PREDICTION OF SCHOLASTIC ACHIEVEMENT IN SECONDARY SCHOOLS:

The Relationship of Third Form Tasman and Otis Test Scores to Fifth Form School Certificate Marks.

A thesis presented as a partial requirement for the degree of

Master of Arts

In

Psychology

In the University of Canterbury,

Christchurch, New Zealand,

by

Robert M. Patchett

1967

CONTENTS

LIST OF TABLES

PREFACE

Testing of new entrants is well established practice in many New Zealand secondary schools. Group tests of intelligence are usually administered, and the results obtained may serve as a basis for assigning pupils to various courses of study at various levels of difficulty.

Little New Zealand research relating new entrant test results to later scholastic achievement has been reported. The validity of assignment decisions made using test results is therefore uncertain.

Mr G. Shouksmith, who acted as initial supervisor of this thesis while Senior Lecturer in Psychology at the University of Canterbury, constructed tests for the specific purpose of improving assignment decisions in New Zealand secondary schools. The major purpose of this study is to compare the predictive validity of Shouksmith's Tasman tests with that of the widely used Otis test using School Certificate examination marks as criteria.

I gratefully acknowledge guidance given me in preparing this thesis by Mr Shouksmith and later by Dr R.A.M. Gregson, Professor of Psychology, University of Canterbury. The co-operation of secondary schools which contributed Otis results is much appreciated. Finally I am indebted to the Department of Education for making School Certificate examination marks available and to the officers of the Department's Examinations Division for their friendly assistance during the extraction of relevant marks from national listings.

CHAPTER I

INTRODUCTION

The Otis Self-Administering Tests of Mental Ability have been used in New Zealand primary and secondary schools since the late 1930's, usually for "ability" grouping or "streaming", but on occasion for assigning pupils to particular courses of study, i.e. for classification purposes.

The author of the Tasman series of tests however, questioned the suitability of the Otis as a guide in the making of educational classification decisions, and the Tasman tests were devised to provide information about a pupil's differential abilities in addition to assessing general ability.

The major problem investigated by this study was whether there exist any significant differences between the Tasman and Otis tests as predictors of both general scholastic achievement and achievement in single subjects.

Tasman scores and Otis IQs obtained by third form new entrants to secondary schools were correlated with single subject and aggregated marks obtained by the same pupils in the nationwide School Certificate examination after three years secondary schooling. Statistical tests examining the significance of differences between predictor tests were then made.

CHAPTER II

EDUCATIONAL CLASSIFICATION: A REVIEW OF RESEARCH, THEORY AND PRACTICES

1. United Kingdom.

An examination of the published literature indicates that little significant work has been reported during recent years. This is particularly so of United Kingdom publications, probably because experience there with the 11+ examination has consistently shown that a reasonably accurate and economical classification of secondary school pupils is possible by using tests of Verbal Intelligence, English and Arithmetic, with the Intelligence test the best single predictor (Yates and Pidgeon 1957, Vernon 1958). The decision to be made in the United Kingdom is whether a pupil will attend a Grammar, Technical or Comprehensive school.

It has been pointed out however, that changes in some children's intelligence and/or scholastic attainments must be expected in the years subsequent to the 11+ examination. Vernon (1958) cites studies which indicate that at least five percent of the placements are unsuitable. This may be because with increasing age and maturity, "the range of ability has become far wider ... and some specialization has to be introduced. Hence, streaming by ability and, to a limited though increasing extent by speciality, would seem legitimate" (Vernon, 1957, p.48).

United Kingdom studies of the prediction of general or overall scholastic achievement at secondary school level have been summarized by Clark (1958, p.8-19) who cites Vernon (1957) in support of his conclusion that the usual combination of Intelligence, English, and Arithmetic tests reaches a high degree of validity. In reviewing the literature on prediction of success in single subjects Clark (p.19-22) concludes that, with much less research data available, confident prediction is much more difficult.

A six year follow-up of a group of 11+ examination candidates by Nisbet and Buchan (1959) supports the continued use of the 11+ battery of tests for the prediction of general scholastic achievement at secondary school leaving level. This was the case when they considered their sample as a whole. But when it was subdivided into groups according to University courses later undertaken substantial differences in the size of the correlation coefficient were found to exist between these groups (the range of r was from -.02 for future medical students to .58 for future Arts students). Thus while their results support the general long-term validity of the tests Nisbet and Buchan warn of the "fallability of early assessments and the need for a highly flexible organization of secondary school courses" (p.7).

Vernon's (1961) description of the "structure of ability" provided a framework which accommodates United Kingdom educational classification practices. He proposed a modification and to a degree, a blending of the two theoretical structurings of ability and traits described by Spearman (1927) and Thurstone (1935, 1938).

