The r_{tt}'s range from .31 to .72 (Table XII). For this population it is

Table XII.

Means, SD's, Test-Retest Reliability Coefficients(r_{tt}), and SEM's for the Behavioural Anchored Rating Scale Criteria Measures.
(Time between Testing: 6 weeks)

Statistics	Behavioural Anchored Rating Scale				
	Self-Communication	Motivation	Tolerance for Stress	Human Relations	Decision Making
N	51	51	51	51	51
Means					
Test	1.33	1.32	1.37	1.24	1.27
Retest	1.37	1.28	1.40	1.23	1.33
SD's					
Test	.21	.35	.37	.32	.35
Retest	.30	.31	.33	.26	.25
r_{tt}	.31	.68	.70	.72	.67
SEM's					
Test	.17	.20	.20	.17	.20
Retest	.24	.18	.18	.14	.14

concluded that the Behavioural Anchored Rating Scale measuring the assessment dimension of communication is unreliable (r_{tt} = .31). The reliability estimates of the other criteria measures are low but acceptable.

2. Method Comparisons for the Dimension of Communication.

The results obtained for predicting managerial performance on-the-job for the assessment dimension of communication are presented in sections A to F.

A. Linearity of Regression

Linearity of regression was tested by comparing the Pearson r and the eta for the regression of the predictors on the criterion and vice versa using Guilford's F test for linearity described in Chapter II. From the data in Table XIII only the regressions of the Personnel Problem on the Behavioural Anchored Rating Scale-communication and vice versa are statistically significant ($p<.05$). All of the other regressions are assumed to be linear.

B. Pure Clinical Method of Assessment[1]

The correlation between the Integration Ratings (Pure Clinical Method of Assessment) and the Behavioural Anchored Rating Scale for the dimension of communication is .007 for

---

[1] Cf. Appendix 11 for the Tables referred to in Section B.

Table XIII.

Linearity of Regression for Each Predictor Measure(X) on the Behavioural Anchored Rating Scale Criterion Measure(Y) for the Assessment Dimension of Communication. (N=54 for each sample)

Predictor[a]	Sample	r	η_{xy}	df	$F(Lin)_{xy}$[b]	η_{yx}	df	$F(Lin)_{yx}$[b]
1	A	.007	.249	2,50	1.651NS	.137	4,48	.232NS
	B	.059	.160	2,50	.552NS	.249	2,50	1.560NS
2	A	-.037	.084	2,50	.142NS	.314	4,48	1.286NS
	B	.042	.127	2,50	.362NS	.123	2,50	.336NS
3	A	-.122	.191	2,50	.574NS	.289	4,48	.905NS
	B	.234	.265	2,50	.410NS	.349	2,50	1.823NS
4	A	-.112	.173	2,50	.450NS	.301	4,48	1.036NS
	B	.349	.387	2,50	.829NS	.407	2,50	1.325NS
5	A	-.026	.103	3,49	.171NS	.067	4,48	.040NS
	B	.174	.256	2,50	.928NS	.270	2,50	1.152NS
6	A	.061	.204	3,49	.653NS	.271	4,48	.911NS
	B	-.019	.346	2,50	3.400*	.142	2,50	4.127*
7	A	.040	.158	3,49	.392NS	.286	4,48	1.051NS
	B	.219	.240	2,50	.266NS	.232	2,50	.160NS
8	A	.110	.191	2,50	.646NS	.160	4,48	.171NS
	B	.102	.100	3,49	-.007NS	.195	2,50	.717NS

a Predictor 1: Integration Ratings
Predictor 2: Assessment Interview
Predictor 3: In-basket (Written)
Predictor 4: In-basket (Oral)
Predictor 5: Individual Case Presentation Oral
Predictor 6: Personnel Problem
Predictor 7: Task Force Exercise
Predictor 8: SIM-Expressional Fluency

b *Significance at $p<.05$ for 2,50 df requires an F value of 3.18.
Significance at $p<.01$ for 2,50 df requires an F value of 5.06.
Significance at $p<.05$ for 3,49 df requires an F value of 2.795.
Significance at $p<.05$ for 4,48 df requires an F value of 2.56.

sample A and .059 for sample B (Table XIV). These validity coefficients are not significantly different from a $\bar{r} = 0$ ($p<.05$).

The z value calculated for testing the difference between two correlation coefficients for independent samples is .26 (Table XX). This value is not significant ($p<.05$) and samples A and B are considered as random samples from a common population.

Mosier's double cross-validation design described in Chapter II is the method used to cross-validate the correlation coefficients. The cross-validated coefficient for sample A is .003 and the shrinkage after cross-validation is .004 (Table XX). For sample B the cross-validated r is .036 and the shrinkage is .023 (Table XX). Both of the cross-validated coefficients are not significantly different from a $\bar{r} = 0$ ($p<.05$). The z value of .17 indicates that the cross-validated r's were derived from a common population.

C. Trait Ratings Method of Assessment [2]

The multiple R for the predictor variables and criterion measure (BARS) for the Trait Ratings Method of Assessment is .182 for sample A (Table XV) and .451 for sample B (Table XVI). The multiple R's for both samples are not significantly different from a $\bar{R} = 0$ ($p<.05$).

---

2 Cf. Appendix II for the Tables referred to in Section C.

None of the correlation coefficients of the individual predictors with the criterion measure are significantly different ($p<.05$) from a $\bar{r} = 0$ for sample A (Table XV). There are, however, many significant correlations between the individual predictors: out of a possible 15 correlations between the predictors, 12 are significant at either $p<.05$ or $p<.01$ for sample A.

The In-basket (Oral) is the only individual predictor which correlates significantly with the criterion for sample B (Table XVI) and 14 out of a possible 15 correlations between the individual predictors are significant at either $p<.05$ or $p<.01$.

The cross-validated multiple R for sample A is -.109 (shrinkage = .073) and -.291 for sample B (shrinkage = .16).[3] Both of the cross-validated R's are not significantly different from a $\bar{R} = 0$ ($p<.05$). The un-cross-validated and cross-validated R's for samples A and B originate from a common population (Table XX).

D. Pure Statistical Method of Assessment[4]

The correlation between the predictor variable (Structure of Intellect Model-Expressional Fluency) and the criterion (BARS) for the Pure Statistical Method of

---

3 Cf. Table XX in Appendix 11.

4 Cf. Appendix 11 for the Tables referred to in Section D.

Assessment is .11 and .102, respectively, for samples A and B (Table XVII). These validity coefficients are not statistically significant from a $\bar{r} = 0$ ($p < .05$). The cross-validated r is identical to the un-cross-validated r for sample A and the cross-validated r for sample B is .101 (shrinkage = .001).[5] The z values indicate that both the un-cross-validated and cross-validated r's originate from a common population.

E. Mechanical Composite Method of Assessment[6]

The predictors for the Mechanical Composite Method consist of a combination of all predictors used in the Pure Clinical, Trait Ratings, and Pure Statistical Methods of Assessment. In sample A the multiple R between the predictors and the criterion (BARS) is .229 (Table XVIII) and .505 for sample B (Table XIX). Both of these multiple R's are not significantly different from a $\bar{R} = 0$ ($p < .05$). The cross-validated R for sample A is -.099 (shrinkage = .130) and -.288 for sample B (shrinkage = .217).[7] The cross-validated R's are not significantly different from $\bar{R} = 0$ ($p < .05$). The z values for the un-cross-validated and cross-validated data indicate that the R's originate from a common population.

---

5 Cf. Table XX in Appendix 11.

6 Cf. Appendix 11 for Tables referred to in Section E.

7 Cf. Table XX in Appendix 11.

Twenty-four of 28 correlations between the predictor variables for sample A (Table XVIII) and also for sample B (Table XIX) are significant at either $p < .05$ or $p < .01$. None of the individual predictors correlate significantly ($p < .05$) with the criterion measure for sample A (Table XVIII). For sample B (Table XIX) the In-basket (Oral) correlates significantly with the criterion ($p < .01$) as was noted earlier in the Trait Ratings Method.

F. Predictor Method Comparisons for the Assessment Dimension of Communication

The statistic (t_{d_r}) which was explained in Chapter II, is used to compare the validity coefficients for each method of assessment. The cross-validated coefficients are used for the Integration Ratings, Trait Ratings, Pure Statistical, and Mechanical Composite Methods. This procedure is used for comparing methods for all of the assessment dimensions.

The intercorrelation matrix of the predicted scores for each method of assessment is presented in Table XXI.[8] These intercorrelations are required for calculating t_{d_r}.

There are no significant differences (two-tail test $p < .05$) between any two of the methods employed to predict managerial performance on-the-job for either sample A or B

---

[8] *Cf.* Appendix 11 for Table XXI.

(Tables XXII and XXIII).[9] Therefore each null hypothesis (1a to 1f)[10] for the dimension of communication is maintained.

3. Method Comparisons for the Dimension of Self-motivation.

The results obtained for predicting managerial performance on-the-job for the assessment dimension of self-motivation are presented in sections A to F.

A. Linearity of Regression

Table XXIV shows that the regression of the predictors on the Behavioural Anchored Rating Scale criterion measure and vice versa are not statistically significant ($p < .05$). Therefore all regressions are assumed to be linear.

B. Pure Clinical Method of Assessment[11]

The Integration Ratings (Pure Clinical Method of Assessment) and the Behavioural Anchored Rating Scale for the dimension of self-motivation correlate .137 and -.101,

---

9 Whenever the validity coeficients are not significantly different from $\bar{R}$ or $r = 0$, statistically it would seem meaningless to compare any of the methods. However, comparisons are made to discover possible trends since this is only a preliminary study.

10 Cf. p. 55.

11 Cf. Appendix 12 for Tables referred to in Section B.

Table XXII.

Statistical Comparisons (t_{dr}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Communication.
(N = 54)

Sample A[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	0.49NS	0.45NS	0.47NS
Trait Ratings		0.004NS	0.04NS
Pure Statistical			0.05NS

1 Significance at $p<.05$ requires a $t \geq 2.00$ (two-tail test).

Table XXIII.

Statistical Comparisons (t_{d_r}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Communication.
(N = 54)

Sample B[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	1.23NS	0.30NS	1.24NS
Trait Ratings		0.85NS	0.01NS
Pure Statistical			0.88NS

1 Significance at $p<.05$ requires a $t \geq 2.00$ (two-tail test).

Table XXIV.

Linearity of Regression for Each Predictor Measure(X) on the Behavioural Anchored Rating Scale Criterion Measure(Y) for the Assessment Dimension of Self-motivation. (N=54 for each sample)

Predictor[a]	Sample	r	η_{xy}	df	$F(Lin)_{xy}$[b]	η_{yx}	df	$F(Lin)_{yx}$[b]
1	A	.137	.152	2,50	.108NS	.265	3,49	.900NS
	B	-.101	.141	3,49	.163NS	.342	4,48	1.451NS
2	A	.035	.176	3,49	.502NS	.298	3,49	1.570NS
	B	-.094	.155	2,50	.388NS	.344	4,48	1.485NS
3	A	.121	.266	2,50	1.517NS	.262	3,49	.954NS
	B	.121	.353	4,48	1.513NS	.390	4,48	1.927NS

a Predictor 1: Integration Ratings
Predictor 2: Assessment Interview
Predictor 3: Self Description Inventory-Achievement Motivation.

b Significance at $p<.05$ for 2,50 df requires an F value of 3.18.
Significance at $p<.05$ for 3,49 df requires an F value of 2.795.
Significance at $p<.05$ for 4,48 df requires an F value of 2.56.

respectively, for samples A and B (Table XXV). The cross-validated r for sample A is -.135 (shrinkage = .002) and -.100 (shrinkage = .001) for sample B (Table XXX). None of the validity coefficients are significantly different from a $\bar{r} = 0$ ($p<.05$). The z values for the un-cross-validated and cross-validated data indicate that the r's come from a common population (Table XXX).

C. Trait Ratings Method of Assessment[12]

The correlation coefficients for the predictor variable (Assessment Interview) and the criterion measure (BARS) for the Trait Ratings Method measuring self-motivation are .035 for sample A and -.094 for sample B (Table XXVI). The cross-validated r's are identical to the un-cross-validated r's and none of the validity coefficients calculated for the un-cross-validated or cross-validated samples are significant at p .05 (Table XXX). Neither of the z values calculated for the un-cross-validated or cross-validated samples are significant ($p<.05$) and it is assumed that the r's originate from a common population (Table XXX).

D. Pure Statistical Method of Assessment[13]

The scale on the Self Description Inventory measuring self-motivation correlates .121 with the criterion (BARS)

---

12 Cf. Appendix 12 for Tables referred to in Section C.

13 Cf. Appendix 12 for Tables referred to in Section D.

for both samples (Table XXVII). The cross-validated r is identical to the original for sample A and is .119 (shrinkage = .003) for sample B (Table XXX). All r's are not significantly different from a $\bar{r} = 0$ ($p<.05$). The z values, which test the difference between two correlation coefficients for independent samples, indicate that the un-cross-validated and cross-validated samples come from a common population (Table XXX).

E. Mechanical Composite Method of Assessment[14]

The multiple R for the predictor variables and the criterion measure (BARS) for the Mechanical Composite Method of Assessment is .188 for samples A and B (Tables XXVIII and XXIX). The cross-validated R's (Table XXX) are, respectively, .021 (shrinkage = .167) and .004 (shrinkage = .184). None of the multiple R's can be taken as genuine multiple correlation coefficients. There is no significant difference between the un-cross-validated or cross-validated z values at $p<.05$ (Table XXX). Only the Integration Ratings and Assessment Interview predictor variables for both samples are significantly intercorrelated at $p<.01$ (Tables XXVIII and XXIX). None of the individual predictors correlate significantly ($p<.05$) with the criterion measure (BARS).

---

14 Cf. Appendix 12 for Tables referred to in Section E.

F. Predictor Method Comparisons for the Assessment Dimension of Self-motivation

The intercorrelation of the predicted scores are presented in Table XXXI.[15] The t_{d_r} values for samples A and B (Tables XXXII and XXXIII) indicate that there are no significant differences (two-tail test $p < .05$) between any two of the methods employed to predict managerial performance. Therefore each null hypothesis (2a to 2f)[16] is maintained for the assessment dimension of self-motivation.

4. Method Comparisons for the Dimension of Tolerance for Stress.

Sections A to F present the results obtained for predicting managerial performance on-the-job for the assessment dimension of tolerance for stress.

A. Linearity of Regression

The regression of the predictors on the Behavioural Anchored Rating Scale criterion measure and vice versa (Table XXXIV) are not statistically significant ($p < .05$). Therefore all regressions are assumed to be linear.

---

15 Cf. Appendix 12 for Table XXXI.

16 Cf. p. 57.

Table XXXII.

