

AN INQUIRY INTO THE USEFULNESS OF
PSYCHOMETRIC TECHNIQUES
IN THE SELECTION OF PRISON OFFICERS

by

RICHARD HAROLD DOWNEY

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
in the Department
of
PHILOSOPHY AND PSYCHOLOGY

We accept this thesis as conforming to the
standard required from candidates for the
degree of MASTER OF ARTS.

Members of the Department of
Philosophy and Psychology.

THE UNIVERSITY OF BRITISH COLUMBIA

April 1956

ACKNOWLEDGMENTS

The writer is most grateful to his Faculty adviser, Dr. E. I. Signori, for his helpful criticisms and suggestions.

The writer is indebted to Warden Hugh Christie of Oakalla Prison Farm for his enthusiastic permission to launch this investigation and his wholehearted cooperation when the study got under way.

Special thanks are due to Mr. Duncan Clark, Deputy Warden of Treatment, of Oakalla who cooperated in every way to arrange testing schedules and interviews.

The writer also wishes to express his appreciation to the subjects of the investigation, whose cooperation in taking the tests made this study possible.

AN INQUIRY INTO THE USEFULNESS OF PSYCHOMETRIC TECHNIQUES IN THE SELECTION OF PRISON OFFICERS

Abstract

The purpose of the study was to explore the use of certain psychometric procedures and to study their value in relation to the problems of selection and prediction of prison personnel. The tests selected were the Wesman Personnel Classification Test, the Kuder Preference Record, form CH, the Minnesota Multiphasic Personality Inventory, and the Manson Evaluation. The criterion used to evaluate the tests was supervisors' ratings based on a forced distribution rating scale which measured only one trait, namely job proficiency.

The total sample consisted of 100 employed prison officers and all were rated for job proficiency. Sub-samples from the main sample were formed from the extreme ratings of the whole group. Thus, the top 27 per cent represented the successful group and the bottom 27 per cent the unsuccessful group.

The tests were analyzed individually. Mean profiles for the total sample were computed for all test variables and were compared with the published norms for each test. Mean scores and standard deviations for both groups of officers were computed and these data were examined for significant differences between the two groups.

From an analysis of the mean scores 14 out of 35 test

variables significantly discriminated between the two groups. The best predictor proved to be the Social Service scale of the Kuder Preference Record. The next best predictor was Part I (verbal) of the Personnel Classification Test.

Biserial correlation coefficients from widespread classes were also computed. These coefficients were generally of a low order, ranging from .04 to .49. Fourteen coefficients were significantly greater than zero.

A scattergram analysis of all tests and subtests was also undertaken to determine the best critical scores. This analysis revealed that, for practical purposes, six scales yielded effective cutting scores. The six scales in descending order of effectiveness for selection purposes were Part I, Personnel Classification Test, the Kuder Social Service scale, the MMPI Psychasthenia scale, the MMPI Depression scale, the MMPI Hostility scale, and the Manson Evaluation Total score.

In general, the results indicate the psychometric tests have value for the screening of prison officers. However, it is indicated that the results must be employed with caution until a further validation study is carried out on a sample of officers that is more representative of the population upon which it is intended to use the tests, viz., a sample of job applicants.

TABLE OF CONTENTS

CHAPTER		PAGE
I	Events Leading Up To The Present Study	1
II	Review Of The Literature	4
	Beginnings in employment psychology	4
	Indirect Contributions	4
	Direct Contributions	5
	Effect of World War I on employment testing	6
	Personnel research following World War I	7
	General	7
	Employment testing	8
	Implications of World War II	17
	Personnel research following World War II	18
	Extent of employment testing	18
	Personnel selection of prison officers	19
	Qualities requisite for effective prison officers	20
III	Tests Used	23
	Personnel Classification Test - Form A	23
	Reason for selection	23
	Structure of the test	24
	The Kuder Preference Record - Vocational	24
	Reason for selection	24
	Structure of the inventory	26
	Minnesota Multiphasic Personality Inventory Test - Group Form	27
	Reason for selection	27
	Structure of the inventory	29

CHAPTER	PAGE
The Manson Evaluation	30
Reason for selection	30
Structure of the Manson Evaluation	30
IV The Research Design	33
The sample	33
Selection of sample	33
Description of obtained sample	34
The criterion	39
Criteria used in this study	40
Other criteria	40
Administration of tests	41
Statistical procedures	43
V Results	45
Personnel Classification Test	45
Mean of total sample	45
Means and differences between means of successful and unsuccessful officers	45
Difference in variability between upper and lower levels of guards	47
Critical scores	47
Intercorrelations	48
Additional data	48
Summary	48
Kuder Preference Test	52
Mean profile of total sample	52
Comparison of mean scores between "good" and "poor" prison officers	52
The reliability of difference between standard deviations	57
Additional data	57
Summary	57

CHAPTER	PAGE
The Minnesota Multiphasic Personality Inventory	61
Profile of total sample based on mean T-scores of the MMPI	61
Comparison of mean scores between the two groups	63
Other MMPI scales	63
Biserial correlations	65
Reliability of standard deviations	65
Scattergrams	67
Summary and discussion of results	67
Manson Evaluation	71
Mean score profiles	71
Mean differences between the two groups	73
Variability between the two groups	73
Critical scores	76
Intercorrelation between subtests	76
Analysis of Manson Evaluation and MMPI scales	76
Analysis of individual scores	78
Conclusion and discussion	80
VI General Conclusions And Implications For Further Research	83
General conclusions	83
Suggestions and implications for further research	88
BIBLIOGRAPHY	92

TABLES

TABLE		PAGE
I	Chronological ages, means, medians and standard deviations for total sample and successful and unsuccessful groups of prison officers	36
II	Educational level, mean grade, median grade and standard deviations for total sample and successful and unsuccessful groups of prison officers	38
III	Means, medians and standard deviations of 100 prison officers on the Personnel Classification Test	45
IV	Means, difference between means, T-ratio and biserial correlations between upper and lower level groups of prison officers on the Personnel Classification Test	46
V	The reliability of differences between standard deviations on two groups of prison officers on the Personnel Classification Test	47
VI	Intercorrelations between total and subtest scores on the Personnel Classification Test	48
VII	Norms for a sample of prison officers on the Personnel Classification Test: raw scores with percentile equivalents	49
VIII	Means and standard deviations of total sample of prison officers on the Kuder Preference Record - Vocational	54
IX	Means and differences of the means for successful guards and unsuccessful guards on the Kuder Preference Record	55
X	The reliability of difference between standard deviations of the successful and unsuccessful groups on the Kuder	58

TABLE		PAGE
XI	Norms for 100 prison officers showing percentile distributions on the Kuder Preference Record - Vocational	59
XII	Mean T-scores and standard deviations of total sample on the MMPI original and additional scales	62
XIII	Mean T-scores, difference between means, and biserial correlations for two groups of prison officers on the MMPI scales and additional MMPI scales	64
XIV	Reliability of differences between standard deviations for successful and unsuccessful prison officers on the MMPI	66
XV	Frequencies (in per cents) of scores over the 65th T-score on the MMPI	70
XVI	Means and standard deviations of 100 prison officers on the Manson Evaluation	72
XVII	Mean T-scores, difference between means, and biserial correlations for successful and unsuccessful prison officers on the Manson Evaluation	74
XVIII	Reliability of differences between standard deviations for the two groups on the Manson Evaluation	75
XIX	Correlations between several Manson Evaluation and MMPI scales	77
XX	Percentages of groups scoring above critical scores on the Manson Evaluation	79
XXI	Effect of six cutting scores in eliminating successful and unsuccessful prison officers	87

FIGURES

FIGURE		PAGE
I	Differentiation achieved between successful prison officers and unsuccessful prison officers using Part I (verbal) of the Personnel Classification Test	51
II	Percentile ranks of mean scores obtained by 100 prison officers on the Kuder Preference Record - Vocational (percentiles from norms supplied by test publisher)	53
III	Percentile ranks of mean scores obtained by groups of successful and unsuccessful prison officers on the Kuder Preference Record - Vocational	56

CHAPTER I

EVENTS LEADING UP TO THE PRESENT STUDY

The selection of personnel at Oakalla Prison Farm is based largely on the results of a personal interview. This method of selection offers little objective evidence of either aptitude for or interest in prison work.

In the spring of 1954 two projects besides the one reported herein were to be undertaken at Oakalla Prison Farm to provide a basis for establishing a scientific program for the selection of prison officers. Part of the research was to consist of a study of biographical data for screening purposes. Another study was to yield a scientific description of the occupation of prison guard by using the critical incidence technique. This was regarded as a very important study to the total project since a scientific job description is obviously a first step in establishing a sound selection program. Hence it was felt that, unless the critical incidents study was completed, the investigation using psychological tests in a prison setting for selection purposes would of necessity have to be exploratory in nature and would be concerned only with the more general features of the occupation of prison guard.

In order to understand what qualities a good guard must possess and to select appropriate tests for use in this study, the author secured employment as a prison guard. This

step later proved to be highly worth while since the critical incidents study unfortunately had to be abandoned. After working as a prison officer for several months, it was concluded that the differentiation between a successful and an unsuccessful officer might possibly be accomplished if attention were focussed on three major personal components.

1. Intelligence or mental capacity - the successful officer seemed to be more alert and quicker to learn and grasp new situations than the unsuccessful officer.

2. Interests - the successful officer generally seemed to like his work and to deal with people in an understanding manner.

3. Personality - the successful officer generally seemed more "adjusted", at peace with himself, more tolerant and open-minded than the unsuccessful officer.

The battery of tests used in this study to measure these areas consisted of the Wesman Personnel Classification Test (which provides three scores of mental ability), the Kuder Preference Record - Vocational (which provides ten scores on interests), the Manson Evaluation Test (a test on alcoholism, which provides eight scores on personality traits), and the Minnesota Multiphasic Personality Inventory (which measures fourteen personality traits).

Several criteria were to have been used to validate the tests but, owing to practical considerations and the time which would be required to collect pertinent data, the present study was limited to the use of one criterion, viz., supervisors forced-distribution ratings.

Formal Statement of Problem

Specifically the purpose of this study was (1) to explore the relationship between a series of tested traits and successful and unsuccessful prison officers, and (2) to set up tentative test norms for these tests which differentiate successful from unsuccessful prison officers.

CHAPTER II

REVIEW OF THE LITERATURE

I. BEGINNINGS IN EMPLOYMENT PSYCHOLOGY

Indirect Contributions

Wilhelm Wundt was one of the first to break away from the field of philosophy and define psychology as an independent experimental science. The school of thought founded by Wundt began in 1879 at Leipzig and established a laboratory specifically devoted to the study of psychological problems (56, p. 242). Two years later, through the work of the laboratory, the first journal for the publication of experimental studies in psychology appeared, namely, Philosophische Studien (20).

This new school at first tended to shift the emphasis from differences between people to specific descriptions of how the human mind works and demonstrated that psychological phenomena can be objectively and quantitatively analyzed.

Probably of even greater significance to employment and personnel psychology was the behaviouristic doctrine, which gave an entirely new perspective to the field of psychology. This school stressed the fact that behaviour could be impersonally observed and measured. Emphasis is placed on the measurement and recording of how people behave in particular situations. Subsequently, tests were developed to measure human behaviour and

differences in aptitude and performance. Such devices are an important contribution to present day personnel methods.

Direct Contributions

Sir Francis Galton recognized the importance of measuring characteristics of the individual. He had been impressed by the Belgian statistician, Adolph Quetelet, who kept many records of social data such as births, deaths, marriages and crimes (5). Quetelet showed that data such as these, gathered from unselected samples, tend to form definite patterns of distribution, termed by statisticians as "normal" or Gaussian curves.

Inspired by the theories of Quetelet, Galton set up his own laboratory in 1882 with the object of measuring characteristics of many individuals and recording the results. He is reported (5) to have formulated some of the most important statistical tools used today: the method of correlation, the rating-scale method, the concept of the median, and the use of standard score. He also helped to develop several mental tests by which certain differences between individuals might be measured. The recognition of variations in individuals and the measurement of such differences as a basis for the selection and retention of workers is one of the goals of the personnel technician.

One of Wundt's students who pioneered in the field of industrial psychology was James McKeen Cattell. In him we see a merging of two movements, viz., the experimental method and the measurement of individual differences. The term "mental testing" as we know it today was first used by Cattell in 1890

in a published article entitled "Mental Tests and Measurements" (10). This article described various tests which were used in his laboratory at Columbia University.

The variety of tests used by Cattell were devised to provide a reliable basis for differentiating individuals from one another. Cattell showed how psychology could be usefully applied. His early work with mental tests for college students set the pattern for other research in the same area.

Intelligence testing as we know it today began in 1905 with the development of Binet and Simons' test of intelligence. Binet probably has contributed more to the development of mental tests than any other individual, although the Binet-type tests are used rarely in the business and industrial fields.

Another student of Wundt who directed the course of the new personnel technology was Hugo Munsterberg who spent the latter part of his life as director of the Psychological Laboratory at Harvard University (5). He also performed experiments which indicated the usefulness of the scientific selection of workers.

II. EFFECT OF WORLD WAR I ON EMPLOYMENT TESTING

Up to the time of World War I the main work in testing had been directed towards individual testing but such tests were impractical for measuring large numbers of men at one time. Just before the war Otis had developed a "group" test of intelligence drawn up on the basis of earlier tests of Binet and others. This test was tried out on a thousand men after which it was

revised and is now known as the Army Alpha Test. The Army Beta, a non-language test, was devised later to test illiterates (74).

Trade tests were also developed to select skilled personnel to carry on specialized tasks in the forces.

III. PERSONNEL RESEARCH FOLLOWING WORLD WAR I

A. General

Personnel techniques made some progress during World War I. In 1919 the journal, Personnel, was first published. During the war a considerable amount of groundwork in personnel testing had been done by army personnel technicians. Later when these technicians went into industry, personnel managers became interested in personnel testing and research, and laboratories for further exploration in various fields of interest were set up.

Also, cooperative programs began between universities and industry, such as the Bureau of Salesmanship Research, which was founded at the Carnegie Institute of Technology in 1922. A number of firms contributed funds to this bureau for the purpose of conducting research on salesmen selection on behalf of the member companies (3).

With the advent of the depression during the 1930's most plans for the development of personnel research in industry were set aside. However, a research group at the University of Minnesota at this time investigated several aspects of the unemployment situation. They attempted to answer such questions as: If one of two men were to be laid off, what should be the basis

for determining who should go? A summary of this research is found in "Men, Women, and Jobs", a report on results of the project of the Minnesota Employment Stabilization Research Institute (52, p. 129).