Spearman maintained originally that all intellectual functioning involved to some degree a common general factor, and that any one particular activity also involved a specific intellectual factor unique to it. Mental ability was accounted for by a two-factor (general 'g' + specific 's') theory. Thurstone on the other hand initially sought to avoid the use of a general factor and instead to account for mental functioning in terms of factors which were intended to represent relatively independent (non-correlated) abilities.

Vernon proposed an hierarchical scheme with a general factor at its top or most general level. At the second level he has two very broad group factors, 'v:ed' (verbal-educational aptitude) and 'k:m' (practical-mechanical aptitude). These two factors are then subdivided into less broad factors at the third level; the 'v:ed' factor yields numerical ('n') and verbal ('v') subfactors. These subfactors are further subdivisable into still narrower subfactors, and ultimately at the lowest level are the specific factors. The 11+ system of testing can be seen to readily fit this scheme.

The appropriateness of the 11+ classification system for the more academic pupils was demonstrated by Lewis (1961). From a centroid analysis of subject marks obtained by 173 boys after three years at Grammar school he obtained one general factor ("educational ability") and three group factors ("mathematical"; "scientific", comprising physics, chemistry and biology; with the third group factor having heavy loadings on History, Geography, English, French and Latin, there being some evidence for partial separation of History and Geography from the others).

Although Armstrong (1964) found that rural children are in some ways at a disadvantage compared with urban children when selection tests are used at the 11+ level such differences were small, and Armstrong specifically suggested that no changes be made in traditional selection methods.

In brief, United Kingdom experience and research indicates that a sufficiently satisfactory working relationship exists between the 11+ selection and classification tests and later general scholastic attainment to justify continuation of this tradition, particularly for Grammar school pupils. Available evidence shows that prediction in single subjects is less satisfactory, and that changing patterns of ability may confound long-term prediction for selected groups.

2. United States.

In the United States there has been considerable interest in the development of multiple aptitude batteries for educational classification. However, in a review of research findings in this area Anastasi (1961) feels that not until the high school age (middle and late teens) has differentiation of stable abilities progressed far enough to allow the practical use of classification tests and batteries.

Anastasi notes, with regard to educational aptitudes, that "the large common contribution of verbal comprehension to achievement in all academic areas has been repeatedly demonstrated" (p.348). Consequently most verbal intelligence tests are relatively poor for classification purposes since they predict success almost equally well in most areas. Thus an individual scoring high on

such a test would be classified as successful for many scholastic assignments and it would be impossible to predict in which he would do best. The single score from a test like the Otis may indicate whether or not a pupil should attempt an academic course but would be of little value in showing for which one of a number of different academic courses (combinations of subjects) the pupil would have most aptitude.

The stimulus for the development of multiple aptitude procedures was initially provided by Thurstone (1935, 1938), who sought to analyse "ability" into discrete factors or "Primary Abilities" which were comparatively independent of one another.

According to Anastasi (1961) the primary abilities which are more frequently confirmed in research are Verbal Comprehension (V), Word Fluency (W), Number (N), Space (S), and Associative Memory (M). Much less support has been found for a Deductive Reasoning factor (D), but there is some for an Inductive Reasoning factor (I), this latter usually being measured by items requiring the testee to establish a general rule logically relating the elements of a given series before the problem can be solved. Factor analytic research by Kettner et al (1956, 1959) also provides evidence of a General Reasoning factor which they found accounted for much of the variance in arithmetic reasoning items.

Guilford (1959) has proposed a three-dimensional factorial model to describe intellect; this makes provision for 120 factors. It is clear that some factorists are dissatisfied with what they feel to be the over-simplified

descriptions of intellectual structure which some of the earlier theorists offered.

Others who have concerned themselves with the theoretical structuring of intellectual functioning do not agree that such refined models as Guilford's provide the basis for practical psychometric application. Of Guilford's work Eysenck (1967) writes: "To some critics, this factorial extension of Thurstone's work has appeared almost as a reductio ad absurdum of the whole approach. There is a possibility of infinite sub-division inherent in the statistical method employed, and evidence is lacking that further and further subfactors add anything either to the experimental analysis of intellectual functioning or the practical aim of forecasting success and failure in intellectual pursuits" (p.82).

McNemar (1964) is of the opinion that "better predictions are possible via old fashioned general intelligence tests than through multi-test batteries" (p.875), and he claims that a model such as Guilford's fails to represent the essentially hierarchical nature of intellectual functions derived from the consistently positive correlations between relevant tests and between the various factors extracted.