Statistical Comparisons (td_r) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Self-motivation.
(N = 54)

Sample A[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	0.41NS	0.07NS	0.46NS
Trait Ratings		0.42NS	0.06NS
Clinical Composite			0.39NS

1 Significance at $p<.05$ requires a $t \geq 2.00$ (two-tail test).

Table XXXIII.

Statistical Comparisons (t_{dr}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Self-motivation.
(N = 54)

Sample B[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	0.02NS	0.08NS	0.36NS
Trait Ratings		0.11NS	0.38NS
Clinical Composite			0.45NS

1 Significance at $p < .05$ requires a $t \geq 2.00$ (two-tail test).

Table XXXIV.

Linearity of Regression for Each Predictor Measure(X) on the Behavioural Anchored Rating Scale Criterion Measure(Y) for the Assessment Dimension of Tolerance for Stress. (N=54 for each sample)

Predictor[a]	Sample	r	η_{xy}	df	$F(Lin)_{xy}$[b]	η_{yx}	df	$F(Lin)_{yx}$[b]
1	A	-.006	.055	2,50	.075NS	.243	3,49	1.024NS
	B	-.032	.136	2,50	.458NS	.191	3,49	.593NS
2	A	.082	.333	2,50	2.932NS	.258	3,49	1.055NS
	B	-.014	.106	2,50	.273NS	.271	3,49	1.302NS
3	A	.036	.243	2,50	1.533NS	.301	3,49	1.612NS
	B	-.007	.074	2,50	.124NS	.277	3,49	1.362NS
4	A	-.067	.086	2,50	.063NS	.177	3,49	.447NS
	B	.270	.285	3,49	.144NS	.284	3,49	.144NS

a Predictor 1: Integration Ratings
Predictor 2: Assessment Interview
Predictor 3: Individual Case Presentation Oral
Predictor 4: Gordon Personal Profile-Emotional Stability

b Significance at $p<.05$ for 2,50 df requires an F value of 3.18.
Significance at $p<.05$ for 3,49 df requires an F value of 2.795.

B. Pure Clinical Method of Assessment [17]

The correlations between the Integration Ratings (Pure Clinical Method of Assessment) and the Behavioural Anchored Rating Scale for the dimension of tolerance for stress are -.0006 for sample A and -.032 for sample B (Table XXXV). The cross-validated r is .0006 (shrinkage =0) for sample A and -.004 (shrinkage = .028) for sample B (Table XLI). All r's are not considered to be genuine validity coefficients. The z values for the un-cross-validated or cross-validated samples are not significant at $p<.05$ (Table XLI).

C. Trait Ratings Method of Assessment [18]

The multiple R for the predictor variables and the criterion measure (BARS) for the Trait Ratings Method of Assessment is .082 and .014, respectively, for samples A and B (Tables XXXVI and XXXVII). The cross-validated R for sample A is -.081 (shrinkage = .001) and -.013 (shrinkage = .001) for sample B (Table XLI). The multiple R's can not be considered genuine validity coefficients. The z values for the un-cross-validated and cross-validated samples indicate that the multiple R's come from a common population (Table XLI). Tables XXXVI and XXXVII show that two predictors, the Assessment Interview and Individual Case Presentation Oral, are significantly correlated at $p<.01$. The

---

17 Cf. Appendix 13 for Tables referred to in Section B.

18 Cf. Appendix 13 for Tables referred to in Section C.

individual predictors do not correlate significantly with the criterion measure (BARS).

D. Pure Statistical Method of Assessment[19]

The Pure Statistical Method of Assessment measuring tolerance for stress (Gordon Personal Profile) correlates -.067 and .270, respectively, for samples A and B with the BARS criterion measure (Table XXXVIII). The cross-validated r's are identical to the un-cross-validated r's (Table XLI). The un-cross-validated and cross-validated r's for sample A are not significantly different from a $\bar{r} = 0$ ($p<.05$) but both the un-cross-validated and cross-validated r's for sample B are significant at $p<.05$ and are taken as genuine correlation coefficients (Table XLI). All of the z values calculated for the un-cross-validated and cross-validated samples are not significant ($p<.05$) and the r's are considered as coming from a common population (Table XLI).

E. Mechanical Composite Method of Assessment[20]

The multiple predictors of the Mechanical Composite Method of Assessment correlate .111 for sample A (Table XXXIX) and .289 for sample B (Table XL) with the criterion measure. After cross-validation the multiple R's were, respectively, -.049 (shrinkage = .062) and -.094 (shrinkage = .195).[21] All multiple R's are not significantly different from $\bar{R} = 0$

---

19 Cf. Appendix 13 for Tables referred to in Section D.
20 Cf. Appendix 13 for Tables referred to in Section E.
21 Cf. Appendix 13, Table XLI.

($p<.05$). The z values for the un-cross-validated or cross-validated samples are not significant at $p<.01$ (Table XXXVII). Three of the correlations between the predictors are significant at $p<.01$ for sample A (Table XXXIX) and four are significant for sample B at $p<.05$ or $p<.01$ (Table XL). Only one predictor in sample B, the Gordon Personal Profile, correlates ($p<.05$) with the criterion measure (Table XL). None are significant for sample A (Table XXXIX).

F. Predictor Method Comparisons for the Assessment Dimension of Tolerance for Stress

Table XLII[22] presents the intercorrelations of the predicted scores for each method of assessment for the assessment dimension of tolerance for stress. Tables XLIII and XLIV indicate that there are no significant differences (two-tail test $p<.05$) between any two of the methods predicting managerial performance on-the-job for either sample A or B. Therefore the null hypotheses (3a to 3f)[23] are maintained for tolerance for stress.

5. Method Comparisons for the Dimension of Human Relations

The results obtained for predicting managerial performance on-the-job for the dimension of human relations are presented in sections A to F.

A. Linearity of Regression

The obtained F values using Guilford's test for linearity are not significant ($p<.05$) for the regression of the predictors on the BARS criterion measure for the

---

22 Cf. Appendix 13 for Table XLII.

23 Cf. p. 57.

Table XLIII.

Statistical Comparisons (t_{d_r}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Tolerance for Stress.
(N = 54)

Sample A[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	0.34NS	0.31NS	0.22NS
Trait Ratings		0.07NS	0.16NS
Pure Statistical			0.07NS

1 Significance at $p < .05$ requires a $t \geq 2.00$ (two-tail test).

Table XLIV.

Statistical Comparisons (t_{dr}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Tolerance for Stress.
(N = 54)

Sample B[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	0.04NS	1.33NS	0.42NS
Trait Ratings		1.29NS	0.31NS
Pure Statistical			0.76NS

1 Significance at $p<.05$ requires a $t \geq 2.00$ (two-tail test).

dimension of human relations (Table XLV). However, the obtained F values for the regression of the criterion measure on the predictors indicate that the Pure Clinical Method of Assessment (Integration Ratings) for sample B is significant ($p<.05$) and is assumed to be nonlinear (Table XLV). It can be safely assumed that the other regressions are linear.

B. Pure Clinical Method of Assessment[24]

The Integration Ratings (Pure Clinical Method of Assessment) and the Behavioural Anchored Rating Scale for the dimension of human relations correlate .093 and .049, respectively, for samples A and B (Table XLVI). The cross-validated r's are .088 (shrinkage = .005) for sample A and .049 (shrinkage = 0) for sample B (Table LII). The r's do not differ significantly ($p<.05$) from a $\bar{r} = 0$. The z values indicate that the un-cross-validated and cross-validated r's originate from a common population (Table LII).

C. Trait Ratings Method of Assessment[25]

The multiple R for the predictor variables (Trait Ratings Method of Assessment) and the BARS criterion measure is .284 for sample A and .344 for sample B (Tables XLVII and XLVIII). The cross-validated R's are, respectively, .131 (shrinkage = .153) and .196 (shrinkage = .148) for

---

24 Cf. Appendix 14 for Tables referred to in Section B.
25 Cf. Appendix 14 for Tables referred to in Section C.

Table XLV.

Linearity of Regression for Each Predictor Measure(X) on the Behavioural Anchored Rating Scale Criterion Measure(Y) for the Assessment Dimension of Human Relations. (N-54 for each sample)

Predictor[a]	Sample	r	η_{xy}	df	$F(Lin)_{xy}$[b]	η_{yx}	df	$F(Lin)_{yx}$[b]
1	A	.093	.183	2,50	.656NS	.176	3,49	.377NS
	B	.049	.349	2,50	3.405*	.229	3,49	.873NS
2	A	.186	.288	2,50	1.320NS	.236	3,49	.370NS
	B	.301	.330	1,51	1.053NS	.431	3,49	1.914NS
3	A	.140	.284	2,50	1.670NS	.168	3,49	.141NS
	B	.018	.048	1,51	.086NS	.208	3,49	.728NS
4	A	.261	.272	3,49	.104NS	.272	3,49	.104NS
	B	.116	.237	2,50	1.127NS	.264	3,49	.993NS
5	A	-.002	.290	3,49	1.500NS	.225	3,49	.878NS
	B	.063	.149	2,50	.461NS	.155	3,49	.335NS

a Predictor 1: Integration Ratings
Predictor 2: Assessment Interview
Predictor 3: Personnel Problem
Predictor 4: Task Force Exercise
Predictor 5: Gordon Personal Inventory-Personal Relations

b *Significance at $p<.05$ for 2,50 df requires an F value of 3.18.
Significance at $p<.01$ for 2,50 df requires an F value of 5.06.
Significance at $p<.05$ for 3,49 df requires an F value of 2.795.
Significance at $p<.05$ for 1,50 df requires an F value of 4.03.

samples A and B (Table LII). The multiple R's are not assumed to be genuine since they do not differ significantly from a $\bar{R} = 0$ at $p<.05$.

None of the correlations of the individual predictors with the criterion measure are significantly different from $\bar{r} = 0$ at $p<.05$ for sample A (Table XLVII). The only predictor which correlates significantly ($p<.05$) with the criterion measure for sample B is the Assessment Interview (Table XLVIII). Two of the correlations between the predictors are significant ($p<.05$ and $p<.01$) for sample A and three are significant ($p<.01$) for sample B (Tables XLVII and XLVIII).

From the data presented in Table LII it can be assumed that the un-cross-validated and cross-validated R's for samples A and B come from a common population.

D. Pure Statistical Method of Assessment[26]

The correlation between the Gordon Personal Inventory -Personal Relations (Pure Statistical Method of Assessment) and the Behavioural Anchored Rating Scale for the dimension of human relations is -.002 for sample A and .063 for sample B (Table XLIX). These validity coefficients are not significantly different from a $\bar{r} = 0$ ($p<.05$). The cross-validated r's are .002 (shrinkage = 0) for sample A and -.062 (shrinkage = .001) for sample B (Table LII). The

---

26 Cf. Appendix 14 for Tables referred to in Section D.

z values for both the un-cross-validated and cross-validated r's are not significant at $p<.05$ (Table LII). Therefore, it is concluded that the un-cross-validated and cross-validated r's originate from a common population.

E. Mechanical Method of Assessment

The multiple R for the predictor variables and the criterion measure (BARS) for the Mechanical Composite Method of Assessment is .322 for sample A and .346 for sample B (Tables L and LI). The cross-validated R's are, respectively, .148 (shrinkage = .174) and .239 (shrinkage = .107).[27] The z values are not significant at $p<.05$ for either the un-cross-validated or cross-validated samples and it can be concluded that the multiple R's come from a common population.

None of the correlations between the individual predictors and the criterion measure are significantly different ($p<.05$) from a $\bar{r} = 0$ for sample A (Table L). Only one predictor, the Assessment Interview, correlates significantly ($p<.05$) with the criterion for sample B (Table LI). Five out of ten correlations between the predictors are significant at either $p<.05$ or $p<.01$ for sample A (Table L), while six out of ten are significant at $p<.01$ for sample B (Table LI).

---

27 Cf. Appendix 14, Table LII.

F. Predictor Method Comparisons for the Assessment Dimension of Human Relations

The intercorrelations of the predicted scores for the assessment methods are presented in Table LIII for samples A and B.[28] The t_{d_r} values for samples A and B (Tables LIV and LV) show that there are no significant differences (two-tail test $p<.05$) between any two of the methods predicting managerial performance on-the-job for the assessment dimension of human relations for either sample A or B. Therefore the null hypotheses **(4a to 4f)**[29] **are maintained.**

6. Method Comparisons for the Dimension of Decision Msking.

The results obtained for predicting managerial performance on-the-job for the dimension of decision making are presented in sections A to F.

A. Linearity of Regression

The regression of the predictors on the BARS criterion measure and vice versa (Table LVI) for the dimension of decision making are assumed to be linear since the obtained F values are not significant ($p<.05$).

---

28 Cf. Appendix 14 for Table LIII.

29 Cf. p. 57.

Table LIV.

Statistical Comparisons (t_{d_r}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Human Relations.
(N = 54)

Sample A[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	0.18NS	0.42NS	0.26NS
Trait Ratings		0.59NS	0.06NS
Pure Statistical			0.66NS

1 Significance at $p<.05$ requires a t = 2.00 (two-tail test).

Table LV.

Statistical Comparisons (t_{dr}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Human Relations.
(N = 54)

Sample B[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	0.60NS	0.07NS	0.98NS
Trait Ratings		0.67NS	0.17NS
Pure Statistical			0.92NS

1 Significance at $p<.05$ requires a $t \geq 2.00$ (two-tail test).

Table LVI.

Linearity of Regression for Each Predictor Measure(X) on the Behavioural Anchored Rating Scale Criterion Measure(Y) for the Assessment Dimension of Decision Making. (N=54 for each sample)

Predictor[a]	Sample	r	η_{xy}	df	$F(Lin)_{xy}$[b]	η_{yx}	df	$F(Lin)_{yx}$[b]
1	A	.217	.219	2,50	.024NS	.312	4,48	.677NS
	B	-.022	.177	2,50	.787NS	.329	5,47	1.145NS
2	A	.334	.409	2,50	1.664NS	.364	4,48	.283NS
	B	.274	.331	3,49	.641NS	.363	5,47	.616NS
3	A	-.158	.239	2,50	.849NS	.330	4,48	1.132NS
	B	-.048	.277	2,50	2.023NS	.265	5,47	.684NS
4	A	.052	.171	3,49	.442NS	.200	4,48	.466NS
	B	-.122	.285	2,50	1.799NS	.286	5,47	.687NS
5	A	-.042	.119	3,49	.204NS	.350	4,48	1.643NS
	B	-.008	.260	3,49	1.191NS	.332	5,47	1.161NS

a Predictor 1: Integration Ratings
Predictor 2: In-basket (Written)
Predictor 3: Self Description Inventory-Decisiveness
Predictor 4: Gordon Personal Profile-Ascendancy
Predictor 5: Gordon Personal Inventory-Cautiousness

b Significance at $p<.05$ for 2,50 df requires an F value of 3.18.
Significance at $p<.05$ for 3,49 df requires an F value of 2.795.
Significance at $p<.05$ for 4,48 df requires an F value of 2.56.
Significance at $p<.05$ for 5,47 df requires an F value of 2.415.