In 1935 the U. S. Employment Service helped to deal with some of the questions posed by the Minnesota Institute by creating the Occupational Research Program. This program draws together the resources of universities, private and public employment agencies, and the collaborative efforts of many industries (58, p. 3). One of the major results of the research was the development of the Dictionary of Occupational Titles (69) which provides definitions of about twenty-one thousand occupations. In addition, several selection devices, such as trade tests and aptitude tests, were constructed which enabled the U.S. Employment Service to supply industry with applicants who had been selected in a more adequate manner. The research improved the technical understanding of occupations and selection devices and contributed to the improvement of industrial and business personnel methods and military personnel management.

B. Employment Testing

The army tests marked a forward step in the progress of industrial psychology and initiated a group testing movement. This led to a wider use of the group testing technique to the measurement of other traits.

Tests used for selection in business and industry are of many kinds and are designed to suit the many special needs of various firms and occupations. In general the tests fall into three main categories, viz., aptitude, self-evaluation and achievement tests. These are used as follows:

1. Aptitude tests yield information relative to the applicant's capacity to learn a job quickly and efficiently.

2. Self-Evaluation tests consist of questionnaires and inventories which provide information about the applicant's suitability for a job in terms of interests and personality traits.

3. Achievement tests help to determine the extent to which the applicant has already acquired the knowledge or skills required in the job.

The subject of this thesis is concerned only with the two first-mentioned areas of employment testing and these are discussed more fully below.

1. Aptitude tests. According to Maier (44) aptitude tests are "designed to measure a person's potentiality for succeeding in certain tasks." The purpose of aptitude testing, then, is to determine before an applicant starts training how well he might succeed in it.

Aptitude tests have been devised for a variety of vocational fields. The history of educational and vocational aptitude testing parallels the developments in mental testing. The aptitude most commonly measured in the field of education

is scholastic aptitude or general mental capacity.

Tests have been developed for the following abilities: Mechanical, motor, clerical, musical, graphic arts, reading, and abilities in specific academic subjects, and in the fields of medicine, law, engineering, teaching, and other areas too numerous to mention.

Aptitude tests might also be divided into various categories as follows:

Specific vs. General Aptitude Tests. On this basis we may distinguish between tests that are designed to detect particular aptitudes from those that are constructed to detect general or average aptitudes. Patten's test for capacity to learn to operate the engine lathe (51) or Deemer's stenographic test are examples of specific aptitude tests. The Stenquist Mechanical Aptitude Tests (13, p. 245) on the other hand apparently are designed to test mechanical aptitude in general. They yield an over-all appraisal of a person's ability to do all kinds of mechanical work. The numerous so-called intelligence tests, such as the Binet-Simon tests, the Terman Group Test of Mental Ability, and others are also examples of general aptitude tests.

Number of units employed in test batteries. A second method of classifying aptitude tests relates to the number of test units employed for a given prognosis. The contrast lies primarily between the use of a single test or a battery of tests. In most cases aptitudes are of such complexity that a single

test will rarely be able to sample enough of the determining factors to make a useful prediction. According to Cronbeck (13) single standardized prognostic tests have declined in use. Test batteries are commonly used for aptitude testing nowadays.

Miniature tests vs. tests of abstract traits. Under this category aptitude tests may be divided into (1) those which attempt to duplicate in one test all of the essential activities of the occupation, and (2) those based on a general psychological analysis of the ability and are then translated into test items.

The job in which the aptitude expresses itself may be analyzed either by factor analysis or by simpler means in order to single out its psychological components. Test items are then organized to represent these components. An application of this procedure is found in the so-called "work sample test" or "miniature" tests which are made up of representative activities involved on the job. The test by Munsterberg (68) for motormen is a good example and was also one of the first aptitude tests to be developed. The Minnesota Test for Clerical Workers is another example of a miniature test which is used in the measurement of clerical aptitude (13, p. 215).

An example of a test which is based on a more general psychological analysis is the Seashore Measures of Musical Talent (9, p. 177) which turns entirely on a psychological conception of musical ability and contains no items from musical activity itself.

There are other ways of classifying aptitude or employment tests but for the present purpose it is felt that the foregoing discussion will suffice.

2. Self-Evaluation tests. Self-evaluation tests may be subsumed under two categories, namely personality and interest tests.

Success in an occupation is not solely determined by ability but is in part also attributable to traits of personality and interest. Separation from a job may be due to deficiencies in personality. Therefore, in employment testing it is necessary to consider various methods that have been devised for evaluating these more intangible factors.

Hunt (34) has described the effect of personality in job turnover. He conducted a survey of 76 corporations on reasons why 4,000 employees were either discharged or failed to receive promotion. He found that lack of ability accounted for only 10 per cent of the discharges and 24 per cent of those who failed to receive promotion. The remainder were due to personality deficiencies.

It appears safe to say that, judging from the voluminous amount of work being done by human engineers and psychologists in industry (29, 30, 35, 47, 61) to satisfy employee-employer relations and bring about effective output, personality is more a determiner of success than is intelligence.

However, personality testing is in its infancy and its accuracy is very limited. Gray (26) notes that we are hardly

yet removed from an age of charlatans, when personality was diagnosed by reference to the stars, palm lines, face contours, or cranial protuberances.

Although Hunt (34) noted that personality was responsible for most of the turnover of employees, he also noted that most of the personality traits were of a long-term type which defied discovery by tests or single interview. Such traits include lack of ambition, dishonesty, and so on. However, although present methods of personality measurement are crude, they are the best available, and, when properly interpreted in light of their deficiencies and used in conjunction with all available data, personality measurement data increase the accuracy of human judgement.

Background of personality and interest tests.

(a) Personality tests. As in the case of intelligence testing, World War I brought into prominence the possibilities of group procedures for obtaining personality self-reports. The first noteworthy questionnaire was Woodworth's Personal Data Sheet which was used in processing World War I recruits (72). Following World War I the enthusiasm for mental testing brought also a demand for tests of personality. As a result, a number of instruments for probing personality were drawn up that were adaptations of Woodworth's questionnaire. Each test consisted of a collection of questions which purported to measure a facet of personality.

Laird and Heidbreder (57, p. 299) independently and

simultaneously published an introversion-extroversion inventory. In 1928, the Allports (1) devised the Allport Ascendancy-Submission Scale. A more objective questionnaire was developed in 1930 by the Thurstones (67) who attempted to validate their items by an empirical method. An integration of several questionnaires was effected by Bernreuter (6) in 1931. He attempted to integrate several questionnaires utilizing items from the Thurstone Personality Schedule, the Laird Introversion-Extroversion Questionnaire and the Allports' test. In 1935, Flannagan (18) applied the factor analysis method to the 125 items of the Bernreuter Personality Inventory.

Some of the most recent instruments have placed their emphasis on scores claiming empirical validity, making psychological interpretation of scores a secondary consideration. This is the case with the Minnesota Multiphasic Inventory and the Humm-Wadsworth Temperament Scale (33, 48). In these scales, items were selected, not because they fitted a definition of a trait but because experimental trial showed that particular groups of mental patients give different responses from normal subjects. Scores on these tests indicate how far the subject deviates from normal on such dimensions as "paranoid", "manic", and so on. Both scales employ check scores to identify undependable self-reports and both utilize correction formulas to compensate for a subject's tendency to give a favourable or over suggestible self-report.

(b) Interest tests. As early as the seventeenth

century the educator Comenius pointed out the need of making subject matter interesting (64, p. 100). Later Rousseau and Dewey made interest the corner-stone of their philosophies of education. However, psychologists did not begin studying interests until about the time of World War I. According to Fryer (19) the measurement of interest began in 1913 and 1914 by Kelly. The former developed an inventory to investigate group interests which called for self-estimates by the individual. The scale was intended to predict success in English, History and Mathematics but no follow-up on the validity of the technique has been reported.

The earliest work upon a standard interest inventory was begun in 1919 by Yoakum, Moore and Freyd (7, p. 5) at the Carnegie Institute of Technology. The Carnegie Interest Inventory published in 1921 is the first standardized inventory in which the validity of each item was carefully checked on the basis of its degree of selectivity in terms of groups.

The method of the Carnegie Interest scale was subsequently adapted or modified by Patterson (50), Cowdery (12) and Kornhauser (36) with few innovations. Later on Strong made extensive use of this work but Garretson (7, p. 6) was the first to develop a widely used questionnaire which did not adhere to the pattern of the Carnegie form. He published a preference questionnaire in 1931 which contained 328 items and which attempts to explore nine areas "useful in attaining educational adjustment" in the technical and commercial fields.

The Kuder Preference Record (37) was published in 1939 after Kuder had carried on research over a period of six years. His inventory consists of 504 items, each of them giving a brief description of a certain type of activity. The listed activities are grouped into 168 triads, in each of which the individual selects the most and least liked items. In 1948 Kuder revised his test using the same method of construction and added two more scales.

In 1943 Lee and Thorpe (40) published their Occupational Interest Inventory. The individual selects the more favoured task from each of 120 pairs of listed activities. Although the method had been used earlier by Miner (49), Lee and Thorpe were the first to use the method as the basis for an entire inventory.

Strong first published his Vocational Interest Blank in 1927 and revised it in 1938 (62). It is a superior example of the empirically constructed interest test. At present it consists of 400 items relating to occupations, school subjects, leisure activities, and types of people to whom one responds in terms of liking, indifference, or disliking. The answer to each item is assigned a weight that indicates its significance for a given occupation. The weight was determined by giving the inventory to a sample of men who were successfully engaged in an occupation, and by determining what percentage indicated a liking, indifference, or dislike for each item in the inventory.

Summary of employment testing. From the foregoing we see that employment testing is not primarily involved in measur-

ing specific aptitudes. Brown and Ghiselli (22) point out that a test is a "stimulating situation designed to elicit behavior of a particular sort". However, the final measurement of an individual is not reflected in single or independent traits but rather involves several abilities, depending on the job. Consequently, a test might be designed that draws principally upon personality traits and interests rather than upon one or another of various abilities. Thus, the term "employment testing" involves any test or inventory or groups of tests which contributes to ultimate prediction and selection for a particular job area.

IV. IMPLICATIONS OF WORLD WAR II

During the war years from 1941 to 1945 a new high was obtained in vocational test development. When war broke out the military forces were faced with the biggest personnel problem in their history. The problem was to recruit, classify, train and assign to duty millions of officers and men who varied greatly in mental ability and physical and emotional make-up. In addition to the use of tests for purposes of general classification and selection, new tests had to be devised for the selection of personnel to be trained for the many occupational specialties that comprise a highly mechanized military organization. Careful studies were made to determine the specific factors entering into success in various assignments and tests were developed to measure these factors. A vast amount of data accumulated from such studies and the results were published at the conclusion of the

war.

The U.S. Airforce (73), for example, published nineteen volumes of research descriptive of personnel investigations made by that single military unit. Another contribution to personnel research by the military was an evaluative summary by Stuit (63) of the selection and classification of personnel conducted by the U.S. Navy's Bureau of Personnel.

Research from World War II provided case histories on the applications of modern techniques to complex personnel problems and proved of value not only for military organizations but also in post-war personnel administration in business and in industry.

V. PERSONNEL RESEARCH FOLLOWING WORLD WAR II

Extent of Employment Testing

The successful use of tests by various government services and private industries during World War II added immediate impetus to the testing movement. Employment tests have been devised for every conceivable kind of work. There are tests for semi-skilled workers, skilled workers, salespersons, clerical workers, managerial and supervisory, professional groups, and so on ad infinitum.

Dorcus and Jones (15) in a study of well over 2,100 employment tests have grouped tests under 227 different occupational titles. However, since many of the studies under consideration had to be eliminated because of minimum criterion

standards and since many industrial settings do not publish their research it follows that many more job titles exist than these that have been listed.

Although psychologists have been reporting on the effectiveness of various employment tests since 1908, in only a few instances has any given test been validated for a particular job more than a few times (22, p. 189). An extremely neglected area is that of the job of prison officer.

Personnel Selection of Prison Officers

Very little scientific experimentation has been attempted on selection of prison officers. J. D. Klinger in New Horizons in Criminology is reported as saying that in 1941 in the United States only ten states selected and employed prison guards by means of either civil service examinations or some other merit system (4, p. 658). Lundberg (41), in the only other survey published in the literature, found that eighteen states made some provision for selecting prison officers by some form of merit system. Several of the states at that time had not yet instituted competitive examinations because it was necessary to hire all applicants. However, it should be noted that the civil service examination does not in itself guarantee adequate standards. A careful job analysis is essential to the development of adequate standards and, outside of the practices in seventeen states, practically no standards have been established for the position of prison officer. Lundberg came to this conclusion :

methods of selection of the Prison Guard are generally

loose and have little experimental study of validity. Of the some 13,000 guards in this country, it is safe to say that over three fourths have been selected by unscientific methods....Several states "validate" their examination by providing a period during which the candidate is in a probationary status - only New Jersey has conducted an experimental validation of their selection methods. (41, p. 179)

The retardation of the scientific selection of prison officers has been due largely to public indifference to the plight of the inmate once he has been sentenced. The purpose of a prison has been looked upon as affording some protection to society or in terms of societal revenge. However, in recent years there has been in Canada, as in other countries, a change in the approach to the treatment of individuals sentenced to imprisonment in penal institutions. Rehabilitation has become a key word. This term refers to those processes of learning or inhibition which may change the personality of the inmate in such a manner that upon release he will conform to the legal requests of society.

Qualities Requisite for Effective Prison Officers

Taft (66, p. 438) aptly describes the guard acceptable in the old-line prison. A guard had to be

physically strong, alert, emotionally dull, courageous, obedient, not averse to monotonous routine, not too intelligent or critically minded, ignorant of social science and amenable to political discipline if otherwise reasonably honest.

However, a prison officer working under a contemporary constructive program has to be

firm and courageous and loyal to the administration,

but their loyalty must be based upon voluntary and enthusiastic cooperation in a piece of difficult but absorbing constructive work. Neither tough men nor tender-hearted sentimentalists will do, but rather men capable of intelligent decisions and with the ability to carry them out by force when necessary, but always by persuasion when possible. They must be examples of self-control, impartial, consistent, objective. They must inspire confidence in men who are prejudiced against them. They must preserve amid discouraging difficulties the belief that some prisoners' attitudes may be changed and that the effort to change them is worth while. (66, p. 438)

It can be seen from the foregoing description that a "modern" prison officer must have qualities which have not characterized the guard of the past.