Mann and Phillips (1967) who also have reservations about efforts to establish "primary" mental abilities and factors state, that despite repeated efforts to isolate subareas of intellectual functioning, "measures of intelligence tapping 'g' rather than 's' still remain the most meaningful assessors and predictors of performance" (p.312).

For the specific purpose of classifying school pupils some factors

have been found more appropriate than others for inclusion in tests. For such classification purposes Anastasi (1961) suggests that "a few broad factors such as verbal comprehension, numerical facility, and general reasoning would be most relevant" (p.345). She considers that the most direct effect of factor analysis upon test construction has been the development of multiple aptitude batteries, and cites the Primary Mental Abilities (P.M.A.) as an example of a battery derived directly from factor analysis and the Differential Aptitude Tests (D.A.T.) as one constructed with the results of factor-analytic research findings in mind although not using factor-analytic tests for its construction.

In brief, both United States and United Kingdom research and practice to date indicates that grouping or streaming of pupils according to general ability at early secondary school level can be carried out with some confidence by using a general intelligence test, usually with a heavy verbal weighting. Classification decisions are much less successful however, even when multiple aptitude batteries are used. Available tests give an adequate indication of a pupil's general academic potential; they do not indicate in which particular area of academic study or in which particular subjects he is most likely to be successful.

3. New Zealand.

At any time during his movement through the education system it is important that a pupil be assigned to courses which are appropriate to his interests, present levels of scholastic functioning, and potential for further

educational progress. Assignment decisions made during the transition from primary to secondary schools when a range of courses is offered are of major importance. Use of tests, together with reference to school progress records, can help ensure that a relatively objective basis for offering guidance exists. Further, by reducing the range of individual differences within any one group of pupils, classification procedures help provide an educational/instructional environment which is more likely to benefit all the individuals comprising that group.

The tendency in New Zealand is for secondary schools to follow a pattern analogous to the United Kingdom 11+ examination, i.e. a general intelligence test plus assorted attainment tests, usually English and Arithmetic, as a guide for making streaming and classification decisions affecting new entrant third form pupils. Typically, such classification is carried out prior to pupils being assigned to any groups.

It is also possible to classify on the basis of school achievement during a trial period. This procedure however may involve subsequent major course or level changes with disruptive social and scholastic effects for some individuals. A full discussion of the advantages of various types of classification of new entrants to secondary schools has been offered by Boag (1960).

Two major group tests which can be used with third form pupils and which were initially constructed and standardized in the United States have been restandardized in New Zealand. They are the Otis Self-Administering (Intermediate Examination) Test, Form A, and the Primary Mental Abilities.

A nationwide restandardization of the former took place in 1936 and

is reported by Redmond and Davies (1940). By analogy with the norms obtained for this Intermediate form New Zealand norms for the Otis Higher Examination were subsequently prepared by the New Zealand Council for Educational Research.

A restandardization of the P.M.A. test using a sample of boys and girls from the Auckland province was reported by Rogers (1956). In contrast to the Otis, which yields a single raw score and an equivalent age-based IQ, the New Zealand standardization of the P.M.A. yields five part scores (Verbal Meaning (V), Spatial (S), Reasoning (R), Number (N), Word Fluency (W)) which are combined by the equation: Total = V+S+2R+2N+W, to give a total score and an equivalent IQ. It is in fact a multiple aptitude battery.

In terms of structure the Tasman tests comprising the Tasman Junior Reasoning Test (TJRT, Appendix B) and the Tasman Verbal Ability Test (TVAT, Appendix C) devised by Shouksmith in the early 1960's fall between the Otis and P.M.A. tests. A total score is obtained by a weighted combination of the two. Part A of the two-part TVAT is very similar to the P.M.A. Verbal Meaning test while the TJRT uses items of the series completion type similar to the P.M.A. Reasoning test.

Another recent development in New Zealand has been the General Ability Test (G.A.T.) also intended as an aid to secondary school pupil classification. This as yet unpublished test being prepared by Arvidson (1959) contains items of English, Arithmetic, and General Intelligence types.

There has been little research into the validity of these tests as tools for aiding educational classification decisions. McGarvey (1962) found for

340 new entrants to secondary school that Arvidson's G.A.T. correlated more highly (r=.81) than did Otis (r=.73) with achievement at the end of the first term. This study also showed that considerably fewer pupils needed reclassifying at the end of the first term when objective tests were used for their initial classification than had been the case when no such testing was done.

Two studies of Otis results in relation to School Certificate Examination (S.C.E.) marks have been reported. For a small sample (N = 56) Riske (1948) found the correlation between Otis IQ on entry and S.C.E. Total mark three years later to be .72. Riske concluded "with some assurance" that "a well-conducted Otis test on entrance gives a good prediction of School Certificate Examination success at the end of the third year" (p.73).