B. Pure Clinical Method of Assessment[30]

The Integration Ratings (Pure Clinical Method of Assessment) and the Behavioural Anchored Rating Scale for the dimension of decision making correlate .217 and -.022, respectively, for samples A and B (Table LVII). The cross-validated r's are -.176 (shrinkage = .041) for sample A and -.022 (shrinkage = 0) for sample B (Table LXIII). All r's are not significantly different ($p<.05$) from a $\bar{r} = 0$. Table LXIII indicates that the un-cross-validated and cross-validated r's come from a common population since the z values are not significant ($p<.05$).

C. Trait Ratings Method of Assessment[31]

The correlations between the In-basket (Written) (Trait Ratings Method of Assessment) and the Behavioural Anchored Rating Scale for the dimension of decision making correlate .334 for sample A and .274 for sample B (Table LVIII). Both of these r's are significantly different from a $\bar{r} = 0$ ($p<.05$). The cross-validated r's are identical to the un-cross-validated ones for both samples (Table LXIII). These r's are significant ($p<.05$) and are considered to be genuine correlations. Both r's for the un-cross-validated and cross-validated data come from a common population (Table LXIII).

---

30 Cf. Appendix 15 for Tables referred to in Section B.
31 Cf. Appendix 15 for Tables referred to in Section C.

D. Pure Statistical Method of Assessment[32]

The multiple R between the predictor variables (Pure Statistical Method of Assessment) and the criterion measure (BARS) for the dimension of decision making is .199 and .144, respectively, for samples A and B (Tables LIX and LX). The R's for both samples are not significantly different from a $\bar{R} = 0$ at $p<.05$. The correlation coefficients between the individual predictors and the criterion measure are not statistically different from $\bar{r} = 0$ at $p<.05$. Only one of the correlations between the predictors is significant ($p<.05$) for sample A and two are significant ($p<.01$) for sample B.

The cross-validated R for sample A is -.013 (shrinkage = .186) and -.025 (shrinkage = .119) for sample B (Table LXIII). Both the un-cross-validated and cross-validated R's originate from a common population since the z values are not significant at $p<.05$ (Table LXIII).

E. Mechanical Composite Method of Assessment[33]

The multiple R for the predictor variables and the criterion measure (BARS) for the Pure Statistical Method of Assessment is .371 for sample A and .387 for sample B

---

32 Cf. Appendix 15 for Tables referred to in Section D.

33 Cf. Appendix 15 for Tables referred to in Section E.

(Tables LXI and LXII). The multiple R's for both samples are not significantly different from a $\bar{R} = 0$ at $p<.05$.

The correlation coefficients of the individual predictors with the criterion measure are not significantly different ($p<.05$) from a $\bar{r} = 0$ for sample A (Table LXI). One is significant ($p<.05$) for sample B (Table LXII). There are significant correlations between the individual predictors: 4 out of 10 are significant at either $p<.05$ or $p<.01$ for sample A (Table LXI); and, 7 out of 10 are significant at either $p<.05$ or $p<.01$ for sample B (Table LXII).

The cross-validated R for sample A is .144 (shrinkage = .227) and .214 for sample B (shrinkage = .173).[34] These are not significantly different from a $\bar{R} = 0$ at $p<.05$ (Table LXIII). The obtained z values from the data presented in Table LXIII indicate that the un-cross-validated and cross-validated R's come from a common population.

F. Predictor Method Comparisons for the Assessment Dimension of Decision Making

The intercorrelations of the predicted scores for the methods of assessment are presented in Table LXIV.[35] The data from Tables LXV and LXVI shows that there are no significant differences (two-tail test $p<.05$) between any

---

34 Cf. Table LXIII in Appendix 15.

35 Cf. Appendix 15 for Table LXIV.

two of the methods employed to predict managerial performance on-the-job for either sample A or B. Therefore the null hypotheses (5a to 5f)[36] **for the assessment dimension of decision making are maintained.**

---

36 Cf.p. 58.

Table LXV.

Statistical Comparisons (t_{dr}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Decision Making.
(N = 54)

Sample A[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	0.69NS	0.78NS	0.14NS
Trait Ratings		1.63NS	0.82NS
Pure Statistical			0.59NS

1 Significance at $p<.05$ requires a $t \geq 2.00$ (two-tail test).

Table LXVI.

Statistical Comparisons (t_{d_r}) of the Validities for Predictor Methods of Managerial Performance for the Dimension of Decision Making.
(N = 54)

Sample B[1]

Method	Trait Ratings	Pure Statistical	Mechanical Composite
Pure Clinical	1.06NS	0.01NS	0.77NS
Trait Ratings		1.26NS	0.23NS
Pure Statistical			0.85NS

1 Significance at $p<.05$ requires a $t \geq 2.00$ (two-tail test).

SUMMARY AND CONCLUSIONS

This summary of research findings in the context of the Canadian Public Service Commission Managerial Assessment Centre Program supports the following conclusions:

1. For the assessment dimension of communication, no significant difference was found between any two of the following methods of predicting managerial performance in either sample A or B: Pure Clinical, Trait Ratings, Pure Statistical, and Mechanical Composite. However, the unreliability of the criterion measure (BARS) for the dimension of communication probably reduced the possibility of finding significant differences between the methods.

2. For the same methods of assessment measuring the dimensions of self-motivation, tolerance for stress, and decision making, no significant difference between any two of the methods was found for either sample A or B. The reliabilities of the predictors were adequate. The possibility does exist that the reliabilities of the predictors, which were not checked for reliability, could have been low enough to reduce the validity coefficients and, hence, reduced the possibility of finding significant differences between the methods of prediction.

3. There was also no significant difference between the predictor methods for the dimension of human relations.

However, the reliability of one of the predictors was inadequate and not all of the predictors were checked for reliability. The reliability of the criterion measure was adequate.

In general, then, there is no evidence that the methods employed for predicting managerial performance are statistically significant from one another. These findings of course apply to the type of population undergoing assessment for each of the five dimensions in a particular Canadian federal government department. Since the population participating in this research is heterogeneous with regard to age, education, job related experience, areas lived in, personal interests, and sincethe standards of pre-selection were limited to geographical mobility and meeting the Public Service classification standards, it might well be that the findings do have some generalizability. However, the writer cautions any generalization of these results for the following reasons:

1. Although most of the predictors that were checked for reliability were reliable, not all of the predictors were checked.

2. The on-the-job performance ratings were taken only six months after the assessees were placed in federal government positions. A longer period of time probably would have given the assessees more time to demonstrate

their abilities and, at the same time, the supervisors would have acquired a greater knowledge of the assessee's on-the-job performance. The on-the-job performance ratings were obtained from each candidate's supervisor. This could also constitute a source of error since each of the supervisors could have rated performance from different perspectives.

3. The constructs which were measured at the Assessment Centre may have been different from the constructs measured on the criterion measure because the behavioural anchors on the predictor scales differed from the behavioural anchors on the criteria measures. The behavioural examples for the predictors were attached to five or six numerically scaled points and were specifically developed and related to each exercise. On the other hand, the behavioural examples on the criteria measures (BARS) were attached to a nine-point numerical scale and the examples were developed and expressed in terms of specific on-the-job behaviours. This difference could have reduced the size of the validity coefficients and could have accounted for the fact that only three of the validity coefficients (cross-validated) were significantly different from a $\overline{R}$ or $\overline{r} = 0$ at $p < .05$.

4. Since the t_{d_r} statistic was used to compare the methods of prediction, generalization is limited to a subpopulation of all possible samples for which X and Y have

exactly the same set of values as those in the observed sample.

Some interesting extensions of this investigation could involve the comparison of assessment methods employing identical behavioural anchors and numerical scales for both the predictor and criterion scales. It further suggests that the claim for superior reliability of the Behavioural Anchored Rating Scale by Peters and McCormick and other authors must be investigated further for the assessment of managerial performance.[1]

Another design could be employed using the data from this investigation, namely, the Campbell and Fiske **convergent** and discriminant validation by the multitrait-multimethod method approach.

In view of the fact that four summary statements were previously made by the writer in which he cautioned generalizing the results, this investigation suggests that many of the claims made by Meehl, Sawyer and Gough for the superiority of mechanical methods over clinical methods for collecting and/or combining data might not be upheld when a cross-validated design is employed, a statistical test is used to compare

---

1 D. L. Peters and E. J. McCormick, "Comparative Reliability of Numerically Anchored Versus Job-Task Anchored Rating Scales," Journal of Applied Psychology, Vol. 50, 1966, p. 92-96.

the methods of assessment, and the same sample of subjects is compared for the methods of assessment.

The writer wishes to point out that all of the validity coefficients obtained in this study do not reflect the validity of the total decision-making process used in this assessment centre since the ratings obtained from the methods of assessment used in this study are incorporated into a global rating of managerial potential (yes, no, or uncertain).

BIBLIOGRAPHY

Campbell, J. P., M. D. Dunnette, E. E. Lawler, and K. E. Weick, Jr., Managerial Behavior, Performance and Effectiveness, New York, McGraw-Hill, 1970, 1-546 p.

Managerial Behavior, Performance and Effectiveness examines current practices in government and industry used in identifying effective managers.

Cohen, B. M., J. L. Moses, and W. C. Byham, "The Validity of Assessment Centers," an article which is to be published in Journal of Industrial and Organizational Psychology, 1-42 p.

A review of all validity studies concerning the assessment centre since 1956 revealed that assessments were valid and were consistently related to several criteria.

Ferguson, G. A., Statistical Analysis in Psychology and Education (3rd ed.), New York, McGraw-Hill, 1971, xii-492 p.

A source book for statistical procedures.

Gordon, L., Gordon Personal Inventory, New York, Harcourt, Brace & World, 1963, 2-20 p.

A description of the personality traits measured by this inventory.

--------, Gordon Personal Profile, New York, Harcourt, Brace & World, 1963, 2-27 p.

A description of personality traits measured by this test.

Gough, H. G., "Clinical versus Statistical Prediction in Psychology," in L. Postman (ed.), Psychology in the Making, New York, Knopf, 1963, p. 526-584.

A lengthy summary of the clinical versus statistical prediction controversy.

Grant, K. W., and L. W. Slivinski, Identification of Assessment Factors in an Assessment Centre, Public Service Commission, Ottawa, 1973, 1-100 p.

A complete summary of the development of assessment dimensions and the formulation of a Behavioural Anchored Rating Scale.

Guilford, J. P., The Nature of Human Intelligence, New York, McGraw-Hill, 1967, 2-538 p.

Description of Guilford's Intellect Model.

Guilford, J. P., Psychometric Methods (2nd ed.), New York, McGraw-Hill, 1954, 1-597 p.
A source book for statistical procedures.

Guilford, J. P., and B. Fruchter, Fundamental Statistics in Psychology and Education (5th ed.), New York, McGraw-Hill, 1973, 1-546 p.
A source book for statistical procedures.

Holt, R. R., "Yet Another Look at Clinical and Statistical Prediction: Or Is Clinical Psychology Worthwhile?" American Psychologist, Vol. 25, 1970, p. 337-349.
A review and critique of Gough's, Meehl's, and Sawyer's empirical surveys on the clinical versus statistical prediction controversy. Very informative.

Huck, J. R., "Assessment Centers: A Review of the External and Internal Validities," Personnel Psychology, Vol. 26, 1973, p. 191-212.
This investigation concludes that research results have been quite impressive in demonstrating both "external" and "internal" validities of multiple assessment procedures but a critical need exists for additional research dealing with the assessment centre approach.

Korman, A. K., "The Prediction of Managerial Performance: A Review," Personnel Psychology, Vol. 21, 1968, p. 295-322.
A review and critical evaluation of the research literature pertaining to the usefulness of various procedures in the prediction of leadership behaviour in formal organizations in a selection context. Very informative.

Leim, M. A., L. W. Slivinski, and K. W. Grant, Identification of Middle Management Job Dimensions and Job Types for Use in an Assessment Centre, Public Service Commission, Ottawa, 1972, 1-55 p.
An investigation which identifies important job dimensions and job types in a Canadian federal government department.

Meehl, P. E., Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence, Minneapolis, University of Minnesota Press, 1954, v-149 p.
A discussion of the issues in the clinical versus statistical prediction controversy and a summary of the empirical evidence. A primary source book.

Meehl, P. E., "Seer over Sign: The First Good Example," Journal of Experimental Research in Personality, Vol. 1, 1965, p. 27-32.

The first and only empirical comparison of the relative efficacy of the clinical and statistical methods showing clear superiority for the clinical judge.

--------, "When Shall We Use Our Heads Instead of the Formula?" Journal of Counseling Psychology, Vol. 4, 1957, p. 268-273.

A further accumulation of empirical evidence supporting his conclusions in 1954 that statistical methods are superior to clinical methods of assessment.

Mosier, C. I., "Symposium: The Need and Means of Cross-validation I. Problems and Designs of Cross-validation," Educational and Psychological Measurement, Vol. 11, 1951, p. 5-11.

A discussion of five distinct designs closely related to cross-validation and a description of the double cross-validational design.

Peters, D. L., and E. J. McCormick, "Comparative Reliability of Numerically Anchored Versus Job-Task Anchored Rating Scales," Journal of Applied Psychology, Vol. 50, 1966, p. 92-96.

It was found that job-task anchored scales could generally be used with significantly greater reliability than simple numerically anchored scales.

Sawyer, J., "Measurement and Prediction," Studies in Personnel Psychology, Vol. 2, 1970, p. 74-105. Also found in Psychological Bulletin, Vol. 66, 1966, p. 178-200.

This investigation suggests that clinical versus statistical prediction is only half the problem--and the last half at that. The first problem is clinical versus statistical measurement. A broader framework is the proposed solution to the problem and an empirical investigation shows the superiority of the mechanical methods.

Slivinski, L. W., and K. W. Grant, Behavioural Validation Prediction Model, paper presented to International Public Personnel Association, San Francisco, 1971, 1-39 p.

The authors present their validation model with the hope that it will lead to research that will lead to greater scientific understanding and assist in developing a meaningful theory of work behaviour.

Smith, P. C., and L. M. Kendall, "Retranslation of Expectations: An Approach to the Construction of Unambiguous Anchors for Rating Scales," Journal of Applied Psychology, Vol. 47, 1963, p. 149-155.

The "retranslation procedure" offers a method of evaluation which is based upon actual observable behaviour, incorporates desirable psychometric characteristics and is acceptable to the people doing the evaluations.

Additions

Bray, D. W., and D. L. Grant, "The Assessment Center in the Measurement of Potential for Business Management," Psychological Monographs, Vol. 80, No. 17 (Whole No. 625), 1966, p. 1-27.