It is obvious that prison officers of high quality are requisite to a successful prison treatment program. In fact, without satisfactory prison officer standards any correction rehabilitation program is ultimately doomed to fail. To emphasize this last point even more, let us note a quotation by a New York State prisoner.

I have all respect for doctors, the social workers, teachers and other workers in the institution and I believe they do a good deal of good, but the greatest influence that can effect a man while in prison is his respect for someone on the prison staff whom he is under. I met such a man in Officer Blank of the Michigan State Prison. My great respect for him and the influence he exerted on me completely reformed my life and I feel that it will continue to exert its influence for my entire future. (66, p. 45)

However, one need hardly mention that establishment of a selection program that is designed to yield a high-grade officer candidate cannot be accomplished unless the salary for such work is commensurate with the abilities of the men hired. The fact that salaries have been inadequate explains why individuals of

good ability have not been attracted to prison work and is in part the reason why it has been difficult to deal with this problem along scientific lines.

Up to 1946 only two states, Michigan and New Jersey, regularly used a standardized mental test in selecting guards (41, p. 143). Hubbard (31) found that objective employment tests used for selecting guards correlated about .70 to .80 with the Army Alpha test. However his study was limited to measuring skills for specific technical jobs such as "painter" and so on, rather than to using the tests in selecting all guard applicants.

So far as the writer has been able to discover in the literature, there have been no attempts to investigate either interests or personality in prison officer candidates other than by interview. It has already been pointed out that interest and personality inventories generally do not yield precise predictions. However, a study of such factors seems necessary in order to determine the principal trends of interests and personality which might prove of value when used as a part of a battery of tests. The purpose of this thesis is to investigate the problem of the selection of prison officers in terms of trends that might be gleaned from a use of personality, interest and intelligence tests.

CHAPTER III

TESTS USED

I. PERSONNEL CLASSIFICATION TEST - FORM A

A. Reason for Selection

(1) Ease of administration and scoring. Directions are simple and easily understood. The test can be conveniently scored in about two minutes.

(2) Moderate time limits. There are separate time limits for the two sections of the PCT. Part I (verbal) requires eighteen minutes and Part II (numerical) ten minutes.

(3) It was standardized with groups of comparable ages and education.

(4) The test is essentially a measure of power rather than speed.

(5) The test was intended primarily for industrial and business use.

(6) The form of the test permits use of a wide variety of subject matter and a consequent reduction of emphasis on mere vocabulary knowledge. This was thought to be important since the minimum educational requirement for a prison officer applicant is Grade VIII.

(7) The name of the test was also taken into consideration to lessen the uneasiness that might result from taking a "mental ability" or "intelligence" test.

(8) The general quality of the technical data supplied in the manual is good. Moreover, normative data is provided for various employment groups and detailed description for each of seventeen occupational groups is also provided.

B. Structure of the Test

The Personnel Classification Test is a short general ability test which yields three scores: a verbal, numerical and a total score. Part I consists of forty multiple-choice verbal analogies items in which the first and fourth terms of the analogy must be chosen with the second and third terms given. According to the author of the test (71) the type of item used to measure verbal reasoning ability was designed so that reasoning through analogy and perception of relationships are needed to respond to each item.

Part II contains twenty arithmetic computation items which were devised to test command of basic arithmetic skills and processes plus general facility in the use of numerical concepts. The content has been so arranged that a premium is placed on one's ability to perceive relationships and to operate with ingenuity while the importance of sheer figure manipulation or number perception, which are better measured by simple clerical tests, is minimized.

II. THE KUDER PREFERENCE RECORD - VOCATIONAL

A. Reason for Selection

(1) This test is characterized by simplicity of adminis-

tration. The Kuder Form CH is self-administering. Directions explaining how to mark the answers are given in each test booklet and are clear and easy to understand. The blank can be given either individually or to groups and the tester need give very little assistance.

(2) There are also moderate time limits in administering and scoring. The Kuder requires about thirty to thirty-five minutes to administer and scoring can be completed in five to six minutes. The ease of scoring was the main reason for selecting this inventory rather than the Strong Vocational Interest Blank since scoring the latter on all possible occupations requires a great deal of clerical work.

(3) Extensive research has been done in the Kuder and it owes much of its success to the research that preceded its introduction to the test market.

Kuder began work on his test in 1934 (37) and it was not published until after six years of research and revision. Kuder attempts to identify certain generalized activity patterns which are psychologically meaningful. According to Brayfield (9, p. 640) this approach to the measurement of interests is in line with factor analysis studies of the problem and is consistent with the "pattern analysis" interpretation of other interest inventories.

(4) The Kuder employs, also, a validity or distortion scale which makes it possible to isolate those individuals who are careless in their answering or who deliberately try to put

themselves in the most favourable light.

(5) It has reasonable reliability. The 1953 revision (39) of the manual gives separate reliability estimates for each of the ten scoring categories based on four different groups. The number of individuals in each group is as follows: 1,000 men, 100 women, 100 boys, 100 girls. On the average, the reported reliabilities are approximately .90, the lowest being .84 and the highest .93.

B. Structure of the Inventory

The testee is presented with 169 groups of activities. in each group there are three activities and the subject is required to rank them in order of preference. From the three items, he selects the most and the least liked activities. This is done by pin perforations made in a record blank consisting of stencils, which facilitate scoring. The various scales contain a different number of items, ranging from 69 on the musical scale to 210 on the persuasive scale.

Nine fields of interest are studied and these are listed on the profile sheet (38) as follows:

0. Outdoor interest means that you prefer work that keeps you outside most of the time and usually deals with animals and growing things. Forest rangers, naturalists, and farmers are among those high in outdoor interests.
1. Mechanical interest means you like to work with machines and tools. Jobs in this area include automobile repairmen, watchmakers, drill press operators, and engineers.
2. Computational interest means you like to work with

numbers. A high score in this area suggests that you might like such jobs as a bookkeeper, accountant, or bank teller.

3. Scientific interest means that you like to discover new facts and solve problems. Doctors, chemists, nurses, engineers, radio repairmen, aviators, and dieticians usually have high scientific interests.
4. Persuasive interest means that you like to meet and deal with people and to promote projects or things to sell. Most actors, politicians, radio announcers, ministers, salesmen, and store clerks have high persuasive interests.
5. Artistic interest means you like to do creative work with your hands. It is usually work that has "eye appeal" involving attractive design, colour, and materials. Painters, sculptors, architects, dress designers, hairdressers, and interior decorators all do "artistic" work.
6. Literary interest shows that you like to read and write. Literary jobs include novelist, historian, teacher, actor, newsreporter, editor, drama critic, and book reviewer.
7. Musical interest shows you like going to concerts, playing instruments, singing, or reading about music and musicians.
8. Social Service interest indicates a preference for helping people. Nurses, Boy or Girl Scouts leaders, vocational counselors, tutors, ministers, personnel workers, social workers, and hospital attendants, spend much of their time helping other people.
9. Clerical interest means you like office work that requires precision and accuracy. Jobs such as bookkeeper, accountant, file clerk, salesclerk, secretary, statistician, and traffic manager fall into this area.

III. MINNESOTA MULTIPHASIC PERSONALITY INVENTORY TEST - (GROUP FORM)

A. Reasons for Selection

- (1) Use of validity scales. As an inventory-type test,

the MMPI has an advantage over other inventories in that it attempts to measure the validity of the test for the individual by showing whether he is taking the test seriously and honestly giving his opinion. The validity indicators are provided on four scales: The Question score, the Lie score, the Validity score, and the K score. The Question score simply records the number of "?" answers and a high score is taken to mean that the true scores on the diagnostic categories would probably be further away from the mean than they are. The Lie score (L) is made up from the answers to a number of questions which make the subject appear in an unfavourable light; subjects claiming the favourable alternative are presumed to have falsified their diagnostic scales also in the direction of greater favourableness. The Validity scale (F) consists of items which are infrequently answered by either normal or abnormal subjects; high scores usually indicate that great care is necessary in interpreting the record. The K scale acts as a suppressor variable and is claimed to sharpen the discriminatory power of the diagnostic scales.

(2) Standardization has been done carefully and conscientiously.

(3) The statistical work is of high quality.

(4) Adaptability to new scales. New diagnostic scoring categories can be added without necessitating a new set of questions. All older records may be scored on any new key. Consequently, the test is a good source of research material and as

new scales are developed standardization can be undertaken on the groups of previous studies. In the present study five additional scales were added to supplement the regular diagnostic scales: Dominance scale (Do); Responsibility scale (Re); Hostility scale (Ho); Social-Economic Status scale (St). These five scales were all developed on normal subjects.

(5) Ease of administration and scoring. Booklets are re-usable. Blank forms are supplied on which a subject working from the booklet can record his decisions. The instructions are easy to understand and little assistance need be given to administer the test. Scoring is made as mechanical as possible and it is thus suitable for regular and routine testing.

B. Structure of the Inventory

The inventory consists of 550 statements which the subject indicates as being true or false or uncertain with respect to himself. Different areas of life experience are covered by the items such as family relations, mood-tone, beliefs, somatic experiences, and so on. The inventory attempts to measure specific clinical syndromes and does not merely attempt to determine whether or not the subject is neurotic, as has been the case with some of the earlier schedules. On the basis of the performance of patients in the various psychiatric groupings, scoring scales based on 351 of the items have been constructed for the following personality trends: (a) Hypochondriasis; (b) Depression; (c) Hysteria; (d) Psychopathic Deviate; (e) Masculinity-Femininity; (f) Psychasthenia; (g) Paranoia; (h) Schizophrenia;

(i) Hypomania. The raw scores of the test are converted into standard scores from which a profile for the subject is made. The standard scales have a common mean of 50 and a standard deviation unit of 10 points. Accordingly, standard scores above 70 would represent the presence of an abnormal amount of the component in question.

A special feature of the test that differentiates it from other inventories is the validity score mentioned above (see section A).

IV. THE MANSON EVALUATION

A. Reasons for Selection

(1) Time consumption in scoring and administration. The test can be completed in fifteen to twenty minutes and scored in about four to five minutes.

(2) It seemed advisable to investigate the sample of prison officers for potential alcoholic tendencies.

(3) To provide a short questionnaire for comparison with the MMPI. The seven traits measured by the Manson Test are clinically comparable to some of the longer MMPI scales. Hence, it was decided to include this test as a supplementary test and determine what relationship it might have with the MMPI (see section B; description of traits).

B. Structure of the Manson Evaluation

The Manson test consists of 72 items which purport to differentiate alcoholics from non-alcoholics. The subject

responds to each question by answering either "yes" or "no".

According to the author (46) the Manson Evaluation was designed to:

(1) Identify individuals whose behavior and personality structure indicated they were alcoholics or had serious alcoholic problems; (2) identify nonalcoholic individuals with personality characteristics often found in alcoholics. Perhaps such individuals would become alcoholic if placed under certain conditions of stress....

The author (45) reports that a subjective analysis of the seventy-two questions resulted in establishing seven neurotic or psychopathic traits. The traits listed in the manual (46) are as follows:

- A. Anxiety. High scores would indicate an excessive number of fears, worries, feelings of insecurity and inadequacy, undue concern over health, easily fatigued.
- B. Depressive Fluctuations. High scores would indicate easily depressed, sadness, frequent mood swings toward depression, prone to quick disappointments.
- C. Emotional Sensitivity. High scores would mean extreme emotional sensitivity with inability to make satisfactory social or emotional adjustments; extreme lability with poor defenses; touchiness.
- D. Resentfulness. High scores would indicate strong and bitter feelings of resentment toward society and individuals; easily irritated; carries chip on shoulder; paranoid ideas.
- E. Incompleteness. High scores would indicate a series of failure to complete commonly accepted social objectives, such as education, work mastery, steady employment, marital adjustments, community, participation, religion, unsteadiness, mobility, and frequent change.
- F. Aloneness. High scores would indicate feelings of being alone in the world, isolated, unique, unwanted, undersocialized, feeling as if there were a barrier between the individual and the world or society.

- G. Interpersonal Relations. High scores would mean lack of close personal and emotional ties, poor family relations, parental rejection, unhappy childhood, lack of real friends, shallow emotional relationships.

CHAPTER IV

THE RESEARCH DESIGN

I. THE SAMPLE

Selection of Sample

The practical aspects of obtaining the sample became important determiners in the final selection of the sample. To secure a satisfactory sample it was hoped that the following criteria could be utilized.

(1) There must be good liaison between the experimenter and officials at Oakalla and good rapport with the prison officers.

(2) The facilities for testing would have to meet a reasonable standard of comfort, quiet and lighting. Moreover, time would have to be set aside during regular working hours to test the men.

(3) The sample must be of reasonable size so as to insure statistical significance in the results.

(4) The sample must be taken from men on the job who have finished at least a three-month probation period in order to make it possible to collect the required proficiency ratings.

(5) The sample must represent equal numbers of men from four different working shifts.

(6) There must be a suitable criterion to differentiate the sample into upper and lower levels of proficiency.

Translating these criteria into action was by far the most difficult part of the whole study.

In order to understand the job of prison officer and establish rapport with prison officials the experimenter hired on as a Prison Officer in November of 1953. Because of attendance at university lectures, only the afternoon shift (4.00 to 12.00 midnight) and the "graveyard" shift (12.00 midnight to 8 a.m.) were available for purposes of this study.

This employment proved invaluable for obtaining the job description data which facilitated the selection of tests and in "learning one's way" around a prison setting. Fellow officers knew that the experimenter was in no way connected with the "administration". This fact in itself helped considerably in establishing the required cooperation when the actual testing got under way.

The variable of the random sampling of different working shifts seems not to have been too well controlled since most testing had to be carried out during the evening shift. However, this problem was overcome to some extent by the fact that it took considerable time to gather the test data and that shifts automatically changed every three months, i.e., the day shift changed to the afternoon shift, the afternoon to the night shift, and so on.

Description of Obtained Sample

The sample for this study was made up of 100 employed prison officers who constituted approximately 33 per cent of the

total population of employed guards. Several criteria concerning the composition of the sample should be noted.

(1) Each guard had undergone a five-day orientation course in basic training.

(2) All guards were in "good" health. A physical examination by the Prison Physician is a prerequisite for employment.

(3) All guards had no civilian crime record.