An extensive study of the assessment process by the American Telephone and Telegraph Company and presentation of the results of the Management Progress Study.

Glaser, R., P. A. Schwartz, and J. C. Flanagan, "The Contribution of the Interview and Situational Performance Procedures to the Selection of Supervisory Personnel," Journal of Applied Psychology, Vol. 42, 1958, p. 69-73.

A study which indicates that situational exercises and clinical procedures do not substantially increase the validity of more traditional methods of assessment.

Moses, J. L., "Assessment Center Performance and Management Progress," a paper presented at the 79th annual meeting of the American Psychological Association, Washington, D.C., 1971.

A research paper presenting results on 5,946 individuals assessed by the Bell System Companies. It indicates that judgmental ratings correlate highly with progress in management.

Wollowick, H. B., and W. J. McNamara, "Relationship of the Components of an Assessment Center to Management Success," Journal of Applied Psychology, Vol. 53, 1969, p. 348-352.

A study involving lower and middle managers selected to participate in IBM's assessment program after being "designated as having above-average potential for advancement."

APPENDIX 1

SUMMARY OF SAWYER'S SURVEY OF 45 CLINICAL-STATISTICAL STUDIES

Table II.

Summary of Sawyer's Survey of 45 Clinical-Statistical Studies.[a]

Number of Comparisons Made and Percentage of Comparisons in Which Each Strategy Was Superior, Equal, or Inferior in Predictive Accuracy to the Comparison Strategy

		Percentage of Comparisons		
	No. of Comparisons	Significantly Superior	Equal	Significantly Inferior
Pure clinical	8	0	50	50
Behavior rating	12	8	76	16
Profile interpretation	12	0	75	25
Pure statistical	32	31	69	0
Clinical composite	24	0	63	37
Mechanical composite	10	60	40	0

Ranking of Strategies from "Best" to "Poorest" According to Proportions of Comparisons Yielding Significantly Superior Predictive Accuracies

	Strategy	Mode of Data Collection	Mode of Data Combination	Percentage of "Superior" Outcomes
BEST	Mechanical composite	Clinical and mechanical	Mechanical	60
	Pure statistical	Mechanical	Mechanical	31
	Behaviour rating	Clinical	Mechanical	8
	Profile interpretation	Mechanical	Clinical	0
	Clinical composite	Clinical and mechanical	Clinical	0
POOREST	Pure clinical	Clinical	Clinical	0

Ranking of Strategies from "Best" to "Poorest" According to Proportions of Comparisons Yielding Significantly Inferior Predictive Accuracies

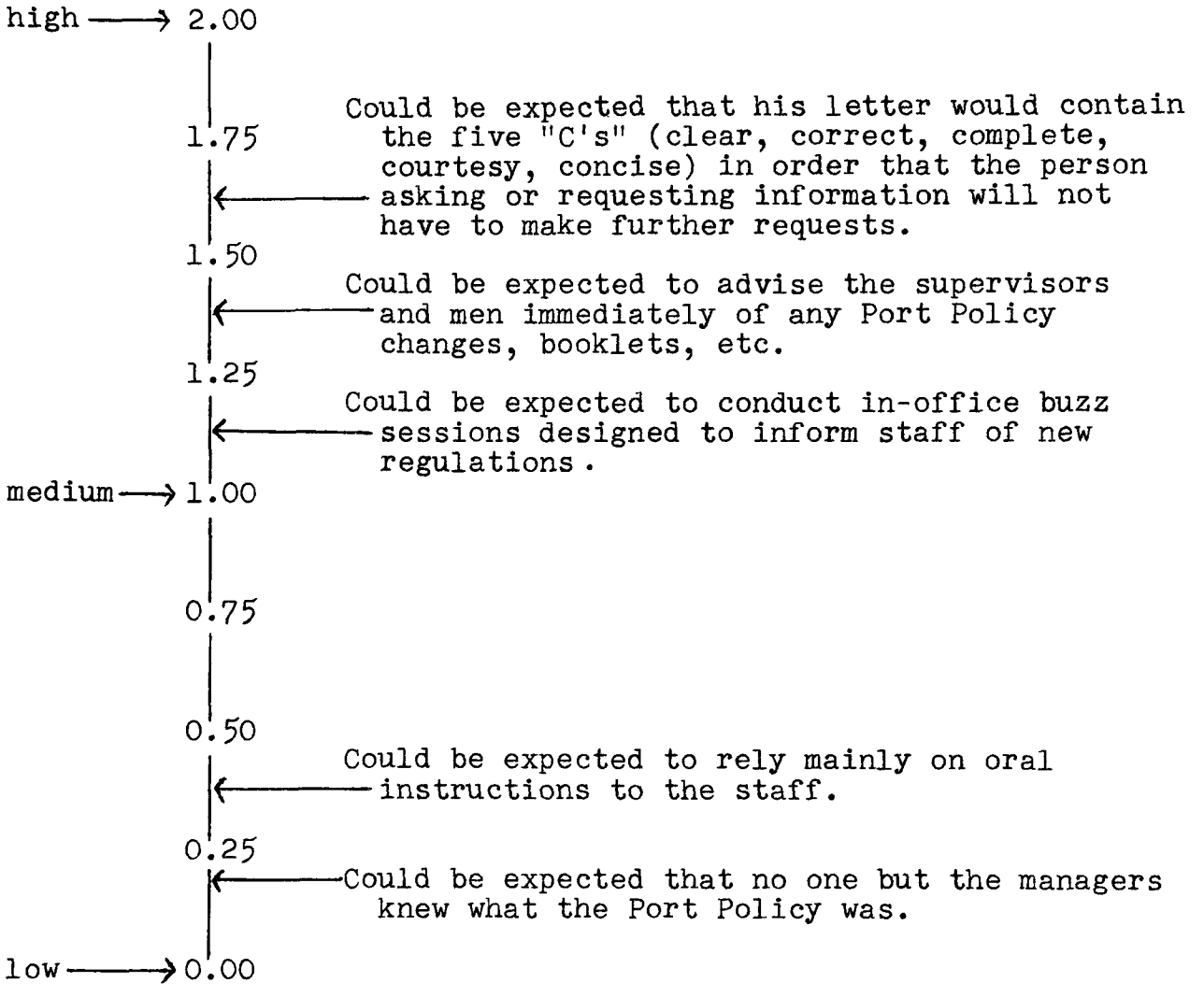
	Strategy	Mode of Data Collection	Mode of Data Combination	Percentage of "Inferior" Outcomes
BEST	Mechanical composite	Clinical and mechanical	Mechanical	0
	Pure statistical	Mechanical	Mechanical	0
	Behavior rating	Clinical	Mechanical	16
	Profile interpretation	Mechanical	Clinical	25
	Clinical composite	Clinical and mechanical	Clinical	37
POOREST	Pure clinical	Clinical	Clinical	50

a J. Campbell et al., Managerial Behavior, Performance and Effectiveness, New York, McGraw-Hill, 1970, p. 155-156.

APPENDIX 2

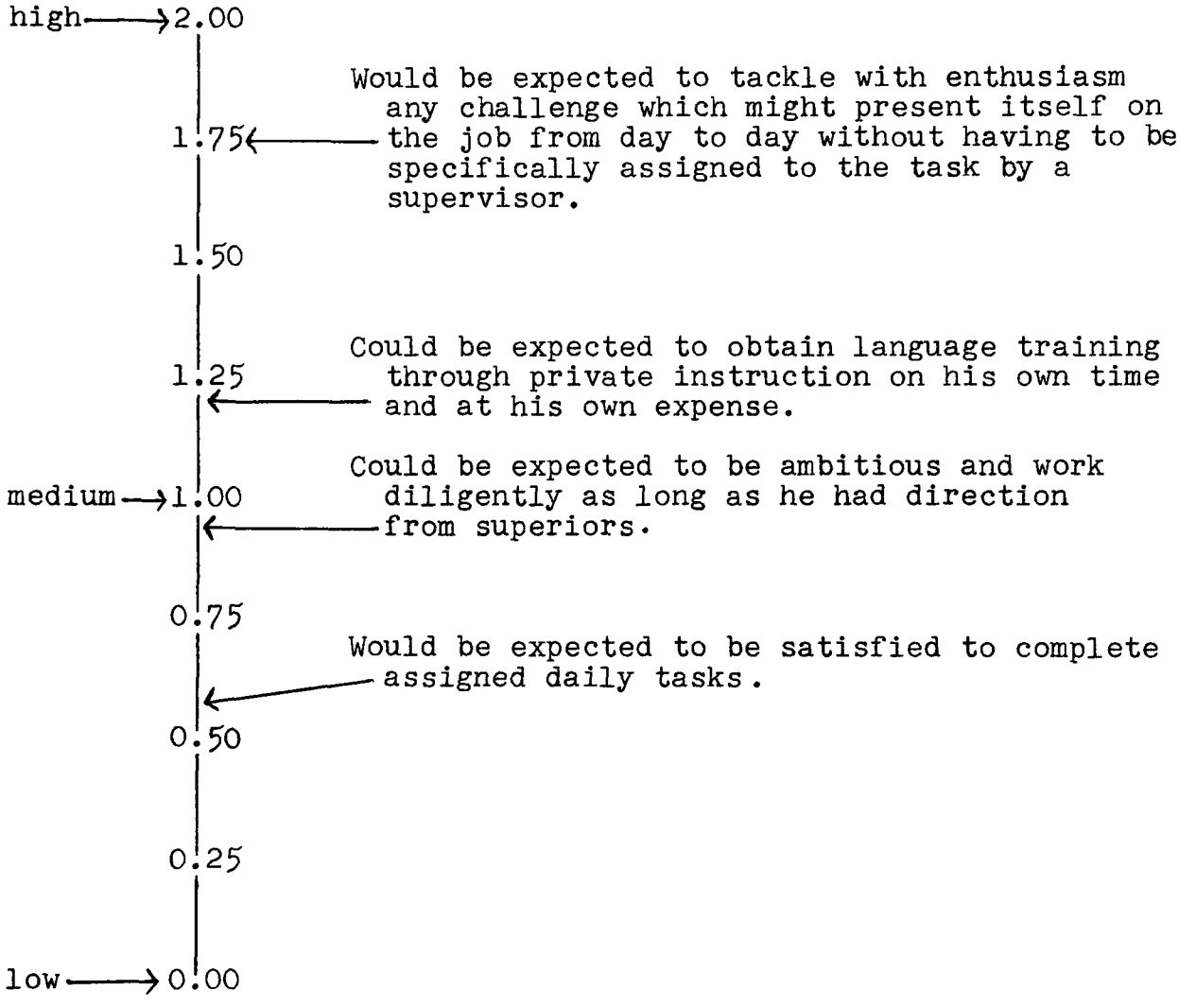
BEHAVIOURAL ANCHORED RATING SCALES

BEHAVIOURAL ANCHORED RATING SCALES



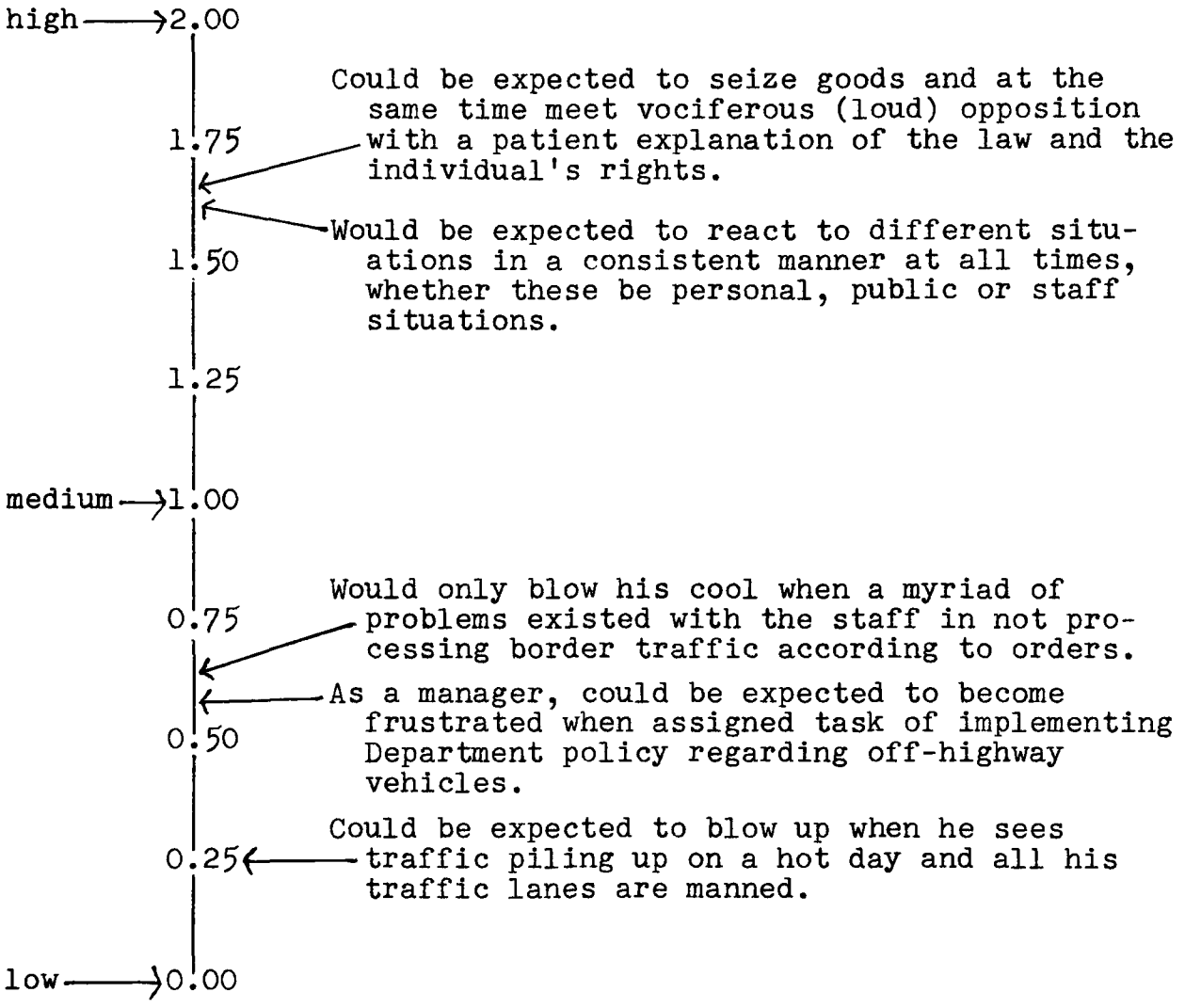
Communication.- The ability to receive and transmit information and ideas in a clear and concise manner.

Figure 1.- Behavioural Anchored Rating Scale for the Dimension of Communication.



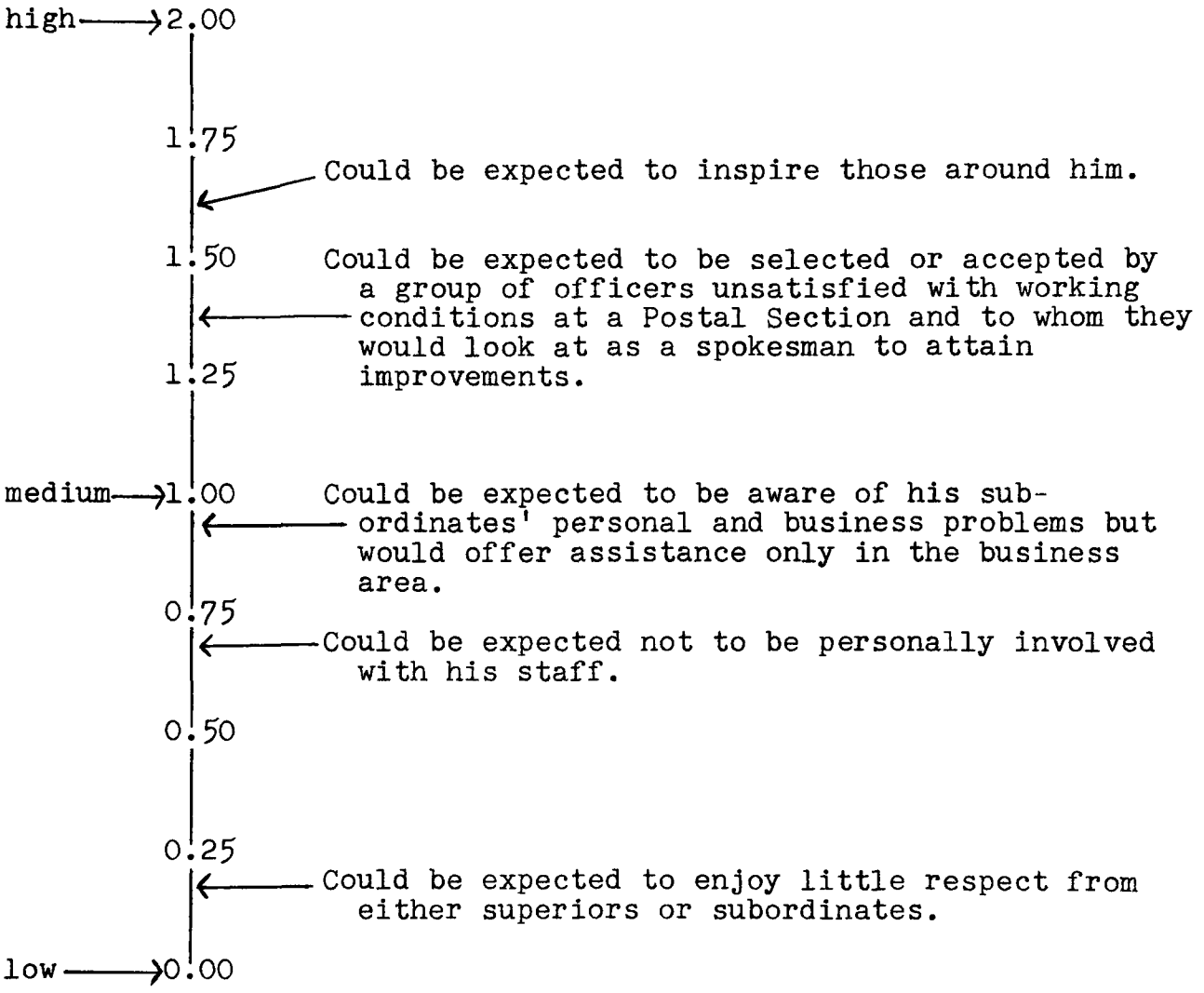
Self-motivation.- The ability to act towards the achievement of a goal, without outside direction or abnormal pressures.

Figure 2.- Behavioural Anchored Rating Scale for the Dimension of Self-motivation.



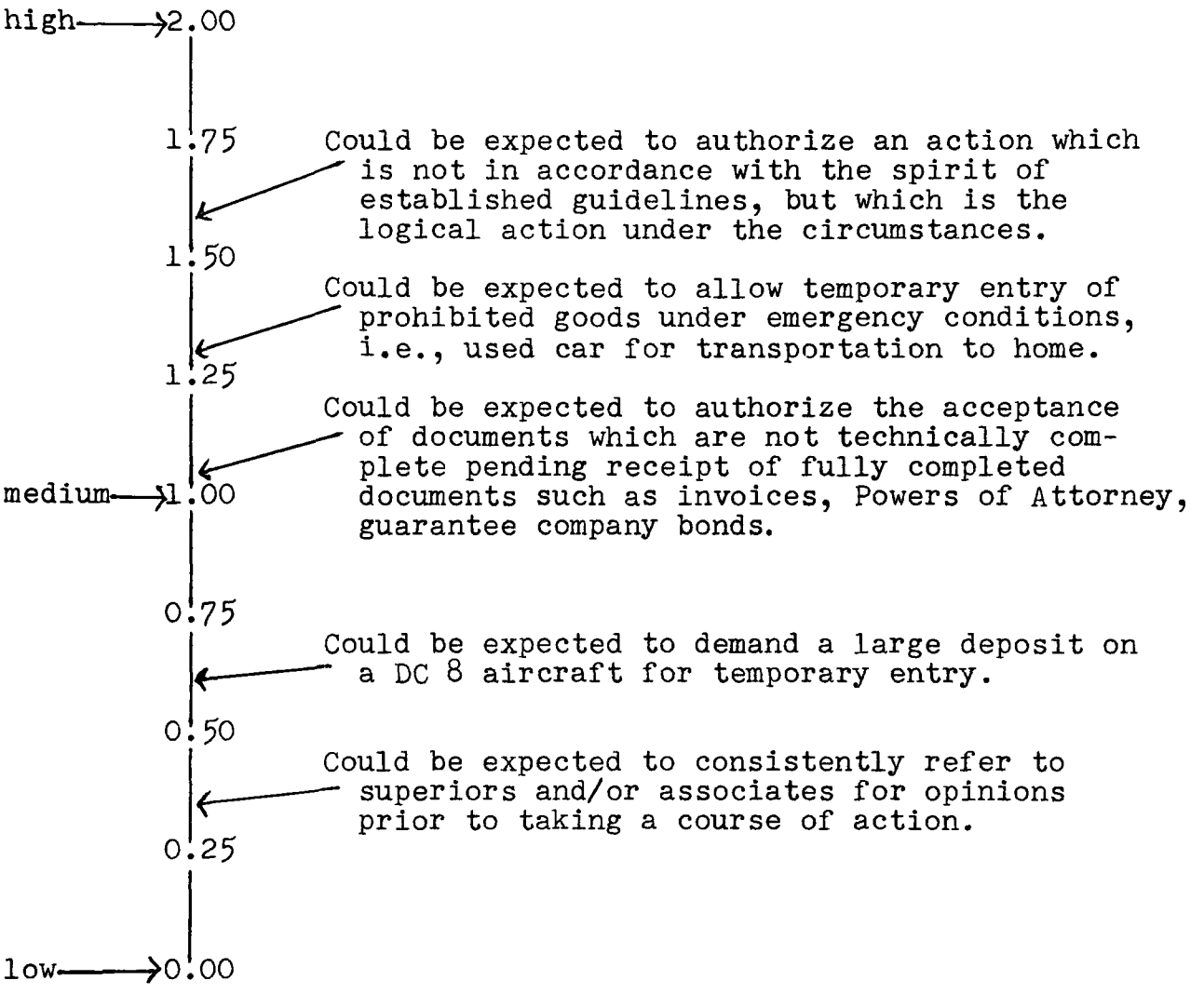
Tolerance to Stress.- The ability to remain calm and objective under stress, and to be consistent in handling all situations, both personal and public.

Figure 3.- Behavioural Anchored Rating Scale for the Dimension of Tolerance for Stress.



Human Relations.- The ability to respond to the needs, feelings, and ideas of others; to accept the shortcomings of others; to motivate people towards a goal; and to be able to inspire in others a positive feeling toward you through tact, persuasiveness, and interest in them.

Figure 4.- Behavioural Anchored Rating Scale for the Dimension of Human Relations.



Decision Making.- The ability to decide a course of action without wasting time or energy, and to accept responsibility for the decision.

Figure 5.- Behavioural Anchored Rating Scale for the Dimension of Decision Making.

APPENDIX 3

A DESCRIPTION OF THE INTEGRATION RATING PROCEDURE

A DESCRIPTION OF THE INTEGRATION RATING PROCEDURE

The Integration process consists of four phases: data collection; evaluation; narrative report; and, overall assessment report. Behavioural data for each assessee is collected in the following manner: each assessor views a number of candidates in a variety of situations, for example, in the Task Force Exercise, Individual Case Presentation, Assessment Interview and the In-basket Interview; and, a behavioural description and a general report is completed by the assessors after each of the above situations. The procedures for rating each assessee are described on page 28. A narrative report is then written up describing each candidate's performance at the assessment centre in terms of each managerial dimension and describes the candidate's strengths, weaknesses, developmental needs and any recommendations the assessors have decided to suggest. The report concludes with a statement of the candidate's overall managerial potential.

APPENDIX 4

AN EXAMPLE OF THE RATING SCALE USED FOR EACH PREDICTOR

AN EXAMPLE OF THE RATING SCALE USED FOR EACH PREDICTOR

COMMUNICATION: the ability to receive and transmit information and ideas in a clear and concise manner

_____ 0. No evidence.

_____ 1. It is frequently necessary to ask for clarification or restatement in order to obtain a clear understanding of what he says.

_____ 2. He can be understood fairly easily, but appears to experience some difficulty in expressing himself.

_____ 3. He is readily understandable and demonstrates a good basic knowledge of vocabulary and grammar; his effectiveness is somewhat diminished by lack of facial or vocal expressiveness, eye contact, gesture, etc., which conveys a lack of forcefulness and enthusiasm.

_____ 4. He is articulate and clear, though not a forceful or dynamic speaker.

_____ 5. He is a highly articulate, as well as dynamic and enthusiastic speaker.

APPENDIX 5

A DESCRIPTION OF THE ASSESSMENT INTERVIEW

APPENDIX 5

A DESCRIPTION OF THE ASSESSMENT INTERVIEW

The Assessment Interview, between assessee and assessor, provides the assessor with an opportunity to evaluate characteristics and qualities of the assessee which are not brought out in the other structured exercises in the assessment program. The assessor has been exposed, at the Assessor Training Session, to the proper procedure for assessment interviewing and has practiced interviewing. Before the assessor conducts the interview he reviews the application form filled out by the assessee and he prepares his questions on the basis of the training he has received and the background information form which the candidate has completed. From the application form the assessor has access to data related to hobbies, interests, organizations to which the candidate belongs, educational background, work experiences, position desired and any additional information given by the candidate which is related to qualifications for employment. An Interview Guide assists the assessor in formulating questions for the interview. It provides guidelines for formulating questions which can be asked about the information obtained from the application form.

After the interview is completed, the assessor

fills out an Interview Rating Form in which he describes and rates the candidate on his work history, education and training, present social adjustment, personality, motivation and character. In addition to the above, the assessor summarizes the candidate's assets, liabilities and gives an overall rating and summary.

The assessment interview was used to measure the following managerial assessment dimensions: communication; self-motivation, tolerance for stress; and, human relations. A rating scale similar to the example provided in Appendix 4 was used to assess the candidates for each assessment dimension.

APPENDIX 6

A DESCRIPTION OF THE TASK FORCE EXERCISE

A DESCRIPTION OF THE TASK FORCE EXERCISE

In the Task Force, six individuals, each assuming a predetermined role and point of view in regard to the problem under discussion, are placed in a group situation which requires both competitive and cooperative behaviour. The participants are presented with the following problem.

The General Manufacturing Company, of which they are all vice-presidents, has lost a major account. Because of this, the president has designated the six vice-presidents to review three alternative courses of action designed to compensate for the lost account.

The vice-presidents represent Administration, Finance, Personnel, Manufacturing, Engineering and Marketing. The exercise is competitive in that each man has the priorities of his own section to defend. This is accomplished by giving a written description to each vice-president of his role in the company and the alternatives which he supports. However, they must ultimately cooperate in order to reach a consensus on one course of action to be followed by the company from the three alternatives suggested. An additional cooperative aspect which is built into the exercise is that each vice-president should have one other vice-president who supports

the same alternative.

The Task Force Exercise is used to assess candidates in terms of the following dimensions:

> Communication: the ability to receive and transmit information and ideas in a clear and concise manner. This is assessed when the candidate makes his presentation from his position and in the group discussion.
>
> Human Relations: the ability to respond to the needs, feelings, and ideas of others; to accept the shortcomings of others; to motivate people towards a goal; and to be able to inspire in others a positive feeling. This is assessed in the group discussion activity.

A rating guide is given to each assessor (Cf. Appendix 4 for a typical example of the rating scale). Although many other managerial dimensions are measured in the Task Force Exercise, only the dimensions of communication and human relations are used in the present investigation.

APPENDIX 7

A DESCRIPTION OF THE PERSONNEL PROBLEM EXERCISE

APPENDIX 7

A DESCRIPTION OF THE PERSONNEL PROBLEM EXERCISE

In the Personnel Problem Exercise the assessees are provided with information which describes their position, the group to which they belong and a description of a Director's situation in a particular government department. The participants are told of three situations which have been brought to the attention of the Director and are asked by the Director to consider the overall situation and formulate a plan of immediate action for making improvements. The Director considers the matter to be of high priority. The assessees are told that they have 45 minutes to discuss and formulate their recommendations.

The Personnel Problem Exercise is used to assess candidates in terms of the following dimensions:

> Communication: the ability to receive and transmit information and ideas in a clear and concise manner.
>
> Human Relations: the ability to respond to the needs, feelings, and ideas of others; to accept the shortcomings of others; to motivate people towards a goal; and to be able to inspire in others a positive feeling.

A rating guide is given to each assessor (Cf. Appendix 4 for a typical example). Although other managerial dimensions are measured in the Personnel Problem Exercise, only the dimensions of communication and human relations are used for the present investigation.

APPENDIX 8

A DESCRIPTION OF THE INDIVIDUAL CASE PRESENTATION ORAL

A DESCRIPTION OF THE INDIVIDUAL CASE PRESENTATION ORAL

In the Individual Case Presentation, the assessee is given a file folder of miscellaneous data from which he is to select, analyze, and report on a survey of student and supervisor attitudes towards a summer student employment program.