(4) All guards had at least three months of on-the-job experience and in most instances over six months' experience. This specification was adhered to in order to ensure getting a more valid rating of performance. Many men who were obviously unfit for work as a guard were either eliminated or dropped out of their own accord by the time the probationary period was over.

Statistical analysis. The sample was appraised statistically for age and formal education level.

(1) Age distribution of total sample. Table I shows that the distribution of ages for the total sample ranged from age 21 to 43. The mean age for the total sample was 28.9 with a S.D. of 5.57. The median age was 29. This table presents the frequency of the distribution in terms of percentages. Analysis of the table shows that 89 per cent of the total sample was distributed between the ages of 21 and 35, and 55 per cent of the sample fell within the 21 to 29 age range. Only 4 per cent of the sample was over the age of 39.

(2) Age distribution of good and poor groups. Table I

TABLE I

CHRONOLOGICAL AGES, MEANS, MEDIANS AND STANDARD DEVIATIONS FOR TOTAL SAMPLE
AND SUCCESSFUL AND UNSUCCESSFUL GROUPS OF PRISON OFFICERS

Sample		Age												Mean Age	Median Age	S.D.
		21 to 20	23 to 22	25 to 24	27 to 26	29 to 28	31 to 30	33 to 32	35 to 34	37 to 36	39 to 38	41 to 40	43 to 42			
Percentage of Total Sample (N = 100)		2	21	14	8	10	12	10	12	2	5	2	2	28.9	29.0	5.75
Successful Group (N = 27)	N	0	4	3	2	5	3	4	4	0	2	0		29.1	30.0	4.72
	%	0	15	11	7	19	11	15	15	0	7	0				
Unsuccess- ful Group (N = 27)	N	1	8	4	2	3	3	2	0	2	2	0		27.6*	27.0*	5.34*
	%	4	31	15	7	11	11	7	0	7	7	0				

* Difference between two groups not
significant.

also shows the mean and median ages, standard deviation and frequency distribution (in per cents) for the successful and unsuccessful groups. The successful group has a distribution of ages ranging from age 22 to 39 and the unsuccessful group from age 21 to 38. The mean age of the successful group was 29.1, the median age 30, and the S.D. was 4.72. The mean age for the unsuccessful group was 27.6, the median 27 and the S.D. was 5.34. There was no significant difference between the two groups either in mean scores or variability.

(3) Education level of the total sample of prison officers. Table II shows that the school grade distribution, expressed in percentages of the total sample, ranged from Grade 6 to 13. The mean school grade level for this sample was 9.4, the median grade 9, and the standard deviation 1.52. Analysis of the frequency distributions in Table II shows that 92 per cent of the guards fall between Grades 8 and 12, and 83 per cent between Grades 8 and 11. Only 6 per cent of this sample finished Grade 7 or below, while 9 per cent finished Grade 12 or had higher education.

(4) Educational differences between the groups of successful and unsuccessful guards. Table II also shows that the mean grade for the successful group was 9.30, the median, 9.00, and the standard deviation 1.34, whereas the mean of the unsuccessful group was 9.40, the median 9, and S.D. 1.64. Very little difference was found in variability.

An analysis of the frequency distribution indicates

TABLE II

EDUCATIONAL LEVEL, MEAN GRADE, MEDIAN GRADE, AND STANDARD DEVIATIONS
FOR TOTAL SAMPLE AND SUCCESSFUL AND UNSUCCESSFUL GROUPS OF PRISON OFFICERS

Sample		Last School Grade Completed								Mean	Median	S.D.
		6	7	8	9	10	11	12	13			
Percentage of Total Sample (N = 100)		2	4	25	26	28	8	9	2	9.4	9.0	1.52
Successful Group (N = 27)	N	0	9	9	5	1	3	0		9.3	9.0	1.34
	%	0	33	33	19	4	11	0				
Unsuccessful Group (N = 27)	N	1	8	8	7	1	1	1		9.4	9.0	1.64
	%	4	29	29	26	4	4	4				

that 87 per cent of the successful guards fall between Grade 8 and 11 while 88 per cent of the unsuccessful guards fall within the same range.

It can be seen from the analysis of the two groups that the samples are quite similar in regard to age and education.

II. THE CRITERION

As Davies notes (14), "In real life situations you have to take the criteria which you can get despite its limitations."

A great many studies have listed the pros and cons of merit testing, and the concensus seems to lie against traditional rating procedures such as the graphic rating scale, the check list, the chart method and so on. The trend in some settings seems to be towards more scientific and objective measures of performance (59) (43) (8) (17).

The method in current use at Oakalla for evaluating or rating a prison officer is the chart method. This method consists of rating an individual in terms of the following qualitative scale, i.e., whether poor, fair, good, or excellent on such variables as dependability, conscientiousness, and so on. The difficulty with all such methods (54, p. 35) is their subjective nature. Little consideration is given to whether the job factors one selects are actually involved in job performance. In most cases no experimentation is undertaken. One may question whether the factors in job performance that are outlined by conference procedures are in fact the most important and the most directly

identifiable. For example, it is easy to demand a quality such as "maturity" in a person, without making a thorough analysis of the trait to arrive at a common understanding of what the term means.

Criteria used in this Study

It was decided to rate this sample by the forced distribution method. This method is probably the simplest of all rating methods since the officers are rated only on one characteristic, viz., that of job performance. Its use is justified by the fact that very extensive statistical work (68, p. 17) with other more complicated systems has shown that it gives similar results with even greater reliability. According to conventional practices a five-point job-performance scale was used based on the normal curve; one end of the scale represents best performance, the other end poorest performance. Officials were asked to allocate 7 per cent of the men in the sample to the best end of the scale, 20 per cent in the next category, 46 per cent in the middle bracket, 20 per cent in the bracket next to the lower end, and 7 per cent in the end. No descriptive phrases were included on the scale.

Because of the homogenous nature of our total sample only the extremes of the sample were used for comparing the means of "good" guards and "poor" guards. The middle group, consisting of about 46 per cent of the sample, was omitted.

Other Criteria

There are many reasons why an officer may be classified

as "good" or "poor". He may have little interest in his work, he may have inferior intelligence or various personality characteristics which would handicap him in "handling" inmates. Therefore, each guard was to have been rated on a number of specific variables by the "chart method" in order to provide a more specific understanding as to why an individual was rated either "good" or "poor" on the forced distribution scale.

The rating form used was devised using the conference method procedure - the very method which has already been discussed above for its lack of reliability. The only defence for the use of this method in this instance was the lack of time to develop a more objective instrument.

The new rating form consisted of 14 critical areas considered by top prison officials to be necessary for success. Actually, however, 25 variables were first offered but, owing to conflicting definitions, the list was reduced to 14 variables.

This part of the project proved to be very time consuming. Moreover, because of administrative duties, Oakalla officials could not rate all of the men in the sample in time to be included in this study.

III. ADMINISTRATION OF TESTS

Approximately 100 trips to Oakalla were required to complete all of the testing. Because of the general nature of the prison setting and the necessity to maintain security, the number of men tested at any one time was limited to two or three.

Hence, it took 12 months to gather the test data. The testing began in April 1954 and was completed in April 1955.

Because of the crowded facilities and the shortage of man power, every officer at Oakalla is kept rather busy. The total time that one might get off during the day was either an hour at noon or perhaps an hour in the afternoon. The evening shift from 9.00 p.m. to 11.45 p.m. proved the most favourable time for the testing since the inmates are locked up for the night at 9.00 p.m. It was difficult to do any test earlier than 9.00 p.m. since a good part of the "treatment" program is under way after supper and every officer has a special role to play in it.

While the battery was completed by fast readers in about two-and-a-half to three hours, a large number of officers averaged around four hours. Consequently, in most instances the testing took two sessions.

For the most part the testing was carried out in the New Gaol in a warm, well lighted section that was used primarily as a gymnasium. However, this location was not the most ideal since the radio was not turned off until 10.00 p.m. and thus proved a distraction for certain individuals. Moreover, the room often became uncomfortable when dust particles were stirred up after the gym had been used for an evening's participation.

Later the testing was conducted in the classroom at the Main Gaol. Thirty-three men were tested there and conditions were good. The testing of each group was prefaced by a brief informal explanation of the purpose of the study, viz., that it

was desired to obtain an average for all prison officers so as to be able to set up minimum requirements that would eliminate certain applicants before coming on the job. Any questions that were asked about the study were answered, and the testing followed.

The standardized instructions, that are provided for the Kuder, MMPI, Manson Evaluation, and the Personnel Classification Test, were followed exactly.

No standard order in the administration of tests was followed. Because of the time factor tests were given which could be finished in the allotted time. Usually and where possible the Personnel Classification Test was given first since it is a timed test. Then followed the Kuder and Manson Evaluation. The MMPI was administered mostly by itself during the second testing session since it took the longest to complete.

IV. STATISTICAL PROCEDURES

The data were first organized by tabulating the test scores for each individual. Opposite each name there were 35 test variables. The possibilities for studying these data were many. However, the statistical procedures finally decided upon were selected in the light of practical considerations and the ease with which the results could be adapted to the actual practical setting. Each test was analyzed individually and the investigation covered the following:

- (1) Computation of a mean profile for the total sample on every test and subtest.

(2) Determination of mean scores for the upper and lower levels of prison officers and examination of these data for significant differences between the two groups.

(3) Computation of biserial correlations from widespread classes for all test and subtest variables based on a formula from Peters and Van Voorhis (53, p. 384).

(4) Statistical analysis of standard deviations for significant differences between groups of good and poor prison guards using formula 36 in Garrett (21).

(5) Scattergram analysis to determine possible critical scores.

(6) Special interpretive analysis depending upon the test under consideration. For example computing percentile ranks for raw scores on the Personnel Classification Test and computing product moment correlations between several MMPI and Manson Evaluation traits and so on.

CHAPTER V

RESULTS

I. PERSONNEL CLASSIFICATION TEST

Mean of Total Sample

Table III shows the means, medians and standard deviations for the total sample of the three parts of the PCT. The mean, median and S.D. for the total score was 22.47, 23.00 and 6.35, respectively.

The mean for Part I was 15.70, the median 16.00, the S.D. 6.40; and for Part II 6.70, 7.00 and 3.47, respectively.

TABLE III

MEANS, MEDIANS AND STANDARD DEVIATIONS OF 100 PRISON OFFICERS
ON THE PERSONNEL CLASSIFICATION TEST

Part	Mean	Median	S.D.
Part I (verbal)	15.70	16.00	6.40
Part II (numerical)	6.70	7.00	3.47
Total	22.47	23.00	8.46

Means and Differences Between Means of Successful and Unsuccessful Officers

Table IV shows that both the Total Score and Part I (verbal) on the Personnel Classification Test significantly differentiate the two sample groups at the .01 level. The

TABLE IV
MEANS, DIFFERENCE BETWEEN MEANS, T-RATIO AND BISERIAL CORRELATIONS
BETWEEN UPPER AND LOWER LEVEL GROUPS OF PRISON OFFICERS
ON THE PERSONNEL CLASSIFICATION TEST

Part	Successful Group Mean	Unsuccessful Group Mean	Difference	T-Ratio	R-Bis
Part I (Verbal)	19.62	12.40	7.22	4.60*	.46#
Part II (Numerical)	7.11	5.81	1.30	1.47	.15
Total	26.74	18.22	8.52	4.00*	.41#

* Significant at the .01 level.

Significantly greater than zero.

biserial correlation for the Total Score was .41, for Part I (verbal) .46, and for Part II (numerical) .15.

Difference in Variability Between Upper and Lower Levels of Guards

Table V shows that the two groups do not differ significantly in variability.

TABLE V

THE RELIABILITY OF DIFFERENCES BETWEEN STANDARD DEVIATIONS
ON TWO GROUPS OF PRISON OFFICERS
ON THE PERSONNEL CLASSIFICATION TEST

Part	Successful Group Mean	Unsuccessful Group Mean	Difference	T-Ratio
Part I (Verbal)	4.38	6.35	1.52	1.38
Part II (Numerical)	2.79	3.57	.78	1.24
Total	6.12	8.98	2.86	1.91

Critical Scores

A scattergram was drawn up representing the "good" and "poor" guards and a cutoff score was determined. Thirteen was found to be the best cutting score for separating "good" prison guards from "poor" prison guards. For predictive purposes, men scoring 13 and above may be considered potentially successful prison guards, whereas men scoring 12 or below may be considered poor risks.

Intercorrelations

Intercorrelations between total score and the two parts were computed in order to determine the extent to which the subtests were independent. Table VI shows that the correlation coefficient between the Verbal and Numerical subtests was .43. The Verbal and Total Test Score correlated .87 while the Numerical and Total Test Score correlated .35.

TABLE VI
INTERCORRELATIONS BETWEEN TOTAL AND SUBTEST SCORES ON
THE PERSONNEL CLASSIFICATION TEST

	Verbal	Numerical	Total
Verbal		.43	.87
Numerical			.35

Additional Data

In order to make the data more useful for actual selection and for further research, a table of percentile norms were calculated for the Verbal subtest. These data are presented in Table VII.

Summary

1. Part I of the Personnel Classification Test differentiated between the group of successful guards and unsuccessful guards significantly at the .01 level. Total Score also differentiated between the two groups at the .01 level but the degree of association as measured by the biserial correlation was .41 for the Total Score

TABLE VII

NORMS FOR A SAMPLE OF PRISON OFFICERS
ON THE PERSONNEL CLASSIFICATION TEST:
RAW SCORES WITH PERCENTILE EQUIVALENTS

Verbal Score	Percentile
40 - 35	
34	
33	
32	99
31	99
30	99
29	98
28	97
27	97
26	97
25	93
24	89
23	87
22	85
21	81
20	76
19	72
18	69
17	65
16	57
15	55
14	42
13	36
12	31
11	24
10	18
9	15
8	12
7	9
6	7
5	6
4	4
3	2
2	1
1	1

versus .46 for Part I alone. Part II (numerical) does not differentiate between the two groups and in this study tends to decrease the Total Score predictions.

2. There was no significant difference in variability between the two groups.

3. Scattergrams showed that the distribution was normal and linear for the total sample and between successful and unsuccessful groups of prison guards.

4. A critical score of 13 was determined on the Verbal subtest. Had such standards been available at the time of hiring and had the optimum cutting score been used, 58 per cent of the "poor" guards would have been rejected at the time of their application. At the same time, only 4 per cent of the "good" guards would have been eliminated. These relationships are presented graphically in Figure I.