The file contains summaries of student and supervisor responses to two 25 item attitude questionnaires. It also contains data pertaining to a study of summer student employment. All of the questionaires and questions used in the studies are included in the file folder. In addition to relevant data, statistics on employee movement and retirement projections are included.

The assessee's task is to analyze the survey results and report his findings in an oral presentation to a committee of senior managers.

The assessors assess the ability of the candidate to receive and transmit information in a clear and concise manner in order to rate the candidate on the dimension of communication. At the same time the assessors observe and rate the candidate's ability to remain calm and objective under stress. It is assumed that the oral presentation

represents a more or less stressful situation. The ability to cope with stress is inferred from certain behaviours which are generally associated with nervousness and anxiety, on the one hand, or calmness and self-assurance on the other.

A rating guide is given to each assessor (Cf. Appendix 4 for a typical example). Although the Individual Case Presentation Oral measures other managerial dimensions, only the above mentioned dimensions are used in the present investigation.

APPENDIX 9

A DESCRIPTION OF THE IN-BASKET TEST

APPENDIX 9

A DESCRIPTION OF THE IN-BASKET TEST

The In-basket Test takes its name from the "in-basket" or tray on the desk of many managers in which letters, reports, memoranda, messages, and other papers are deposited for their attention and action. Many of the problems a manager faces on his job are presented to him in this manner. A management job in a prescribed setting and at a predetermined level, under given conditions, is simulated by a judicious choice of paperwork representing a cross-section of the range of problems and responsibilities that this manager may encounter on a day-to-day basis.

The person taking the test is provided with selected background material and references and a package of problems which have built-in priorities, relationships, and required decision-making. He is asked to work on the problems in a specific time period as if he were indeed on the job as described in the instructions. In this way, a sample of his administrative behaviour is obtained.

The test used in assessment is the Consolidated Fund In-basket Test, developed by Educational Testing Service, Princeton, New Jersey. It has the general background of a community fund, in which the assessee assumes the role of the Fund's paid director. He must deal with a small paid

staff, a Board of Directors composed of leading citizens, and a large volunteer organization in the community. This background and situation is considered to be familiar to most North Americans, and not to require any particular background of experience or education beyond secondary school level. The assessee has to act on a range of problems, some of them interrelated by common problems or by persons involved. Opportunity is given him to demonstrate such management skills as the ability to organize discrete pieces of information, to analyze the problems implicit in a situation, to consider the effect of alternative courses of action, to make decisions on a variety of problems involving people and accomplishing an objective, and to assume a leadership role in implementing the decisions.

Background information on the organization and current status of the Fund is given. Instructions to the subject are to take action on the problems as if he were, in fact, sitting at the desk of the director with a period of $2\frac{1}{2}$ hours in which to do as much work as he can.

The duties of the candidate as Director of the Bradford Consolidated Fund can be summarized in terms of analysis of the organization and organizational problems, planning and organizing of work, selection and allocation of staff, leadership, direction and control over the

organization.

The performance of each candidate is assessed on a rating scale similar to the example given in Appendix 4. In the present study, each candidate is assessed on the following dimensions using the In-basket Test: communication; human relations; and, decision making.

APPENDIX 10

A DESCRIPTION OF THE SELF DESCRIPTION INVENTORY

A DESCRIPTION OF THE SELF DESCRIPTION INVENTORY

The Self Description Inventory was developed by Edwin E. Ghiselli and consists of 64 pairs of personally descriptive adjectives. The adjectives were chosen so that both members of each pair are similar in terms of the social desirability of the traits they symbolize. In half the pairs the individual checks that adjective which he believes most characterizes him, both adjectives in these pairs referring to socially desirable traits. In the other half he checks the one he believes least characterizes him, both adjectives in these pairs referring to socially undesirable traits.

Measures of the following traits are given by the inventory: supervisory ability, intelligence, initiative, self-assurance, decisiveness, masculinity-feminity, maturity, working class affinity, achievement motivation, need for self actualization, need for power, need for high financial reward, and need for security. Only the decisiveness and achievement motivation scales were used in the present investigation to measure the respective managerial dimensions of decision making and self-motivation. All scales were developed empirically, each scale being based upon an item analysis of

the responses of specified criterion groups. The items in any given scale are those which yielded significant differences between criterion groups, and hence the different scales include different numbers of items. The items in the various scales carry different weights, the weights being proportional to the magnitudes of the differences in the responses of the criterion groups.

The criterion groups for the decisiveness scale consist of persons in top management positions and persons in middle management positions. For the achievement motivation scale the criterion groups consist of 119 men and 50 women holding top management and professional jobs, and 134 men and 37 women employed in semiskilled and unskilled work.

APPENDIX 11

DATA FOR THE DIMENSION OF COMMUNICATION

DATA FOR THE DIMENSION OF COMMUNICATION

Table XIV.

Means, SD's and Pearson r's for the Predictor Variable (Pure Clinical Method-Communication) and the Criterion Measure (BARS Ratings-Communication).

Sample A (N = 54)

Variable*	Mean	SD	Pearson r
1	2.95	0.85	.007[1] NS
2	1.30	0.31	

Beta weight = .003
Constant = 1.30
* Variable 1: Integration Ratings (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
1 Significance at $p<.05$ requires an $r \geq .269$.

Sample B (N = 54)

Variable*	Mean	SD	Pearson r
1	2.97	0.93	.059[1] NS
2	1.27	0.36	

Beta weight = 1.20
Constant = .02
* Variables 1 and 2 are the same as Sample A.
1 Significance at $p<.05$ requires a $r \geq .269$.

Table XV.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Trait Ratings Method-Communication) and the Criterion Measure (BARS Ratings-Communication).

Sample A (N = 54)

Variable[1]	1	2	3	4	5	6	7
1	1.00	0.19	0.57	0.55	0.53	0.52	-0.04
2	0.19	1.00	0.45	0.43	0.19	0.17	-0.12
3	0.57	0.45	1.00	0.49	0.33	0.49	-0.11
4	0.55	0.43	0.49	1.00	0.54	0.57	-0.03
5	0.53	0.19	0.33	0.54	1.00	0.73	0.06
6	0.52	0.17	0.49	0.57	0.73	1.00	0.04
7	-0.04	-0.12	-0.11	-0.03	0.06	0.04	1.00
Means	3.85	3.76	3.80	3.20	3.46	3.26	1.30
SD's	0.79	0.75	0.98	0.92	0.98	0.91	0.31
Beta weights	-0.01	-0.04	-0.04	0.003	0.03	0.02	
Constant =	1.47						
Multiple R =	.182[a] NS						

1 Variable 1: Assessment Interview)
Variable 2: In-basket (Written))
Variable 3: In-basket (Oral))
Variable 4: Individual Case Presentation Oral) Predictors
Variable 5: Personnel Problem)
Variable 6: Task Force Exercise)
Variable 7: Behavioural Anchored Rating Scale (Criterion)

For the intercorrelation of any two variables significance at $p < .05$ requires a $r \geq .269$ and significance at $p < .01$ requires a $r \geq .348$.

a Significance at $p < .05$ requires a $R \geq .485$.

Table XVI.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Trait Ratings Method-Communication) and the Criterion Measure (BARS Ratings-Communication).

Sample B (N = 54)

Variables*	1	2	3	4	5	6	7
1	1.00	0.31	0.30	0.36	0.19	0.36	0.04
2	0.31	1.00	0.62	0.53	0.30	0.40	0.23
3	0.30	0.62	1.00	0.52	0.40	0.45	0.35
4	0.36	0.53	0.52	1.00	0.46	0.63	0.17
5	0.19	0.30	0.40	0.46	1.00	0.62	-0.02
6	0.36	0.40	0.45	0.63	0.62	1.00	0.22
7	0.04	0.23	0.35	0.17	-0.02	0.22	1.00
Means	3.85	3.69	3.80	3.07	3.46	3.40	1.27
SD's	0.90	0.93	0.88	0.99	0.75	0.77	0.36
Beta weights	-0.05	0.01	0.16	-0.02	-0.16	0.15	
Constant =	0.92						
Multiple R =	.451[1] NS						

* Variables 1 - 7 are the same as Sample A (Table XV).

For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .348$.

1 Significance at $p<.05$ requires a $r \geq .485$.

Table XVII.

Means, SD's and Pearson r's for the Predictor Variable (Pure Statistical Method-Communication) and the Criterion Measure (BARS Ratings-Communication).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	6.43	3.14	.110[a] NS
2	1.30	0.31	

Beta weight = .01
Constant = 1.24
1 Variable 1: Structure of Intellect Model Expressional Fluency (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
a Significance at $p<.05$ requires a $r \geq .269$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	6.24	2.86	.102[a] NS
2	1.27	0.35	

Beta weight = .01
Constant = 1.19
1 Variables 1 and 2 are the same as Sample A.
a Significance at $p<.05$ requires a $r \geq .269$.

Table XVIII.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Communication) and the Criterion Measure (BARS Ratings-Communication).

Sample A (N = 54)

Variable[1]	1	2	3	4	5	6	7	8	9
1	1.00	0.29	0.18	0.33	0.48	0.34	0.27	0.46	0.11
2	0.29	1.00	0.19	0.57	0.63	0.55	0.53	0.52	-0.04
3	0.18	0.19	1.00	0.45	0.44	0.43	0.19	0.17	-0.12
4	0.33	0.57	0.45	1.00	0.57	0.49	0.33	0.49	-0.11
5	0.48	0.63	0.44	0.57	1.00	0.71	0.67	0.63	0.01
6	0.34	0.55	0.43	0.49	0.71	1.00	0.54	0.57	-0.03
7	0.27	0.53	0.19	0.33	0.67	0.54	1.00	0.73	0.06
8	0.46	0.52	0.17	0.49	0.63	0.57	0.73	1.00	0.04
9	0.11	-0.04	-0.12	-0.11	0.01	-0.03	0.06	0.04	1.00
Means	6.43	3.85	3.76	3.80	2.95	3.20	3.46	3.26	1.30
SD's	3.14	0.79	0.75	0.98	0.85	0.92	0.98	0.91	0.31
Beta weights	0.01	-0.02	-0.05	-0.04	0.01	-0.005	0.03	-0.003	
Constant = 1.49									
Multiple R= .229[a] NS									

1 Variable 1: Structure of Intellect Model (Expressional Fluency))
Variable 2: Assessment Interview)
Variable 3: In-basket (Written))
Variable 4: In-basket (Oral))
Variable 5: Integration Ratings) Predictors
Variable 6: Individual Case Presentation Oral)
Variable 7: Personnel Problem)
Variable 8: Task Force Exercise)
Variable 9: Behavioural Anchored Rating Scale (Criterion)

For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p \leq .01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .526$.

Table XIX.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Communication) and the Criterion Measure (BARS Ratings-Communication).

Sample B (N = 54)

Variables*	1	2	3	4	5	6	7	8	9
1	1.00	0.11	0.46	0.30	0.35	0.33	0.13	0.09	0.10
2	0.11	1.00	0.31	0.30	0.45	0.36	0.19	0.36	0.04
3	0.46	0.31	1.00	0.62	0.61	0.53	0.30	0.40	0.23
4	0.30	0.30	0.62	1.00	0.63	0.52	0.40	0.45	0.35
5	0.35	0.45	0.61	0.63	1.00	0.74	0.66	0.67	0.06
6	0.33	0.36	0.53	0.52	0.74	1.00	0.46	0.63	0.17
7	0.13	0.19	0.30	0.40	0.66	0.46	1.00	0.62	-0.02
8	0.09	0.36	0.40	0.45	0.67	0.63	0.62	1.00	0.22
9	0.10	0.04	0.23	0.35	0.06	0.17	-0.02	0.22	1.00
Means	6.24	3.85	3.68	3.80	2.97	3.07	3.46	3.40	1.27
SD's	2.86	0.90	0.93	0.88	0.93	0.99	0.75	0.77	0.36
Beta weights	0.01	-0.02	0.04	0.18	-0.18	0.04	-0.09	0.17	
Constant = .63									
Multiple R = .505[1] NS									

* Variables 1 - 9 are the same as sample A (Table XVIII).

1 Significance at $p<.05$ requires a $R \geq .526$.

Table XX.

Multiple R's, Pearson r's, Differences between Two Correlation Coefficients (z), Cross-validated r's and R's, and "Drop" in Cross-validation for the Dimension of Communication. (N=54 for each sample)

Method	Sample	Multiple R^{a} # Pearson r^{b} #	z_r	z^{1}	Cross-validation	R^{a} # r^{b} #	z_r	z^{1}	"Drop"
Pure Clinical	A	$.007^{b}$	.007	.26	B to A	$.003^{b}$	.003	.17	.004
	B	$.059^{b}$	.059		A to B	$-.036^{b}$	.036		.023
Trait Ratings	A	$.182^{a}$	.184	1.53	B to A	$-.109^{a}$	.109	.96	.073
	B	$.451^{a}$	.486		A to B	$-.291^{a}$	.300		.160
Pure Statistical	A	$.110^{b}$	.110	.04	B to A	$.110^{b}$	.110	.05	.000
	B	$.102^{b}$	.102		A to B	$.101^{b}$	.101		.001
Mechanical Composite	A	$.229^{a}$	.233	1.63	B to A	$-.099^{a}$	.099	1.00	.130
	B	$.505^{a}$	.556		A to B	$-.288^{a}$	.298		.217

The required levels of significance at p .05 and p .01 for the Multiple R's or Pearson r's are listed in the tables for each method of prediction for the dimension of Communication.

1 *Significance at $p<.05$ requires a $z \geq 1.96$.
**Significance at $p<.01$ requires a $z \geq 2.58$.

Table XXI.

Intercorrelation Matrix, Means and SD's of the Predicted Scores for the Methods Measuring the Dimension of Communication.
(N = 54)

Method*	1	2	3	4
Sample A				
1	1.00	0.19	0.45	0.09
2	0.19	1.00	0.29	0.91
3	0.45	0.29	1.00	0.30
4	0.09	0.91	0.30	1.00
Means	1.27	1.25	1.24	1.26
SD's	0.02	0.19	0.03	0.22
Sample B				
1	1.00	-0.20	0.22	-0.16
2	-0.20	1.00	-0.40	0.74
3	0.22	-0.40	1.00	0.25
4	-0.16	0.74	0.25	1.00
Means	1.30	1.30	1.28	1.30
SD's	0.005	0.06	0.03	0.07

* Method 1: Pure Clinical
Method 2: Trait Ratings
Method 3: Pure Statistical
Method 4: Mechanical Composite

APPENDIX 12

DATA FOR THE DIMENSION OF SELF-MOTIVATION

APPENDIX 12

DATA FOR THE DIMENSION OF SELF-MOTIVATION

Table XXV.

Means, SD's and Pearson r's for the Predictor Variable (Pure Clinical Method-Self-motivation) and the Criterion Measure (BARS Ratings-Self-motivation).