5. The correlation coefficient between Verbal and Numerical abilities is low enough (.35) to suggest that these abilities are sufficiently independent to warrant measuring them separately in those situations where both qualities are considered critical requirements. This correlation is similar to that reported by the author of the test (71). However, it is evident that Verbal ability contributes most to the total Score since the intercorrelation between Total Score and Verbal was .87 versus only .35 for Total Score and Numerical.

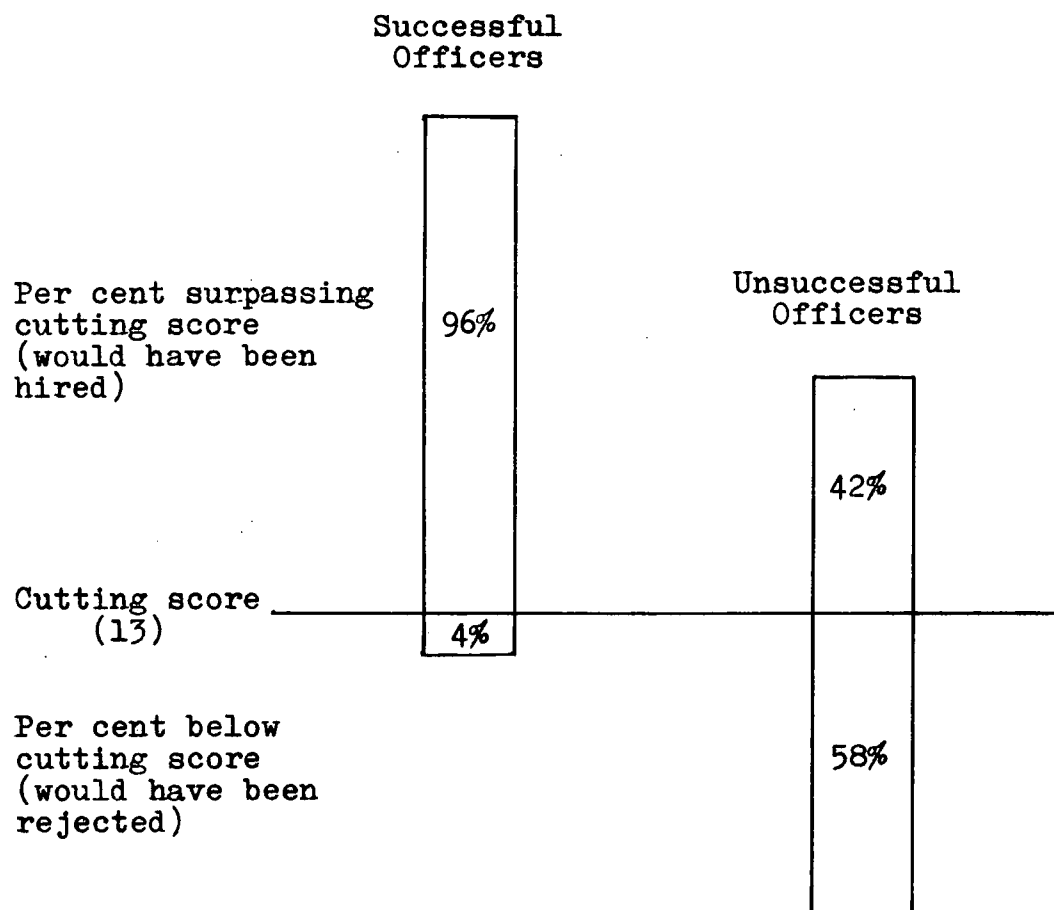


Figure I. Differentiation achieved between successful prison officers and unsuccessful prison officers using Part I (verbal) of the Personnel Classification Test.

II. KUDER PREFERENCE TEST

Mean Profile of Total Sample

Figure II shows in graphic form the mean Kuder profile for a sample of 100 prison officers. The percentile scales of the published norms (38) were employed as a basis for plotting the profile. Scale eight, the Social Service scale on the graph, shows that in general prison officers have more preference for working with and helping people than the average man. There were no other mean scores above the seventy-fifth percentile of the published norms. Table VIII shows the mean profiles and standard deviations in raw score form.

Comparison of Mean Scores Between "Good" and "Poor" Prison Officers

Only three of the scales yielded results that make it possible to differentiate between the two groups. Table IX shows that successful guards have significantly greater scores on the Social Service scale. The Social Service scale difference shows a t-value of 4.71 significant at the .01 level of confidence, and the Computational and Clerical scale differences produced t-values of 2.01 and 2.58 respectively, both significant at the .05 level. The Kuder test profiles for the two groups are depicted graphically in Figure III.

Table IX also shows the degree of relationship expressed as biserial coefficients for each scale in relation to the successful and unsuccessful prison officer categories. The Social Service scale yielded the highest biserial correlation, .49; next

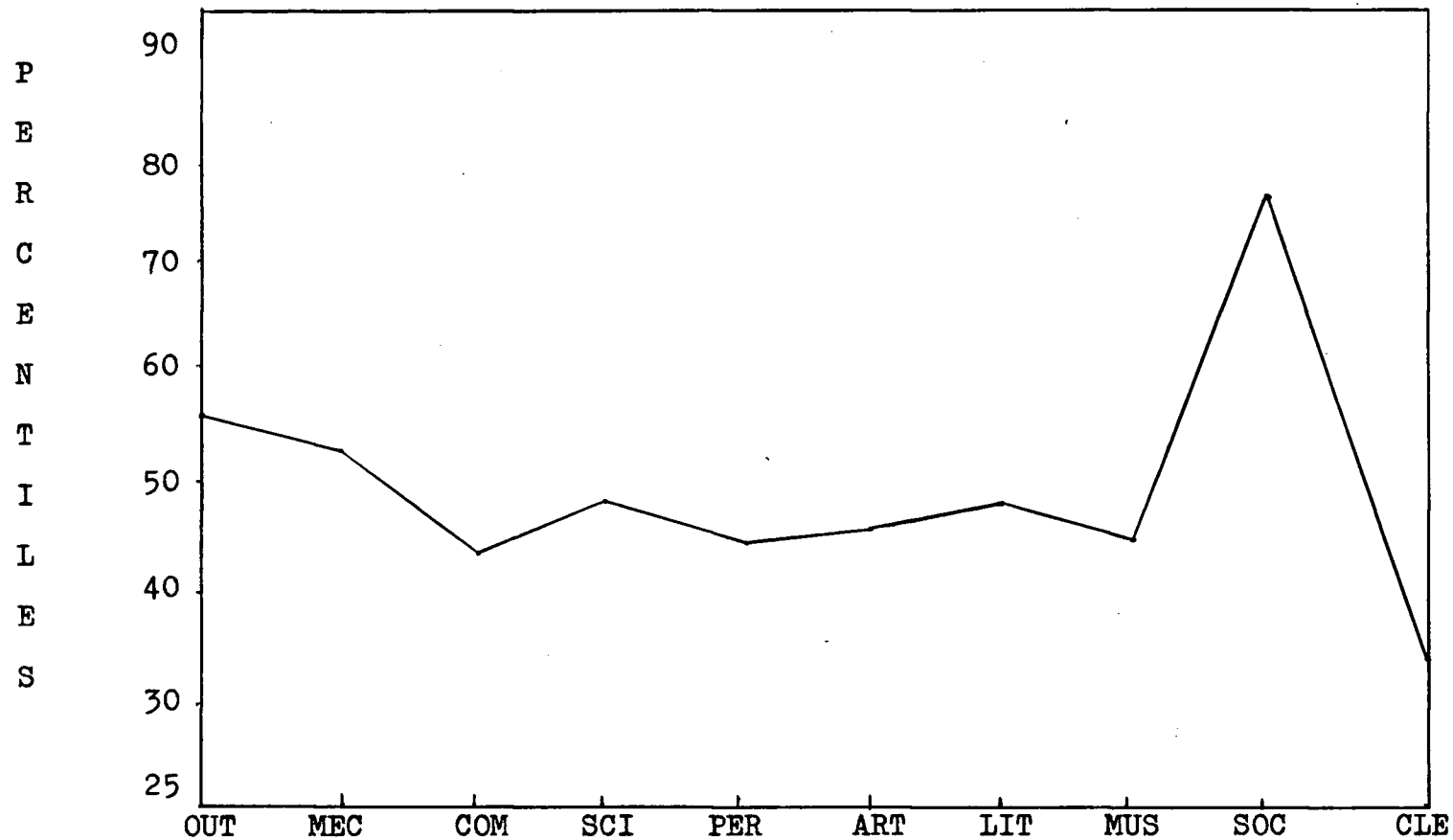


Figure II. Percentile ranks of mean scores obtained by 100 prison officers on the Kuder Preference Record - Vocational (percentiles from norms supplied by test publisher).

TABLE VIII

MEANS AND STANDARD DEVIATIONS OF TOTAL SAMPLE OF PRISON OFFICERS
ON THE KUDER PREFERENCE RECORD - VOCATIONAL

Scales	0 Out	1 Mec	2 Com	3 Sci	4 Per	5 Art	6 Lit	7 Mus	8 Soc	9 Cle
Mean	45.55	45.95	26.20	39.50	37.60	21.65	18.93	10.71	51.10	40.95
S.D.	14.60	10.83	7.89	8.90	11.17	7.65	7.19	10.11	14.65	10.30

TABLE IX

MEANS AND DIFFERENCES OF THE MEANS FOR SUCCESSFUL GUARDS AND UNSUCCESSFUL GUARDS
ON THE KUDER PREFERENCE RECORD

Kuder Scales	Successful Guards Mean	Unsuccessful Guards Mean	Diff.	T-Test	R Bis
Outdoor	48.22	42.48	5.74	1.48	.16
Mechanical	47.70	44.77	2.93	1.02	.11
Computational	24.74	28.74	4.00	2.01*	.21#
Scientific	39.92	41.11	1.19	0.48	.06
Persuasive	38.37	37.18	1.19	0.39	.04
Artistic	20.55	23.81	3.26	1.69	.18
Literary	19.11	21.00	1.89	0.99	.11
Musical	8.74	10.55	1.81	1.01	.12
Social Service	58.88	40.81	18.07	4.71**	.49#
Clerical	38.11	44.26	6.15	2.58*	.24#

* Significant to the .05 level.

** Significant to the .01 level.

Significantly greater than zero.

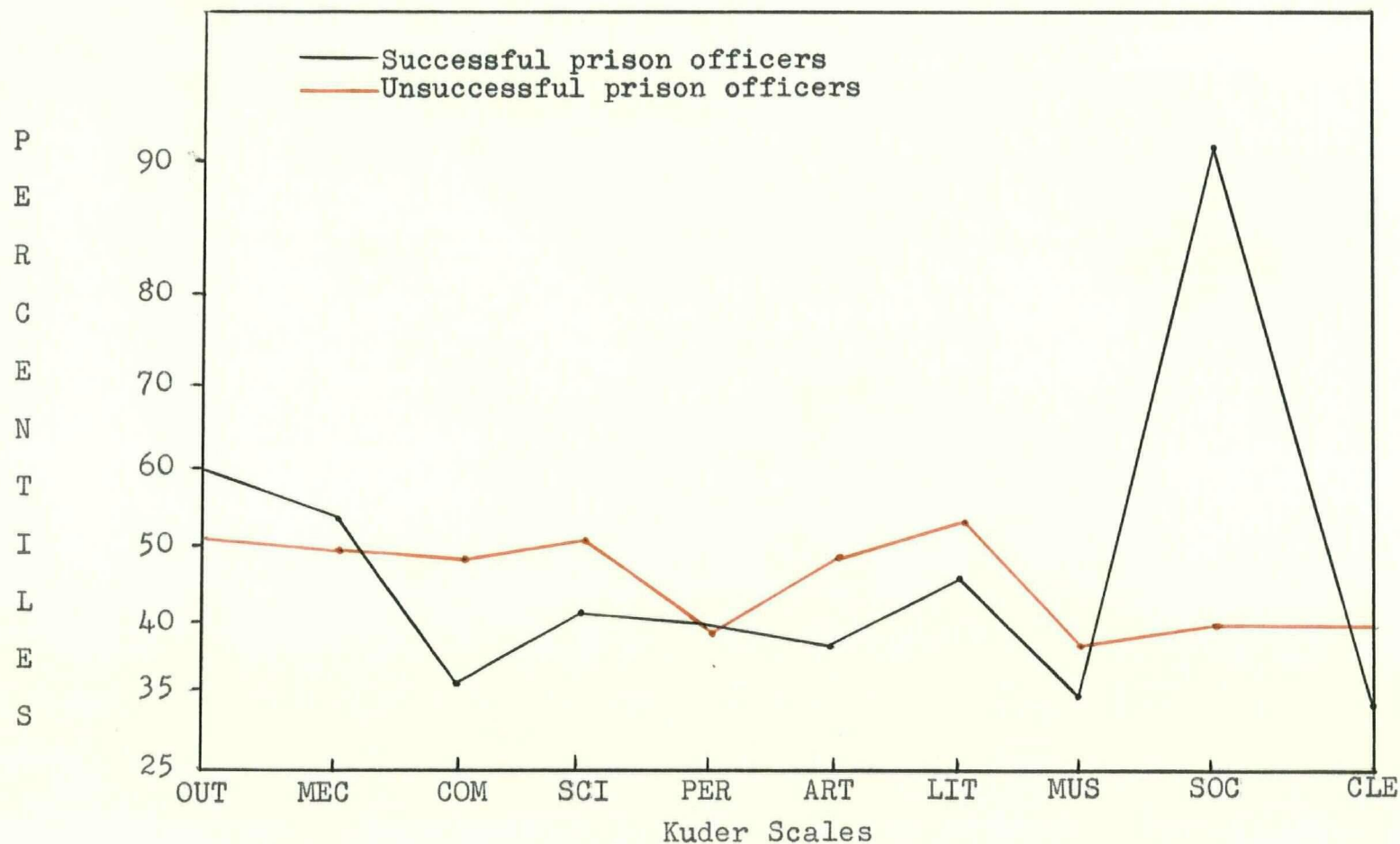


Figure III. Percentile ranks of mean scores obtained by groups of successful and unsuccessful prison officers on the Kuder Preference Record - Vocational.

in order was the Clerical scale with a coefficient of .24, followed by the Computational scale, .21. These biserial correlation coefficients were all significantly greater than zero. The other correlations were below .20 and were not significantly greater than zero and, thus, do not contribute anything to differentiating between "good" and "poor" prison officers.

The Reliability of Difference between Standard Deviations

Differences in variability between the successful and unsuccessful groups of prison officers is summarized in Table X. Only on the Social Service and Clerical scales were variabilities significantly different between the two groups. On the former scale the difference was significant at the .01 level with the greater variability occurring in the unsuccessful group. Likewise, there was more variability in the lower group on the Clerical scale. This difference was significant at the .05 level.

Additional Data

As a practical aid in selection, tentative norms were computed based on the scores of the total sample. Table XI shows the percentile ranks for each of the ten preference scales.

Summary

1. When an analysis was made of the total sample in relation to the published norms, only one scale appeared significantly different. This was the Social Service scale. In regard to the strength of Social Service interest the prison

TABLE X
THE RELIABILITY OF DIFFERENCE BETWEEN STANDARD DEVIATIONS
OF THE SUCCESSFUL AND UNSUCCESSFUL GROUPS
ON THE KUDER

Kuder Scales	Successful Group S.D.	Unsuccessful Group S.D.	Difference	Critical t-Ratio
Outdoor	14.55	13.94	.61	.22
Mechanical	9.25	11.26	-2.01	.49
Computational	7.35	6.99	.36	.26
Scientific	10.15	7.34	2.81	1.62
Persuasive	9.56	12.10	-2.54	1.19
Artistic	6.67	7.25	-.58	.42
Literary	6.38	7.41	1.03	.76
Musical	4.72	4.94	.22	.23
Social Service	8.54	17.61	-9.07	3.35**
Clerical	6.25	10.45	-4.20	2.47*

* Significant at the .05 level.

** Significant at the .01 level.

TABLE XI

NORMS FOR 100 PRISON OFFICERS SHOWING PERCENTILE DISTRIBUTIONS
ON THE KUDER PREFERENCE RECORD - VOCATIONAL

Score	Out	Mec	Com	Sci	Per	Art	Lit	Mus	Soc	Cle
80	100								100	
78	98								99	
76	98								99	
74	97								98	
72	96								97	
70	95								96	
68	94	100			100				94	
66	92	99			99				92	100
64	89	99		100	99				85	99
62	86	95		99	99				80	99
60	83	90		99	98				72	98
58	81	87		98	97				66	97
56	80	83		97	96				60	95
54	70	77	100	96	94				55	94
52	65	71	99	94	92	100			48	90
50	60	65	99	90	89	99			44	85
48	55	57	99	86	85	99			37	80
46	50	50	98	78	80	99			31	72
44	46	42	97	70	75	99	100		26	65
42	42	35	96	62	68	98	99		22	57
40	35	29	93	53	60	98	99		18	47
38	23	17	91	43	53	96	99		17	37
36	23	15	85	34	45	95	98		15	30
34	23	7	78	25	39	92	97		14	23
32	13	7	71	18	31	88	96	100	13	16
30	13	3	63	13	25	83	94	99	12	12
28	7	3	55	6	20	77	90	99	8	8
26	6	3	45	5	17	70	84	99	8	5
24	5	1	35	3	11	61	77	98	6	4
22	4	1	25	2	6	52	68	96	6	1
20	3		20	1	5	44	58	92	6	1
18	2		14		2	35	48	88	4	
16	2		12		2	27	37	80	4	
14	1		5		1	20	26	70	4	
12	1		4		1	15	18	56	1	
10	1		2		5	10	14	45	1	
8	1		2			3	7	33	.5	
6			0			3	7	23	.5	
4						3	7	15	.5	
2										
0										

officer, in this sample, may be classified along with individuals employed as social workers, welfare workers, and teachers (secondary school). It is interesting to note also that scale four, viz., Persuasive, yields one of the lower ranges of scores and this might suggest that the group as a whole has little need for "aggressive outlets" on the job. This finding would seem to support the view that the present-day officer is different from the guard of the "old school" and to reflect modern penological thinking in that, although the modern prison officer must be firm and often stern, the emphasis lies in helping or rehabilitating the inmate rather than using aggressive methods to keep the inmate in "line".

All remaining scales are near the medians of norms used and thus have little significance from the viewpoint of selection.

2. A critical score of 45 was determined on the Social Service scale. Had such a cut-off standard been used at time of hiring, 59 per cent of the unsuccessful group would have been rejected, whereas only 4 per cent of the successful group would have been eliminated.

3. Three scales on the Kuder discriminated between the two groups. The successful officers had significantly higher scores than the unsuccessful officers on the Social Service scale and significantly lower scores on the Computational scale and the Clerical scale. Biserial correlation coefficients on all three scales were significantly greater than zero. However, the

correlations of the other scales were of a low order ranging from .05 to .18 and were not significantly greater than zero.

From the analysis of the two groups in this study it appears that "good" prison officers have a markedly greater interest in social service activities, such as working with and helping people, and are less interested in activities that require precision and accuracy of detail or working with numbers and figures.

4. Differences in variability were found between the two groups on the Social Service and Clerical scales. An analysis of individual Social Service scores revealed that in the unsuccessful group two men received extremely high scores, while three men scored extremely low. These five extreme scores represent over 20 per cent of the poor group sample and undoubtedly account for much of the greater variability in the unsuccessful group.

III. THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

Profile of Total Sample based on Mean T-Scores of the MMPI

Table XII shows the mean T-scores and standard deviations of the total sample. The mean T-scores for the sample of prison officers all fall within the "normal" range. The highest mean T-score was 57.06 on the Psychopathic Deviate (Pd) scale. In descending order the Hypomanic scale was next with a mean T-score of 55.20, followed by Hypochondriasis (Hs) and Depression (D) scales both with a T-score of 52. The mean scores for the

TABLE XII

MEAN T-SCORES AND STANDARD DEVIATIONS OF TOTAL SAMPLE
ON THE MMPI ORIGINAL AND ADDITIONAL SCALES

MMPI Scales	Mean T-Scores	S.D.
L #	4.64	2.30
K #	15.15	4.30
F #	3.76	2.30
Hs	48.92	6.97
D	52.30	9.77
Hy	52.70	7.14
Pd	57.06	9.97
Mf	52.62	8.44
Pa	49.06	8.81
Pt	49.02	8.63
Sc	50.01	9.42
Ho	46.38	9.75
Ma	55.20	9.07
Si	46.85	7.45
Dom.#	17.33	2.76
RE #	21.32	3.58
Soc.St.#	21.74	3.21

Raw scores

other scales fall below a T-score of 50.

The three validity scales, L, K, and F, all centred around a mean T-score of 50.

Comparison of Mean Scores Between the Two Groups

Table XIII gives the means, differences between means and biserial correlation coefficients between the successful and unsuccessful prison officers. Of the nine original MMPI clinical scales the means on only three were significantly different between the two groups. A difference significant at the .01 level was observed for the means on the Depression scale, and a difference significant at the .05 level of confidence was found for the means on the Psychasthenia scale and Validity scales.

Other MMPI Scales

Several other scales which have been developed on normal groups were also included in the study. Cook (11) developed a Hostility scale; Gough developed a Dominance scale (23), a Responsibility scale (24), and a Social Status scale (22), and Drake developed the Social Introversion scale (16). All five scales were developed from existing MMPI items. Home-made scoring stencils were made and the scales treated statistically in the same manner as the regular MMPI scales. The results are presented in Table XIII.

Three of the five additional scales yielded significant differences between the two groups. A difference significant at

TABLE XIII

MEAN T-SCORES, DIFFERENCE BETWEEN MEANS, AND BISERIAL CORRELATIONS
FOR TWO GROUPS OF PRISON OFFICERS ON THE MMPI SCALES AND ADDITIONAL MMPI SCALES

Original MMPI Scales	Successful Group Mean	Unsuccessful Group Mean	Difference	T-Ratio	R Bis
L #	4.00	4.33	.33	1.56	.05
F #	2.30	3.80	1.50	2.39'	.27**
K #	15.44	14.93	.51	.41	.05
Hs + .5K	47.51	50.30	2.80	1.57	.16
D	48.40	55.59	7.19	3.09*	.30**
Hy	51.62	52.92	1.30	.58	.07
Pd + .4K	55.52	56.67	1.15	.43	.05
Mf	53.11	51.52	1.61	.66	.08
Pa	47.07	50.44	3.47	1.57	.16
Pt + 1K	47.27	52.90	5.63	2.47''	.27**
Sc + 1K	48.00	51.77	3.77	1.56	.16
Ma + .2K	54.77	58.15	3.38	1.32	.15
Additional MMPI Scales					
Si	45.81	51.26	5.45	2.52''	.30**
Ho	45.64	51.85	6.21	2.28'	.26**
Dominance#	18.81	16.25	2.56	3.77*	.38**
Respon- sibility#	21.11	20.00	1.11	1.24	.14
Social Status#	22.18	21.48	.64	.77	.08

Raw scores

' Significant at the .05 level

'' Significant at the .02 level

* Significant at the .01 level

** Significantly greater than
zero

the .01 level of confidence was found for the Dominance scale (Do), a difference at the .02 level for the Social Introversion scale (Si), and a difference significant at the .05 level of confidence for the Hostility scale (Ho). The unsuccessful group received significantly higher mean scores on the Si and Ho scales whereas the successful group received a significantly higher mean score on the Do scale. Nearly identical means were found for both groups on the Responsibility and Social Status scales.

Biserial Correlations

Table XIII also shows the differentiating capabilities of each scale expressed in terms of biserial correlation coefficients. The Dominance scale had the highest correlation coefficient with .38, next in descending order is the Depression scale with a coefficient of .30, then the Social Introversion scale with .30, the Validity scale and Psychasthenia scale both yield a biserial correlation of .27 and finally the Hostility scale with .26. These biserial correlations were all significantly greater than zero. The other correlations were below .20 and were not significant.

Reliability of Standard Deviations

Differences in variability on the MMPI between successful and unsuccessful guards were investigated. These data are presented in Table VIV. With the exception of the Validity scale (F) there were no significant differences in variability

TABLE XIV

RELIABILITY OF DIFFERENCES BETWEEN STANDARD DEVIATIONS
FOR SUCCESSFUL AND UNSUCCESSFUL PRISON OFFICERS ON THE MMPI

MMPI Scales	Successful Group S.D.	Unsuccessful Group S.D.	Difference	T
L #	2.47	2.37	.10	.21
F #	1.48	2.75	1.27	2.95*
K #	4.32	4.64	.32	.36
Hs	6.41	6.87	.46	.35
D	7.45	9.19	1.74	1.06
Hy	7.77	8.32	.55	.34
Pd	10.22	9.09	1.12	.59
Mf	9.93	7.48	2.45	1.42
Pa	8.08	7.89	1.20	.12
Pt	7.24	9.08	1.84	1.15
Sc	8.44	8.99	.55	.33
Ma	7.43	10.84	3.41	1.93
Si	6.12	9.12	3.00	1.97
Ho	10.14	9.46	.68	.35
Dom.#	2.59	2.28	.31	.65
Respons.#	3.37	3.59	.22	.32
Soc. St.#	2.76	3.24	.48	1.02

Raw scores

* Significant at the .01 level

between the two groups.

Scattergrams

Scattergrams were carefully examined to establish cutting off scores where possible. Because of the degree of overlap between the scores of the successful and unsuccessful groups on most of the scales, cutting off scores could not be applied for all MMPI scales.

Of the five scales which differentiated between successful and unsuccessful guards, three scales permitted reasonable cutting off scores. These are the Depression, Psychasthenia and Hostility scales. The cutting off scores were 55, 57, and 52 respectively. Cutting off scores were arrived at according to the conventional practice of determining the score point which results in the smallest loss of successful and the maximum loss of unsuccessful personnel.

Summary and Discussion of Results

1. The mean profile of the total sample suggests that as a group, prison officers are less inhibited, more capable of acting out impulses and reducing tensions, and are more active and enthusiastic than the general population norm group for the MMPI. The values of the validity scales suggest that as a group they did not attempt to place themselves in the most favourable light when answering questions and were neither over-defensive nor over-critical in their test-taking attitude.

2. The successful group differed from the unsuccessful

group on six scales: the Depression scale, the Psychasthenia scale, the Social Introversion scale, the Hostility scale, the Dominance scale and the Validity scale. It is interesting to note that the last three MMPI additional scales which were derived from a normal population seem to have more validity than the original MMPI scales. However, this is not too surprising since the total prison officer sample is also a normal group and thus differences within this group compare more favourably with the population on which the additional scales were based rather than with the clinical populations from which the regular scales were derived. The unsuccessful officers as a group scored higher on all the MMPI scales, with the exception of the Male-Female interest scale where the means almost coincide.

3. Qualitatively speaking, successful officers appear to have more confidence; more ability to concentrate; less tendency to worry and become introverted; greater interests; they appear to be more optimistic about the future and generally more extroverted compared to unsuccessful officers. Moreover, successful officers appear to like and trust and have more confidence in their fellow men, or, in other words, are more tolerant and less critical than unsuccessful officers. The successful officer tends to be more self-assured and assertive in his dealing with people, that is, he tends to be "stronger" in face-to-face personal situations than the unsuccessful officer, who by contrast tends to lean in the other direction toward a state of submissiveness and being "over-kindly" so to speak. This

characteristic of assertiveness and self assurance, if not excessive, is essential to a prison setting, since a satisfactory prison officer must have the ability to assert himself and administer sound discipline when the occasion calls for it.

4. With the exception of the Validity scale the two groups were not significantly different in variability. A difference significant at the .01 level was found on the Validity scale. In this instance an analysis of the two groups showed a range of scores from 1 to 13 for the unsuccessful group and from 0 to 6 for the successful group.

5. When analyzing individual profiles it was apparent that more of the "poor" officers seemed to get "high" MMPI scores than did the "good" officers. Therefore a second method of comparison was decided upon. A tally was made of the percentage of persons in each group obtaining a given number of scores over the 65th T-score. These percentages are shown in Table XV. A test of the significance of difference of percentages revealed that significantly more successful officers had only one or less high scores. Only 37 per cent of the successful officers had scores over the 65th T-score on any scale, and only 11 per cent had more than one high score. On the other hand, 66.8 per cent of the unsuccessful officers had scores over the 65th T-score and 36.8 per cent had more than one high score.