Sample A (N - 54)

Variable[1]	Mean	SD	Pearson r
1	2.56	1.14	.137[a]NS
2	1.30	0.37	

Beta weight = .04
Constant = 1.19
1 Variable 1: Integration Ratings (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
a Significance at $p<.05$ requires a $r \geq .269$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	2.95	1.21	-.101[a]NS
2	1.30	0.39	

Beta weight = -.03
Constant = 1.40
1 Variables 1 and 2 are the same as Sample A.
a Significance at $p<.05$ requires a $r \geq .269$.

Table XXVI.

Means, SD's and Pearson r's for the Predictor Variable (Trait Ratings Method-Self-motivation) and the Criterion Measure (BARS Ratings-Self-motivation).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	3.24	0.95	.035[a] NS
2	1.30	0.37	

Beta weight = .01
Constant = 1.26
1 Variable 1: Assessment Interview (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
a Significance at $p<.05$ requires a $r \geq .269$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	3.39	0.79	-.094[a] NS
2	1.30	0.39	

Beta weight = -.05
Constant = 1.46
1 Variables 1 and 2 are the same as Sample A.
a Significance at $p<.05$ requires a $r \geq .269$.

Table XXVII.

Means, SD's and Pearson r's for the Predictor Variable (Pure Statistical Method-Self-motivation) and the Criterion Measure (BARS Ratings-Self-motivation).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	36.52	9.03	.121[a] NS
2	1.30	0.37	

Beta weight = .005
Constant = 1.12
1 Variable 1: Self-description Inventory-Achievement Motivation (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
a Significance at $p<.05$ requires a $r \geq .269$

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	37.74	9.08	.121[a] NS
2	1.30	0.39	

Beta weight = .05
Constant = 1.11
1 Variables 1 and 2 are the same as Sample A.
a Significance at $p<.05$ requires a $r \geq .269$.

Table XXVIII.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Self-motivation) and the Criterion Measure (BARS Ratings-Self-motivation).

Sample A (N = 54)

Variable[1]	1	2	3	4
1	1.00	0.07	0.06	0.12
2	0.07	1.00	0.59	0.04
3	0.06	0.59	1.00	0.14
4	0.12	0.04	0.14	1.00
Means	36.52	3.24	2.56	1.30
SD's	9.03	0.95	1.14	0.37
Beta weights =	0.005	-0.03	0.06	
Constant =	1.08			
Multiple R =	.188[a] NS			

1 Variable 1: Self-description Inventory-Achievement Motivation) Predictors
Variable 2: Assessment Interview)
Variable 3: Integration Ratings)
Variable 4: Behavioural Anchored Rating Scale (Criterion)

For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .379$.

Table XXIX.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Self-motivation) and the Criterion Measure (BARS Ratings-Self-motivation).

Sample B (N = 54)

Variable[1]	1	2	3	4
1	1.00	0.25	0.23	0.12
2	0.25	1.00	0.65	-0.09
3	0.23	0.65	1.00	-0.10
4	0.12	-0.09	-0.10	1.00
Means	37.74	3.39	2.95	1.30
SD's	9.08	0.79	1.21	0.39
Beta weights =	0.007	-0.04	-0.03	
Constant =	1.26			
Multiple R =	.188[a] NS			

1 Variables 1 - 4 are the same as Sample A (Table XXVIII). For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .379$.

Table XXX.

Multiple R's, Pearson r's, Differences between Two Correlation Coefficients (z), Cross-validated r's and R's, and "Drop" in Cross-validation for the Dimension of Self-motivation. (N=54 for each sample)

Method	Sample	Multiple R^a # Pearson r^b #	z_r	z^1	Cross-validation	R^a r^b #	z_r	z^1	"Drop"
Pure Clinical	A	$.137^b$	.138	.19	B to A	$-.135^b$	.136	.18	.002
	B	$-.101^b$	.101		A to B	$-.100^b$	.100		.001
Trait Ratings	A	$.035^b$	.035	.30	B to A	$-.035^b$	.035	.30	.000
	B	$-.094^b$	.094		A to B	$-.094^b$	.094		.000
Pure Statistical	A	$.121^b$	.122	.00	B to A	$.121^b$	.122	.02	.000
	B	$.121^b$	.122		A to B	$.118^b$	.119		.003
Mechanical Composite	A	$.188^a$	.190	.00	B to A	$.021^a$	.021	.09	.167
	B	$.188^a$	.190		A to B	$.004^a$	.004		.184

The required levels of significance at $p<.05$ and $p<.01$ for the Multiple R's or Pearson r's are listed in the tables for each method of prediction for the dimension of Self-motivation.

1 *Significance at $p<.05$ requires a $z \geq 1.96$.
**Significance at $p<.01$ requires a $z \geq 2.58$.

Table XXXI.

Intercorrelation Matrix, Means and SD's of the Predicted Scores for the Methods Measuring the Dimension of Self-motivation.
(N = 54)

Method*	1	2	3	4
Sample A				
1	1.00	0.57	-0.06	0.58
2	0.57	1.00	-0.07	0.61
3	-0.06	-0.07	1.00	0.69
4	0.58	0.61	0.69	1.00
Means	1.32	1.29	1.47	1.31
SD's	0.03	0.05	0.09	0.09
Sample B				
1	1.00	0.65	0.23	0.82
2	0.65	1.00	0.26	0.69
3	0.23	0.26	1.00	0.69
4	0.82	0.42	0.69	1.00
Means	1.30	1.28	1.30	1.33
SD's	0.05	0.01	0.05	0.08

* Method 1: Pure Clinical
Method 2: Trait Ratings
Method 3: Pure Statistical
Method 4: Mechanical Composite

APPENDIX 13

DATA FOR THE DIMENSION OF TOLERANCE FOR STRESS

APPENDIX 13

DATA FOR THE DIMENSION OF TOLERANCE FOR STRESS

Table XXXV.

Means, SD's and Pearson r's for the Predictor Variable (Pure Clinical Method-Tolerance for Stress) and the Criterion Measure (BARS Ratings-Tolerance for Stress).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	2.68	0.84	-.0006[a] NS
2	1.35	0.35	

Beta weight = -.003
Constant = 1.35

1 Variable 1: Integration Ratings (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)

a Significance at $p<.05$ requires a $r \geq .269$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	2.85	0.94	-.032[a] NS
2	1.29	0.37	

Beta weight = -.01
Constant = 1.33

1 Variables 1 and 2 are the same as Sample A.

a Significance at $p<.05$ requires a $r \geq .269$.

Table XXXVI.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Trait Ratings Method-Tolerance for Stress) and the Criterion Measure (BARS Ratings-Tolerance for Stress).

Sample A (N = 54)

Variable[1]		1	2	3
1		1.00	0.53	0.08
2		0.53	1.00	0.04
3		0.08	0.04	1.00
Means		3.48	2.86	1.35
SD's		0.79	0.97	0.35
Beta weights	=	0.04	-0.003	
Constant	=	1.23		
Multiple R	=	.082[a] NS		

1 Variable 1: Assessment Interview)
Variable 2: Individual Case Presentation) Predictors
Oral)
Variable 3: Behavioural Anchored Rating Scale (Criterion)
For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .336$.

Table XXXVII.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Trait Ratings Method-Tolerance for Stress) and the Criterion Measure (BARS Ratings-Tolerance for Stress).

Sample B (N = 54)

Variable[1]	1	2	3
1	1.00	0.56	-0.01
2	0.56	1.00	-0.01
3	-0.01	-0.01	1.00
Means	3.63	2.84	1.29
SD's	0.83	0.97	0.37
Beta weights =	-0.007	0.001	
Constant =	1.32		
Multiple R =	0.14[a] NS		

1 Variables 1 - 3 are the same as Sample A (Table XXXVI). For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .336$.

Table XXXVIII.

Means, SD's and Pearson r's for the Predictor Variable (Pure Statistical Method-Tolerance for Stress) and the Criterion Measure (BARS Ratings-Tolerance for Stress).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	26.15	5.23	-.067[a] NS
2	1.35	0.35	

Beta weight = -.004
Constant = 1.47

1 Variable 1: Gordon Personal Profile-Emotional Stability (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
a Significance at $p<.05$ requires a $r \geq .269$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	28.04	4.75	.270*
2	1.29	0.37	

Beta weight = .02
Constant = .70

1 Variables 1 and 2 are the same as Sample A.
*Significance at $p<.05$ requires a $r \geq .269$.
**Significance at $p<.01$ requires a $r \geq .348$.

Table XXXIX.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Tolerance for Stress) and the Criterion Measure (BARS Ratings-Tolerance for Stress).

Sample A (N = 54)

Variable[1]	1	2	3	4	5
1	1.00	-0.04	0.16	-0.02	-0.07
2	-0.04	1.00	0.50	0.53	0.08
3	0.16	0.50	1.00	0.74	-0.001
4	-0.02	0.53	0.74	1.00	0.04
5	-0.07	0.08	-0.001	0.04	1.00
Means	26.15	3.48	2.68	2.86	1.35
SD's	5.23	0.79	0.84	0.97	0.35
Beta weights =	-0.003	0.04	-0.03	0.01	
Constant =	1.34				
Multiple R =	.111[a] NS				

1 Variable 1: Gordon Personal Profile-Emotional Stability)
Variable 2: Assessment Inverview-Rating Guide)
Variable 3: Integration Ratings) Predictors
Variable 4: Individual Case Presentation Oral)
Variable 5: Behavioural Anchored Rating Scale (Criterion)

For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .412$.

Table XL.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Tolerance for Stress) and the Criterion Measure (BARS Ratings-Tolerance for Stress).

Sample B (N = 54)

Variable[1]		1	2	3	4	5
1		1.00	-0.10	-0.14	-0.29	0.27
2		-0.10	1.00	0.57	0.56	-0.01
3		-0.14	0.57	1.00	0.73	-0.03
4		-0.29	0.56	0.73	1.00	-0.01
5		0.27	-0.01	-0.03	-0.01	1.00
Means		28.04	3.63	2.85	2.84	1.29
SD's		4.75	0.83	0.94	0.97	0.37
Beta weights	=	0.02	-0.01	-0.04	0.06	
Constant	=	0.59				
Multiple R	=	.289[a] NS				

1 Variables 1 - 5 are the same as Sample A (Table XXXIX). For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .412$.

Table XLI.

Multiple R's, Pearson r's, Differences between Two Correlation Coefficients (z), Cross-validated r's and R's, and "Drop" in Cross-validation for the Dimension of Tolerance for Stress. (N=54 for each sample)

Method	Sample	Multiple R^a Pearson r^b #	z_r	z^1	Cross-validation	R^a r^b #	z_r	z^1	"Drop"
Pure Clinical	A	$-.0006^b$	.0006	.16	B to A	$.0006^b$	.0006	.02	.000
	B	$-.032^b$	.032		A to B	$-.004^b$	.004		.028
Trait Ratings	A	$.082^a$	.082	.34	B to A	$-.081^a$	.081	.34	.001
	B	$.014^a$	.014		A to B	$-.013^a$	.013		.001
Pure Statistical	A	$-.067^b$	.067	1.06	B to A	$-.067^b$	.067	1.06	.000
	B	$.270^{b*}$	.277		A to B	$.270^{b*}$	.277		.000
Mechanical Composite	A	$.111^a$	.110	.95	B to A	$-.049^a$	.049	.23	.062
	B	$.289^a$	.299		A to B	$-.094^a$	.094		.195

The required levels of significance at $p<.05$ and $p<.01$ for the Multiple R's or Pearson r's are listed in the tables for each method of prediction for the dimension of Tolerance for Stress.

1 *Significance at $p<.05$ requires a $z \geq 1.96$.
**Significance at $p<.01$ requires a $z \geq 2.58$.

Table XLII.

Intercorrelation Matrix, Means and SD's of the Predicted Scores for the Methods Measuring the Dimension of Tolerance for Stress.
(N = 54)

Method*	1	2	3	4
Sample A				
1	1.00	0.42	-0.20	-0.27
2	0.42	1.00	-0.02	-0.07
3	-0.20	-0.02	1.00	0.95
4	-0.27	-0.07	0.95	1.00
Means	1.29	1.29	1.21	1.26
SD's	0.01	0.01	0.10	0.13
Sample B				
1	1.00	-0.49	0.10	-0.18
2	-0.49	1.00	-0.09	0.80
3	0.10	-0.09	1.00	-0.47
4	-0.18	0.80	-0.47	1.00
Means	1.34	1.35	1.58	1.36
SD's	0.01	0.03	0.02	0.04

* Method 1: Pure Clinical
Method 2: Trait Ratings
Method 3: Pure Statistical
Method 4: Mechanical Composite

APPENDIX 14

DATA FOR THE DIMENSION OF HUMAN RELATIONS

DATA FOR THE DIMENSION OF HUMAN RELATIONS

Table XLVI.

Means, SD's and Pearson r's for the Predictor Variable (Pure Clinical Method-Human Relations) and the Criterion Measure (BARS Ratings-Human Relations).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	2.83	0.73	.093[a] NS
2	1.22	0.35	

Beta weight = .04
Constant = 1.10

1 Variable 1: Integration Rating (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
a Significance at $p < .05$ requires a $r \geq .269$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	2.89	0.91	.049[a] NS
2	1.17	0.38	

Beta weight = .02
Constant = 1.11

1 Variables 1 and 2 are the same as Sample A.
a Significance at $p < .05$ requires a $r \geq .269$.

Table XLVII.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Trait Ratings Method-Human Relations) and the Criterion Measure (BARS Ratings-Human Relations).

Sample A (N = 54)

Variable[1]		1	2	3	4
1		1.00	0.15	0.32	0.19
2		0.15	1.00	0.62	0.14
3		0.32	0.62	1.00	0.26
4		0.19	0.14	0.26	1.00
Means		3.72	3.03	2.98	1.22
SD's		0.76	0.78	0.67	0.35
Beta weights	=	0.05	-0.01	0.13	
Constant	=	0.690			
Multiple R	=	.284[a] NS			

1 Variable 1: Assessment Interview)
Variable 2: Personnel Problem) Predictors
Variable 3: Task Force Exercise)
Variable 4: Behavioural Anchored Rating Scale (Criterion)

For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .379$.

Table XLVIII.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Trait Ratings Method-Human Relations) and the Criterion Measure (BARS Ratings-Human Relations).

Sample B (N = 54)

Variable[1]		1	2	3	4
1		1.00	0.46	0.36	0.30
2		0.46	1.00	0.60	0.02
3		0.36	0.60	1.00	0.12
4		0.30	0.02	0.12	1.00
Means		3.83	3.12	3.01	1.17
SD's		0.64	0.55	0.76	0.38
Beta weights	=	0.22	-0.15	0.06	
Constant	=	0.64			
Multiple R	=	.344[a] NS			

1 Variables 1 - 4 are the same as Sample A (Table XLVII). For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .379$.

Table XLIX.