6. Critical scores were derived for three scales:

Depression scale. A T-score of 55 was found to be the best cutting score on the depression scale. Had this score been

TABLE XV
FREQUENCIES (IN PER CENTS) OF SCORES
OVER THE 65TH T-SCORE ON THE MMPI

MMPI Scales	Successful Officers	Unsuccessful Officers	Difference
Hs	0	3.7	3.7
D	0	26.0	26.0
Hy	11.1	14.8	3.7
Pd	14.8	14.8	0
Mf	14.8	7.4	7.4
Pa	0	7.4	7.4
Pt	0	7.4	7.4
Sc	3.7	11.1	7.4
Ma	14.8	33.3	18.5
Si	0	14.8	14.8
Ho	0	7.4	7.4
Number of Scales	Summary of Percentages of Groups Scoring above Critical Scores on a Different Number of Scales		
0	63.0	33.2	29.8*
1	26.0	30.0	4.0
2	3.7	18.4	14.7
3 and over	7.3	18.4	11.1

* Significant at the .01 level

used at time of hiring, 52 per cent of the unsuccessful officers would have been rejected, whereas only 22 per cent of the successful officers would have been eliminated.

An analysis of the ten best officers and ten poorest officers in the sample showed that this cutting score would have eliminated 60 per cent of the poorest men but only 10 per cent of the best.

Psychasthenia scale. A T-score of 57 proved to be the best cutting score on the Psychasthenia scale. A cutting score of 57 would have eliminated 30 per cent of the "poor" officers while rejecting only 4 per cent of the "good" officers.

An analysis of the ten best officers and ten poorest officers in the total sample revealed that such a critical score would have eliminated 30 per cent of the poorest men with no loss at all in number of the best men.

Hostility scale. A T-score of 52 was the best cutting score on the Hostility scale. Such a cutoff point would have eliminated 52 per cent of the unsuccessful officers and 33 per cent of the successful officers. However, an analysis of the ten best and ten poorest officers in the total sample revealed that a critical score of 52 would have eliminated 50 per cent of the poorest officers but only 10 per cent of the best officers.

IV. MANSON EVALUATION

Mean Score Profiles

Table XVI shows the means and standard deviations of

TABLE XVI
MEANS AND STANDARD DEVIATIONS OF 100 PRISON OFFICERS
ON THE MANSON EVALUATION

Manson Scales	Mean	S.D.
AN (Anxiety)	2.17	2.51
DF (Depressive Fluctuation)	1.91	1.68
ES (Emotional Sensitivity)	1.74	1.75
RE (Resentfulness)	2.51	2.01
IN (Incompleteness)	4.60	2.02
AL (Aloneness)	1.49	1.16
IR (Interpersonal Relations)	1.24	1.18
ME Total	12.96	7.31

100 prison officers on the Manson Evaluation test which purports to measure alcoholic traits. The total mean score on the test was 12.96 with a standard deviation of 7.31. The highest mean score on the subtests was the Incompleteness scale with 4.60 and, next highest, the Resentfulness scale with 2.51, and Anxiety scale with 2.17. The scores ranged from a total score of 2 to 39. Most variability appeared on the Anxiety scale, with a standard deviation of 2.51, the Resentfulness and Incompleteness scales follow with S.D. values of 2.01 and 2.02, respectively.

Mean Differences Between the Two Groups

Table XVII gives the means, difference between means, t-values, and biserial correlations for successful and unsuccessful prison officers. Differences were found at the .05 level as follows: Total score resulted in a t-score of 2.35, the Depressive scale, 2.03, and the Incompleteness scale, 2.36. No significant differences were found at the .01 level.

The Manson total score and the Resentfulness scale both had a biserial correlation coefficient of .26, the Aloneness scale, .22, and the Depressive Fluctuations scale, .21.

Variability Between the Two Groups

Table XVIII shows the reliability of the differences between standard deviations of the two groups. The only significant result between the two groups was on the Depressive Fluctuations scale. The t-score of 2.06 is significant at the .05

TABLE XVII

MEANS, DIFFERENCE BETWEEN MEANS, AND BISERIAL CORRELATIONS
FOR SUCCESSFUL AND UNSUCCESSFUL PRISON OFFICERS ON THE MANSON EVALUATION

Manson Scales	Successful Officers Mean	Unsuccessful Officers Mean	Difference	T-Ratio	R Bis
AN (Anxiety)	1.85	3.30	1.44	.74	.13
DF (Depressive)	1.55	2.40	.85	2.03*	.21**
ES (Emotional Sensitivity)	1.44	2.18	.74	1.53	.17
RE (Resentfulness)	2.11	2.77	.66	1.27	.16
IN (Incompleteness)	4.07	5.40	1.33	2.36*	.26**
AL (Aloneness)	1.48	2.11	.63	1.83	.22**
IR (Interpersonal Relations)	1.00	1.70	.70	2.00	.19
Total Score	11.70	16.44	4.74	2.35*	.26**

* Significant at the .05 level of confidence.

** Significantly greater than zero.

TABLE XVIII
RELIABILITY OF DIFFERENCES BETWEEN STANDARD DEVIATIONS
FOR THE TWO GROUPS ON THE MANSON EVALUATION

Manson Scales	Successful Group S.D.	Unsuccessful Group S.D.	Difference	T
AN	1.98	2.47	.49	1.12
DF	1.17	1.79	.62	2.06*
ES	1.65	1.85	.20	.58
RE	1.87	1.88	.01	.03
IN	1.71	2.30	.59	1.47
AL	1.07	1.40	.33	1.35
IR	1.05	1.27	.22	.97
Total	5.84	8.48	2.64	1.86

* Significant at the .05 level

level.

Critical Scores

Critical scores were obtained for the Total score and several of the subtests as follows: A critical score of 17 on the Total test score, 5 on the Anxiety scale, 3 on the Depressive Fluctuation scale and 7 on the Incompleteness scale.

Intercorrelation Between Subtests

The Depressive Fluctuation scale and the Incompleteness scale correlated .46.

Analysis of Manson Evaluation and MMPI Scales

The scales on the Manson which differentiated between the two groups were correlated with the scales on the MMPI which differentiated between the two groups. A tabulation of the test correlations from which the following analysis is made is found in Table XIX.

Manson Total score and MMPI scales. The MMPI Hostility scale correlated the highest with the Manson total score, yielding a coefficient of .47; the Social Introversion scale was next with .40. Both of these correlations are significant beyond the .01 level of confidence. The Psychasthenia scale correlated .21 which is significant at the .05 level. The Depression and Dominance scales are of low order, .10 and -.08, respectively, neither of which were significant.

Manson "Depressive Fluctuation" and MMPI scales. Of all the intercorrelations, the Social Introversion and Depressive

TABLE XIX
CORRELATIONS BETWEEN SEVERAL MANSON EVALUATION
AND MMPI SCALES

MMPI Scales	Manson Evaluation Scales		
	Total Score	Depressive Fluctuation	Incomplete- ness
Psychasthenia	.21**	.26**	.29**
Depression	.10	.27**	-.21**
Social Introv.	.40**	.52**	.10
Hostility	.47**	.21*	.22*
Dominance	-.08	-.17	-.38**

* Significant at the .05 level

** Significant at the .01 level

Fluctuation scale correlated the highest with .52, significant well beyond the .01 level. The MMPI Depression and Psychasthenia scales correlated .27 and .26, respectively, both significant at the .01 level. The Dominance scale had a negative correlation of -.17 which was not significant.

Manson Incompleteness scale and MMPI scales. The highest correlation was on the Dominance scale with a negative correlation of -.38, significant at the .01 level of confidence. The Psychasthenia scale had the highest positive correlation, .29, which was significant at the .01 level. The Hostility scale correlated .22 and the Depression scale, -.21, both significant at the .05 level.

Analysis of Individual Scores

An analysis of individual scores was undertaken for each group. A comparison was made by tallying the percentage of each group obtaining a given number of scores over critical points. The critical scores that are regarded as high are the ones given in the test manual (46). The upper portion of Table XX shows the percentages of each group who received scores above the critical points on each scale. The bottom part of Table XX shows that 67 per cent of the successful officers did not exceed the critical scores on any of the scales. This compares with 44 per cent of the unsuccessful officers who did not exceed any of the critical scores. This difference was significant at the .05 level of confidence. The other comparisons in terms of one,

TABLE XX
PERCENTAGES OF GROUP SCORING ABOVE CRITICAL SCORES
ON THE MANSON EVALUATION

Manson Scales	Successful Officers	Unsuccessful Officers	Difference
AN	7.4	22.0	14.6
DF	0	7.4	7.4
ES	3.7	13.3	9.6
RE	22.0	30.0	8.0
IN	7.4	37.0	29.6**
IR	3.7	7.4	3.7
Total Score	14.8	30.0	15.2
Number of Scales	Summary of Percentages of Groups Scoring above Critical Scores on a Different Number of Scales		
0	67.0	44.0	23.0*
1	14.0	22.0	8.0
2	7.0	12.0	5.0
3 and over	12.0	22.0	10.0

* Significant at the .05 level

** Significant at the .01 level

two and three or more scales are not significant.

Conclusion and Discussion

1. The total sample in this study obtained a total mean score of 12.96 on the Manson which compared with the mean score of a group of social drinkers (45) used in the validation study of the Manson test. The mean score was significantly below the critical score of 21 for alcoholics.

2. The poorer group of officers received significantly higher scores on the Depressive Fluctuation scale, the Incompleteness scale and the Total Test score. From this, the clinical picture as measured by the Manson would seem to indicate that the "poor" group tend to be more easily depressed, have more frequent mood swings and tend to be more restless and unsteady than the good group of officers.

3. Scattergrams were investigated to determine possible critical scores on some of the scales. A cutting score of 17 was found on the Total score. Such a score would have eliminated 50 per cent of the "poor" group but only 18 per cent of the "good" group.

The most effective cutting score on the Anxiety scale was 5. Such a cutting score eliminated 30 per cent of the "poor" men while retaining 89 per cent of the "good" officers.

On the Depressive Fluctuation scale a critical score of 3 proved the most effective. If this score had been utilized at the time of hiring, 44 per cent of the unsuccessful officers in

this sample would have been rejected before coming on the job, whereas only 15 per cent of the successful officers would have been rejected.

Investigation of the Incompleteness scale favoured 7 as the cutting score which eliminates 37 per cent of unsuccessful officers but only 7 per cent of the successful officers. An analysis of the ten best officers and the ten poorest officers on the same scale revealed that a critical score of 7 would reject 66 per cent of the very poor men while retaining 75 per cent of the best men.

4. An analysis in terms of the critical scores (taken from the test Manual) for each of the Manson scales revealed that more of the unsuccessful officers exceeded such critical scores. However, a similar analysis in terms of Total scores on the Manson revealed that there were no statistically significant differences between the two groups. According to the test author (46) a Total score of 21 is the critical score which differentiates alcoholics from non-alcoholics. On the total test, 30 per cent of the unsuccessful group obtained critical scores of 21 or over while 14 per cent of the successful group obtained similar scores.

5. Correlations between the MMPI and the Manson Evaluation reveal that, for the most part, the scales under question are fairly independent measures. The MMPI Social Introversion and Hostility scales correlated the highest with the Manson scales. The MMPI Social Introversion scale and the Manson Depressive Fluctuation scale yielded a correlation of .52. The

Social Introversion scale correlated .40 with the Manson Total score. The next highest correlation. .47 was found between the MMPI Hostility scale and the Manson Total score.

It appears that the Manson Depressive scale and the MMPI Depression scale are measuring two different characteristics. The Social Introversion scale appears to have more in common with the Manson Depressive scale than does the MMPI Depression scale.

One concludes that the important scales form the two tests studied, although not yielding completely independent measures, supplement one another in important respects.

CHAPTER VI

GENERAL CONCLUSIONS AND IMPLICATIONS FOR FURTHER RESEARCH

General Conclusions

1. Of the 35 test variables used in this study, 14 variables significantly discriminated between successful and unsuccessful prison officers.

The following 5 variables were significant beyond the .01 level of confidence: Part I and Total score on the Personnel Classification Test; the Kuder Social Service scale; the MMPI Dominance scale; and the MMPI Depression scale. Except for the Depression scale the successful group in each instance scored higher means than the unsuccessful group.

Two scales significantly differentiated the two groups at the .02 level of confidence. These were the MMPI Psychasthenia scale and the MMPI Social Introversion scale. The unsuccessful group received higher means in both instances.

The following 7 scales significantly differentiated the two groups at the .05 level of confidence: the Kuder Computational scale; the Kuder Clerical scale; the MMPI Validity (F) scale; the MMPI Hostility scale; the Manson Depressive scale; the Manson Incompleteness scale; and the Manson Evaluation Total score. Except for the MMPI Validity scale the unsuccessful group received higher means on each of these variables.

2. In this study the best predictor in differentiating successful and unsuccessful prison officers proved to be the Social Service scale of the Kuder Preference Record. The Social Service scale significantly discriminated between the two groups well beyond the .01 level with a t-ratio of 4.71 and a biserial correlation coefficient of .49. The next best predictor was the Verbal scale of the Personnel Classification Test which also significantly differentiated the two groups at the .01 level of confidence with a t-ratio of 4.60 and with a biserial correlation of .46. These two scales alone make a significant contribution to increased proficiency in practical selection procedures. (See (5) below.)

3. The correlation coefficients were generally of a low order, the largest was .49, but this is not unusual in the measurement of psychological characteristics. In regard to low correlation coefficients, Cronbach states (13, p. 257) that the only fair standard for an acceptable validity coefficient is the question: "Does the test permit us to make a better judgment than we could make without it - sufficiently better to justify its cost?" The evaluation of validity coefficients for selection purposes nowadays is usually made in terms of selection cost. According to Cronbach (13, p. 257), "a test which increases the proportion of good employees is a test worth using; but the validity of the test must be balanced against attrition, the number of potentially good employees discarded in screening."

However, percentage comparisons are often more valuable

than correlations. In practice one is usually more interested in the problem of how well one can differentiate between the poor and the best workers and not so much in the excellence of differentiation at the extremes of test scores. Percentage comparisons are thus especially useful in showing how well prediction can be made at some particular part of the total range of test scores.

4. In this study scattergrams were investigated to determine possible critical scores on those test variables which differentiated the two groups. The ideal critical point is that score point which differentiates the greatest number of probable failures from the greatest number of probable successes. Because of the degree of overlap between the samples of successful and unsuccessful groups, it did not seem justified to determine critical scores for a number of the scales. However, in spite of the overlap, six scales yielded reasonably effective critical scores and their practical significance is discussed below. Since these critical scores are based on the results from small samples of successful and unsuccessful officers, they must be considered tentative until corroborated by further research.