Means, SD's and Pearson r's for the Predictor Variable (Pure Statistical Method-Human Relations) and the Criterion Measure (BARS Ratings-Human Relations).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	27.46	5.73	-.002[a] NS
2	1.22	0.35	

Beta weight = -.0001
Constant = 1.22

1 Variable 1: Gordon Personal Inventory-Personal Relations (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)

a Significance at $p<.05$ requires a $r \geq .269$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	28.07	4.47	.063[a] NS
2	1.17	0.38	

Beta weight = .005
Constant = 1.02

1 Variable 1 and 2 are the same as Sample A.

a Significance at $p<.05$ requires a $r \geq .269$.

Table L.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Human Relations) and the Criterion Measure (BARS Ratings-Human Relations).

Sample A (N = 54)

Variable[1]		1	2	3	4	5	6
1		1.00	0.18	0.06	-0.15	-0.09	-0.002
2		0.18	1.00	0.66	0.15	0.32	0.19
3		0.06	0.66	1.00	0.44	0.53	0.09
4		-0.15	0.15	0.44	1.00	0.62	0.14
5		-0.09	0.32	0.53	0.62	1.00	0.26
6		-0.002	0.19	0.09	0.14	0.26	1.00
Means		27.46	3.72	2.83	3.03	2.98	1.22
SD's		5.73	0.76	0.73	0.78	0.67	0.35
Beta weights	=	-0.0002	0.11	-0.11	0.01	0.15	
Constant	=	0.64					
Multiple R	=	.322[a] NS					

1 Variable 1: Gordon Personal Inventory-Personal Relations)
Variable 2: Assessment Interview)
Variable 3: Integration Ratings)
Variable 4: Personnel Problem) Predictors
Variable 5: Task Force Exercise)
Variable 6: Behavioural Anchored Rating Scale (Criterion)

For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .440$.

Table LI.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Human Relations) and the Criterion Measure (BARS Ratings-Human Relations).

Sample B (N = 54)

Variable[1]	1	2	3	4	5	6
1	1.00	0.14	0.06	-0.06	-0.12	0.06
2	0.14	1.00	0.42	0.46	0.36	0.30
3	0.06	0.42	1.00	0.66	0.57	0.05
4	-0.06	0.46	0.66	1.00	0.60	0.02
5	-0.12	0.36	0.57	0.60	1.00	0.12
6	0.06	0.30	0.05	0.02	0.12	1.00
Means	28.07	3.83	2.89	3.12	3.01	1.17
SD's	4.47	0.64	0.91	0.55	0.76	0.38
Beta weights =	0.002	0.22	-0.02	-0.13	0.07	
Constant =	0.57					
Multiple R =	.346[a] NS					

1 Variables 1 - 6 are the same as Sample A (Table L). For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .485$.

Table LII.

Multiple R's, Pearson r's, Differences between Two Correlation Coefficients (z), Cross-validated r's and R's, and "Drop" in Cross-validation for the Dimension of Human Relations. (N=54 for each sample)

Method	Sample	Multiple R^a # Pearson r^b #	z_r	z^1	Cross-validation	R^a # r^b	z_r	z^1	"Drop"
Pure Clinical	A	$.093^b$	.093	.20	B to A	$.088^b$	.088	.18	.005
	B	$.049^b$	.053		A to B	$.049^b$	.053		.000
Trait Ratings	A	$.284^a$	.293	.34	B to A	$.131^a$	.131	.34	.153
	B	$.344^a$	.360		A to B	$.196^a$	.198		.148
Pure Statistical	A	$-.002^b$	.002	.31	B to A	$.002^b$	.002	.30	.000
	B	$.063^b$	.063		A to B	$-.062^b$	.062		.001
Mechanical Composite	A	$.322^a$	.332	.14	B to A	$.148^a$	.151	.42	.174
	B	$.346^a$	.360		A to B	$.239^a$	.234		.107

The required levels of significance at $p<.05$ and $p<.01$ for the Multiple R's or Pearson r's are listed in the tables for each method of prediction for the dimension of Human Relations.

1 *Significance at $p<.05$ requires a $z \geq 1.96$.
**Significance at $p<.01$ requires a $z \geq 2.58$.

Table LIII.

Intercorrelation Matrix, Means and SD's of the Predicted Scores for the Methods Measuring the Dimension of Human Relations.
(N = 54)

Method*	1	2	3	4
Sample A				
1	1.00	0.43	0.08	0.41
2	0.43	1.00	0.23	0.99
3	0.08	0.23	1.00	0.29
4	0.41	0.99	0.29	1.00
Means	1.16	1.15	1.15	1.16
SD's	0.02	0.19	0.03	0.18
Sample B				
1	1.00	0.60	-0.09	-0.01
2	0.60	1.00	-0.06	0.77
3	-0.09	-0.06	1.00	-0.004
4	-0.01	0.77	-0.004	1.00
Means	1.21	1.22	1.19	1.20
SD's	0.04	0.11	0.01	0.13

* Method 1: Pure Clinical
Method 2: Trait Ratings
Method 3: Pure Statistical
Method 4: Mechanical Composite

APPENDIX 15

DATA FOR THE DIMENSION OF DECISION MAKING

DATA FOR THE DIMENSION OF DECISION MAKING

Table LVII.

Means, SD's and Pearson r's for the Predictor Variable (Pure Clinical Method-Decision Making) and the Criterion Measure (BARS Ratings-Decision Making).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	2.66	0.86	.217[a]
2	1.29	0.35	

Beta weight = .09
Constant = 1.06

1 Variable 1: Integration Ratings (Predictor)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
a Significance at $p<.05$ requires a $r \geq .269$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	2.64	0.89	-.022[a] NS
2	1.24	0.42	

Beta weight = -.01
Constant = 1.26

1 Variables 1 and 2 are the same as Sample A.
a Significance at $p<.05$ requires a $r \geq .269$.

Table LVIII.

Means, SD's and Pearson r's for the Predictor Variable (Trait Ratings Method-Decision Making) and the Criterion Measure (BARS Ratings-Decision Making).

Sample A (N = 54)

Variable[1]	Mean	SD	Pearson r
1	3.24	0.87	.334*
2	1.29	0.35	

Beta weight = .14
Constant = .85
1 Variable 1: In-basket (written)
Variable 2: Behavioural Anchored Rating Scale (Criterion)
* Significance at $p<.05$ requires a $r \geq .269$.
** Significance at $p<.01$ requires a $r \geq .348$.

Sample B (N = 54)

Variable[1]	Mean	SD	Pearson r
1	3.20	1.09	.274*
2	1.24	0.42	

Beta weight = .11
Constant = .90
1 Variables 1 and 2 are the same as Sample A.
* Significance at $p<.05$ requires a $r \geq .269$.
** Significance at $p<.01$ requires a $r \geq .348$.

Table LIX.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Pure Statistical Method-Decision Making) and the Criterion Measure (BARS Ratings-Decision Making).

Sample A (N = 54)

Variable[1]		1	2	3	4
1		1.00	0.32	-0.06	-0.16
2		0.32	1.00	-0.01	0.05
3		-0.06	-0.01	1.00	-0.04
4		-0.16	0.05	-0.04	1.00
Means		19.54	23.11	27.00	1.29
SD's		4.83	5.01	5.88	0.35
Beta weights	=	-0.01	0.01	-0.003	
Constant	=	1.47			
Multiple R	=	.199[a] NS			

1 Variable 1: Self Description Inventory-Decisiveness)
Variable 2: Gordon Personal Profile-Ascendancy) Predictors
Variable 3: Gordon Personal Inventory-Cautiousness)
Variable 4: Behavioural Anchored Rating Scale (Criterion)
For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .377$.

Table LX.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Pure Statistical Method-Decision Making) and the Criterion Measure (BARS Ratings-Decision Making).

Sample B (N = 54)

Variable[1]	1	2	3	4
1	1.00	0.36	-0.16	-0.05
2	0.36	1.00	-0.48	-0.12
3	-0.16	-0.48	1.00	-0.01
4	-0.05	-0.12	-0.01	1.00
Means	19.39	28.85	27.52	1.24
SD's	3.94	4.84	5.13	0.42
Beta weights =	0.0003	-0.01	-0.01	
Constant =	1.77			
Multiple R =	.144[a] NS			

1 Variables 1 - 4 are the same as Sample A (Table LIX). For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .379$.

Table LXI.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Decision Making) and the Criterion Measure (BARS Ratings-Decision Making).

Sample A (N = 54)

Variable[1]	1	2	3	4	5	6
1	1.00	0.32	-0.06	-0.10	0.29	-0.16
2	0.32	1.00	-0.01	0.23	0.27	0.05
3	-0.06	-0.01	1.00	-0.16	-0.26	-0.04
4	-0.01	0.23	-0.16	1.00	0.56	0.33
5	0.29	0.27	-0.26	0.56	1.00	0.22
6	-0.16	0.05	-0.04	0.33	0.22	1.00
Means	19.54	23.11	27.00	3.24	2.66	1.29
SD's	4.83	5.01	5.88	0.87	0.86	0.35
Beta weights =	-0.01	0.001	0.001	0.10	0.05	
Constant =	1.02					
Multiple R =	.371 NS[a]					

1 Variable 1: Self Description Inventory-Decisiveness)
Variable 2: Gordon Personal Profile-Ascendancy)
Variable 3: Gordon Personal Inventory-Cautiousness)Predictors
Variable 4: In-basket (Written))
Variable 5: Integration Ratings)
Variable 6: Behavioural Anchored Rating Scale (Criterion)

For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .440$.

Table LXII.

Intercorrelation Matrix, Means, SD's, and Multiple R for the Predictor Variables (Mechanical Composite Method-Decision Making) and the Criterion Measure (BARS Ratings-Decision Making).

Sample B (N = 54)

Variable[1]		1	2	3	4	5	6
1		1.00	0.36	-0.16	0.10	0.34	-0.05
2		0.36	1.00	-0.48	0.16	0.46	-0.12
3		-0.16	-0.48	1.00	-0.27	-0.44	-0.01
4		0.10	0.16	-0.27	1.00	0.65	0.27
5		0.34	0.46	-0.44	0.65	1.00	-0.02
6		-0.05	-0.12	-0.01	0.27	-0.02	1.00
Means		19.39	23.85	27.52	3.20	2.64	1.24
SD's		3.94	4.83	5.13	1.09	0.89	0.42
Beta weights	=	0.004	-0.01	-0.01	0.18	-0.16	
Constant	=	1.30					
Multiple R	=	.387[a] NS					

1 Variables 1 - 6 are the same as Sample A (Table LXI). For the intercorrelation of any two variables significance at $p<.05$ requires a $r \geq .269$ and significance at $p<.01$ requires a $r \geq .348$.

a Significance at $p<.05$ requires a $R \geq .440$.

Table LXIII.

Multiple R's, Pearson r's, Differences between Two Correlation Coefficients (z), Cross-validated r's and R's, and "Drop" in Cross-validation for the Dimension of Decision Making. (N=54 for each sample)

Method	Sample	Multiple R^a Pearson r^b #	z_r	z^1	Cross-validation	R^a r^b#	z_r	z^1	"Drop"
Pure Clinical	A	$.217^b$	.220	1.00	B to A	$-.176^b$	.177	.78	.041
	B	$-.022^b$	.022		A to B	$-.022^b$	.022		.000
Trait Ratings	A	$.334^{b*}$	.347	.33	B to A	$.334^{b*}$	.347	.33	.000
	B	$.274^{b*}$	.281		A to B	$-.274^{b*}$	.281		.000
Pure Statistical	A	$.199^a$	.202	.29	B to A	$-.013^a$	.013	.06	.186
	B	$.144^a$	.145		A to B	$-.025^a$	.025		.119
Mechanical Composite	A	$.371^a$	.389	.10	B to A	$.144^a$	.145	.36	.227
	B	$.387^a$	.408		A to B	$.214^a$	.217		.173

The required levels of significance at $p<.05$ and $p<.01$ for the Multipe R's or Pearson r's are listed in the tables for each method of prediction for the dimension of Decision Making.

1 *Significance at $p<.05$ requires a $z \geq 1.96$.
**Significance at $p<.01$ requires a $z \geq 2.58$.

Table LXIV.

Intercorrelation Matrix, Means and SD's of the Predicted Scores for the Methods Measuring the Dimension of Decision Making.
(N = 54)

Method*	1	2	3	4
Sample A				
1	1.00	-0.52	0.16	0.34
2	-0.52	1.00	-0.11	0.56
3	0.16	-0.11	1.00	0.30
4	0.34	0.56	0.30	1.00
Means	1.23	1.23	1.24	1.24
SD's	0.01	0.09	0.08	0.14
Sample B				
1	1.00	0.63	0.10	0.70
2	0.63	1.00	0.08	0.93
3	0.10	0.08	1.00	0.32
4	0.70	0.93	0.32	1.00
Means	1.29	1.28	1.30	1.27
SD's	0.08	0.15	0.06	0.14

* Method 1: Pure Clinical
Method 2: Trait Ratings
Method 3: Pure Statistical
Method 4: Mechanical Composite

APPENDIX 16

ABSTRACT OF

A Test of Sawyer's Model for the Comparison of Predictive Methods for Five Managerial Dimensions in an Assessment Centre

APPENDIX 16

ABSTRACT OF

A Test of Sawyer's Model for the Comparison of Predictive Methods for Five Managerial Dimensions in an Assessment Centre

Two samples of fifty-four subjects who were self-nominated, met the Public Service classification requirements and were willing to be geographically mobile were assessed for managerial potential in the Managerial Assessment and Research Division Assessment Centre, Public Service Commission, Ottawa. They were assessed in the context of a Behavioural Validation Prediction Model employing four methods of assessment (Pure Clinical, Trait Ratings, Pure Statistical, and Mechanical Composite) based on a predictor classification model developed by Sawyer. For the purpose of this investigation five assessment dimensions were measured (communication, self-motivation, tolerance for stress, human relations, and decision making). At the completion of assessment procedures, all of the subjects were placed in job positions in a Canadian federal government department. After six months, supervisors' on-the-job performance ratings were obtained for each of the five

---

1 Lyle D. Pederson, Master's thesis presented to the School of Graduate Studies of the University of Ottawa, Ontario, April 1974, xv-184p.

assessment dimensions using a Behavioural Anchored Rating Scale developed by managers of the federal government department based on a "retranslation" procedure suggested by Smith and Kendall.

The data for the Pure Clinical, Trait Ratings, Pure Statistical and Mechanical Composite Methods of assessment was cross-validated using a double cross-validation design.

Statistically there were no significant differences ($p<.05$) between any two of the methods for each of the five assessment dimensions for either sample.