The results showed that: (1) If a critical score of 13 on Part I of the Personnel Classification Test had been used at the time of hiring, 58 per cent of the unsuccessful group would have been rejected whereas only 4 per cent of the successful group would have been eliminated. (2) A critical score of 45 on the Kuder Social Service scale would have rejected 59 per

cent of the unsuccessful group but only 4 per cent of the successful officers. (3) A cutting score of 57 on the MMPI Psychasthenia scale would have eliminated 30 per cent of the unsuccessful officers and only 4 per cent of the successful group. (4) A critical score of 56 was determined on the MMPI Depression scale. Such a cutting score would have eliminated 51 per cent of the unsuccessful group used in the study and 22 per cent of the successful group. (5) A cutting score of 52 on the MMPI Hostility scale would have eliminated 50 per cent of the unsuccessful group and 20 per cent of the successful group. (6) Finally, a critical score of 17 on the Manson Total scale would have eliminated 50 per cent of the unsuccessful officers but only 18 per cent of the successful officers.

5. An analysis was made of the combined effect of the six critical scores on the two groups. Table XXI presents the number of successful and unsuccessful officers that are eliminated when various scores are applied to the two groups. The results indicate that if the two best predictors (Social Service scale and Part I PCT) had been used at the time of hiring, 93 per cent of the unsuccessful group would have been rejected, whereas only 8 per cent of the successful group would have been eliminated. If all six critical scores had been used, 100 per cent of the unsuccessful group would have been eliminated, while 63 per cent of the successful group would also have been eliminated. It can be seen from the table that the greater the number of critical scores used for screening purposes, the less effective are the

TABLE XXI
EFFECT OF SIX CUTTING SCORES IN ELIMINATING
SUCCESSFUL AND UNSUCCESSFUL PRISON OFFICERS

Various Critical Scores Employed* (in order of merit)	Successful Group Percentage Eliminated (cumulative)	Unsuccessful Group Percentage Eliminated (cumulative)
1	4	58
1,2	8	93
1,2,3	12	93
1,2,3,4	30	93
1,2,3,4,5	52	100
1,2,3,4,5,6	63	100

* 1. Part I (verbal) PCT
2. Soc. Service scale
3. Psychasthenia scale

4. Depression scale
5. Hostility scale
6. Manson Total scale

results. However, in actual practice the three measured areas (intelligence, interest and personality) would be considered separately and critical scores would be selective and pertain to that one area only. That is, an applicant would have to pass the "hurdles" or critical scores in each of the three predictive areas before being hired on.

6. In summary, the results seem to indicate that for practical purposes personality, interest and intelligence tests are probably more useful in screening out candidates who should be rejected rather than in indicating which ones can be safely accepted. Because of the significant t-ratios obtained, and the significant relationships between tests and criterion ratings, it is felt that the results warrant further investigation.

Suggestions and Implications for Further Research

1. General interpretation of the results from this study must be guarded and must be considered tentative since more refined research is necessary before further generalizations may be drawn. The next step must be a cross-validation study utilizing applicants.

The purposes of such a cross-validation study is to protect one from putting too much confidence in relationships which may just "happen" to hold true for the present group, but which may not hold true in the long run. A cross-validation study would prove whether the results derived from this study would be truly effective for screening purposes, when applied to another independent sample of officers. However, the study must

consist of a representative sample of applicants rather than presently employed officers.

In regard to this, the use of the "present employee" method of validation raises the question of whether the cutting scores suggested above can be used to predict job proficiency when applied to applicants. When using presently employed officers, the sample should represent similar age, vocational interests and attitudes and other characteristics that the battery will ultimately be used on, i.e., prospective job applicants. However, this principle is violated when using on-the-job officers, since these men markedly differ from applicants in their ability to do the job and in the fact that they are not similar in training, age, interest or attitude, and thus results from such a validation might differ greatly from the results obtained from a group of applicants. In this regard, Ghiselli and Brown (22, p. 173) state that correlation coefficients will be lower for prognostic purposes when established workers are used rather than applicants, and the distribution of the scores on the tests will not be the same. However, if the test battery stands up under this kind of validation it will tend to have more differentiating power (than suggested by the results from the group used in this study) when used with applicants. This would seem to be so since psychological tests will reflect a greater differentiation in a large representative group rather than in a partially screened smaller group; and when measuring applicants the variations in ability, interest, and so on, will be greater than employed

workers because of the reasons mentioned above. Thus, the critical scores determined from the present group of employed officers will be somewhat different from those obtained from applicants. Consequently, any critical scores established in this study must be considered wholly tentative and should be corrected from the results of a cross-validation study utilizing a representative sample of applicants.

2. A comprehensive job analysis of the position of prison officers using the critical incidents technique should be undertaken to determine in a scientific way what abilities and habits contribute to or limit success.

3. The MMPI should be revalidated on applicants to develop a specific occupational key utilizing item analysis, in much the same way as the keys were developed for the existing categories on the test.

Also a further study of the MMPI should be undertaken using mean profile analysis interpretations of high and low patterns, rather than placing the emphasis upon individual mean scores.

4. Reliabilities of the test measures must also be obtained by retest methods or by using equivalent test forms.

5. An important research project for the development of adequate selection procedures is the need of a thorough investigation of the validity and reliability of various criteria of success. Such a study would strive to attain objective criteria measures so far as possible and also develop more reliable merit

rating procedures.

6. In conclusion one should point out that although a test battery is an important supplement in selection procedures, other methods of selection are still necessary adjuncts, for example the interview and the probationary training period.

BIBLIOGRAPHY

1. Allport, G. W. A test for ascendance-submission. J. abnorm. soc. Psychol., 1928, 23, 118-136.
2. Andrews, A. G. A year's experience of selection tests for engineering operatives. Occup. Psychol., 1937, 21, 139-172.
3. Armstrong, T. The Life Insurance Bureau: what it is, how it operates. Sales Mgmt., 1944, LIII, 23, 83-89.
4. Barnes, H. E., & Teeters, N. K. New horizons in criminology. New York: Prentice-Hall Inc., 1952.
5. Bellows, Roger M. Psychology of personnel in business and industry. New York: Prentice-Hall, Inc., 1949, pp. 23-25.
6. Bernreuter, R. G. Theory and construction of the personality inventory. J. soc. Psychol., 1933, 4, 387-405.
7. Blewett, D. B. A comparison of certain measures of interest in armed service grade selection. Unpublished M. A. thesis, The University of British Columbia, 1950.
8. Bittner, Reign. Developing an employee merit rating procedure. Personnel, 1949, 1, 275-279.
9. Buros, O. K. (ed.) The third mental measurements yearbook. New Brunswick: Rutgers University Press, 1949.
10. Cattell, J. McK. Mental tests and measurements. Mind, 1890, XV, 373-380.
11. Cook, W. W., & Medley, D. M. Proposed hostility and pharisaic-virtue scales for the MMPI. J. appl. Psychol., 1954, 38, 414-418.
12. Cowdery, K. M. Measurement of professional attitudes, difference between lawyers, physicians and engineers. J. pers. Res., 1926-1927, 5, 131-141.
13. Cronbach, Lee J. Essentials of psychological testing. New York: Harper and Bros., 1949.
14. Davies, J. G. W. What is occupational success? Occup. Psychol., 1950, 24, 233-241.

15. Dorcus, Roy M., & Jones, M. H. Handbook of employee selection. New York: McGraw-Hill Book Co., 1950.
16. Drake, L. E. A social I. E. scale for the MMPI. J. appl. Psychol., 1946, 30, 51-54.
17. Driver, R. S. The validity and reliability of ratings. Personnel, 1941, 17, 185-191.
18. Flanagan, John C. Factor analysis in the study of personality. Stanford: Stanford University Press, 1935.
19. Fryer, D. Measurement of interests. New York: Holt, 1931.
20. Fryer, D. H., & Henry, E. R. Handbook of applied psychology, Vol. I. New York: Rinehart and Company, Inc., 1950.
21. Garrett, H. E. Statistics in psychology and education. New York: Longmans, Green and Co., 1947.
22. Ghiselli, Edwin E., & Brown, Clarence W. Personnel and industrial psychology. New York: McGraw-Hill Book Co., Inc., 1948.
23. Gough, H. G. A new dimension of status: 1. Development of a personality scale. Amer. soc. Rev., 1948, 13, 401-409.
24. Gough, H. G., McClosky, H., & Meehl, P. E. A scale for measuring dominance. J. abnorm. soc. Psychol., 1951, 46, 360-366.
25. Gough, H. G., McClosky, H., and Meehl, P. E. A scale for measuring social responsibility. J. abnorm. soc. Psychol., 1952, 47, 73-80.
26. Gray, J. S. Psychology in human affairs. New York: McGraw-Hill Book Company, Inc., 1946.
27. Griffith, J. E., Kerr, W. A., Mayo, T. P., Jr., & Topal, J. R. Changes in subjective fatigue and readiness for work during the eight-hour shift. J. appl. Psychol., 1950, 34, 163-166.
28. Halsey, George D. Making and using industrial service ratings. New York: Harper and Brothers, 1944.
29. Harris, F. J. Can personality tests identify accident-prone employees? Personnel Psychol., 1950, 3, 455-459.
30. Hersey, R. Some neglected aspects of accident prevention. Personnel, 1950, 27, 231-236.

31. Hubbard, H. F. A study of objective employment tests in the New Jersey Civil Service. George Washington University doctoral dissertation, 1934.
32. Hull, C. L. Aptitude testing. New York: World Book Company, 1928.
33. Humm, Doncaster G., & Humm, Kathryn A. Compensation for subjects' response-bias in a measure of temperament. Amer. Psychologist, 1947.
34. Hunt, H. C. Why they couldn't hold their jobs. Pers. J., 1935, 14, 227-228.
35. Kerr, W. A. Accident proneness of factory departments. J. appl. Psychol., 1950, 34, 167-170.
36. Kornhauser, A. W. Results from a quantitative questionnaire on likes and dislikes used with a group of college freshmen. J. appl. Psychol., 1927, 21, 85-94.
37. Kuder, G. F. Kuder Preference Record, Form B.B., rev. ed., Chicago: Science Research Assoc., 1942.
38. Kuder, G. F. Profile sheet for the Kuder Preference Record, Vocational. Chicago: Science Research Assoc., 1950.
39. Kuder, G. F. Revised manual for the Kuder Preference Record, Vocational. Chicago: Science Research Assoc., 1953.
40. Lee, E. A., & Thorpe, L. P. Manual of directions, Occupational Interest Inventory, advanced theories.
41. Lundberg, D. E. Some aspects of the job of prison guard. Cornell University doctoral dissertation, 1946.
42. McNemar, Q. The revision of the Stanford Binet Scale. Boston: Houghton Mifflin, 1942.
43. Mahler, W. Twenty years of merit rating. New York: Psychological Assoc., 1946.
44. Maier, Norman R. F. Psychology in industry. New York: Houghton Mifflin Company, 1955, p. 243.
45. Manson, M. P. A psychometric differentiation of alcoholics from non-alcoholics. Quart. J. Stud. Alc., 1948, 9, 175-206.

46. Manson, M. P. Manual of directions and norms for the Manson Evaluation. Los Angeles: Western Psychol. Services, 1949.
47. Marrow, J. F. Human factors in production. Personnel, 1949, 25, 34-35.
48. Meehl, P. E., & Hathaway, S. R. The K factor as a suppressor variable in the MMPI. J. appl. Psychol., 1946, 30, 525-564.
49. Miner, J. B. An aid to the analysis of vocational interests. J. educ. Res., 1922, 5, 311-323.
50. Paterson, D. G. Vocational interest inventories in selection. Occupations, 1946, 25, 152.
51. Patten, E. F. An experiment in testing engine-lathe aptitude. J. appl. Psychol., 1923, 7, 16-29.
52. Paterson, D. G., Darley, John G., & Elliott, R. M. Men, women, and jobs. Minneapolis: The University of Minnesota Press, 1936.
53. Peters, C. P., & Van Voorhis, W. R. Statistical procedures and their mathematical bases. New York: McGraw-Hill Book Co., Inc., 1940.
54. Richardson, M. W. Forced-choice performance reports. New York: American Mgmt. Assoc., 1950.
55. Sawatsky, J. C. Psychological factors in industrial organization affecting employee stability. Can. J. Psychol., 1951, 5, 29-38.
56. Scott, W. D., Clothier, R. C., & Spriegel, W. R. Personnel Management. New York: McGraw-Hill Book Co., Inc., 1954.
57. Shaffer, L. F. The psychology of adjustment. New York: Houghton Mifflin Co., 1936.
58. Shartle, C. L. Occupational information. New York: Prentice-Hall, Inc., 1946.
59. Sisson, Donald E. Forced-choice, the new army rating. Personnel Psychol., 1948, 3, 365-381.
60. Stenquist, John L. Measurements of mechanical ability. New York: Columbia University, Teachers College Contributions to Education, No. 130, 1923.

61. Stockford, L. O., & Kunze, K. R. Psychology and the pay check. Personnel, 1950, 27, 2-15.
62. Strong, E. K., Jr., Vocational interests of men and women. Stanford University, Calif.: Stanford Press, 1943.
63. Stuit, D. B. (ed.) Personnel research and test development in the bureau of naval personnel. New Jersey: Princeton University Press, 1947.
64. Super, Donald E., & Dunlap, Jack W. Handbook of applied psychology (interest in work and play). New York: Rinehart and Company, Inc., 1950.
65. Symonds, P. M. Diagnosing personality and conduct. New York: The Century Co., 1931.
66. Taft, Donald R. Criminology, a cultural interpretation. New York: The Macmillan Company, 1950.
67. Thurstone, L. L., & Thurstone, T. G. Neurotic inventory. J. soc. Psychol., 1930, 1, 1-29.
68. Tiffin, J. Merit rating: its validity and techniques. New York: American Mgmt. Assoc., 1950.
69. U.S. Department of Labor. Dictionary of occupational titles. U.S. Government Printing Office, 1939.
70. Viteles, Morris S. Research in selection of motormen. J. pers. Res., 1925, 4, 100-115.
71. Wesman, A. G. Revised manual for the Personnel Classification Test. New York: Psychol. Assoc., 1951.
72. Woodworth, R. S. Dynamic psychology. New York: Columbia University, 1918.
73. World War II Research. Army Air Forces Aviation Psychology Program Research - Reports. Wash., D.C.: U.S. Government Printing Office, 1947.
74. Yerkes, R. M. Psychological examining in the United States Army. Memoirs of the National Academy of Sciences, XV, 1921.