Clark (1958) reported on the relationship between Otis Higher Form B IQ's obtained by new entrants at an urban co-educational school and S.C.E. total and subject marks obtained three years later. His sample (N=670) was drawn from those pupils who entered the third form from 1948 to 1955 and who sat the S.C.E. in the period 1950 - 1957.

Table 1 gives the correlation coefficients (r) between Otis IQ and S.C.E. marks obtained by Clark for combined (boys and girls) samples (N). Total mark was the sum of the marks obtained in English and the three best of the other subjects offered for examination. In calculating r for an individual subject Clark included in his sample <u>all</u> candidates who offered the subject for examination, i.e. N included those candidates whose marks in that subject were their lowest and were not incorporated in their Total marks.

Table 1 Correlation Coefficients: Otis IQ with S.C.E. Marks (from Clark, 1958)

Mark	N	r with Otis IQ
Total	670	.577
English	670	.544
Geography	651	.373
History	559	.394
Chemistry	427	.370
Mathematics	336	.546
Book-keeping	166	.392
French	151	.507

The S.C.E. subjects tabulated above were all "sat by both girls and boys in substantial proportions". No significant sex differences were found at the 1% level of confidence, and only two at the 5% level: both English and Geography marks correlated more highly with Otis IQ for girls than for boys.

It was noted by Clark that his study "suffers to some extent from the fact that only one school has been covered", and he suggested that it could be "worthwhile to extend the present investigation to schools of the same and other types" (p.59). His sample was "a selection of the more able pupils from the total school entry" (p.50). The sample mean IQ was 105.7 and the standard deviation (S.D.) 9.1, and Clark argued that this would have caused the obtained correlations to be lower than would be expected if the whole school population had been considered. It is of interest that although "School Certificate syllabuses remained relatively stable during the period 1950 - 1957" (p.24) a footnote referred to new syllabuses in Geography (1954) and History (1956).

In brief, New Zealand practices in classifying secondary school new

entrants have largely followed the United Kingdom 11+ testing pattern with only limited research into the relationships between test scores and later scholastic achievement. The results of New Zealand research however, are essentially consistent with United Kingdom findings.

CHAPTER III

THE PROBLEM

In general terms the major problem investigated was whether or not there exist any differences between the Tasman and Otis tests as predictors of scholastic attainment at secondary school. As used here "attainment" refers both to an overall level of scholastic achievement and to levels of achievement in single subjects. In the event of Tasman being shown to be equal or superior to Otis in making such predictions further refinement of the as yet embryonic Tasman would be indicated as a worthwhile project in local test development.

Because few detailed theoretical justifications have been advanced for using either of these tests as predictors of attainment (and particularly in single subjects) there appeared to be little point in investigating hypotheses of the form "Tasman will predict attainment in subject X better than will Otis" or vice-versa. The stating of hypotheses in such specific unidirectional forms could have encouraged speculative and probably unprofitable after-the-event "explanations" of obtained results, explanations which could not have been at all substantial in view of the lack of adequate theoretical rationale for using these tests as achievement predictors.

Even if Tasman were to be better than Otis as a predictor of attainment in subject X such a result would not necessarily validate Tasman rationale and disprove Otis rationale — both could be equally unsatisfactory as

theoretical explanations of the empirically obtained prediction coefficients and differences between these. Little will, therefore, be added to psychological or educational theory and principles by this study.

There remained, however, a need to examine New Zealand testing practice in secondary schools, particularly since the major tool currently used (the Otis test) was not designed specifically for this purpose, lacks recent norms and adequate validation, and takes no account of the movement towards the differential testing of abilities. Further, in view of the self-acknowledged limitations of the one substantial New Zealand study in this area (Clark, 1958) another validation of the Otis using School Certificate Examination marks as criteria seemed appropriate.

The essential problem to be examined then, was whether there exist any significant differences between the Tasman and Otis tests as long-term predictors of scholastic achievement.

CHAPTER IV

INVESTIGATING THE PROBLEM

1. General Procedure.

In February 1963 Tasman and Otis tests were administered to 2,180 third form new entrants at ten secondary schools. Tasman tests were administered as part of a larger testing programme which was intended to provide test standardization data. Otis tests were administered as part of each school's regular testing and classification programme for new entrants.

In November 1965 after nearly three full years of secondary school attendance 1,324 of these pupils became candidates for the School Certificate by sitting the nationwide School Certificate Examination (S.C.E.) in four or more subjects. Detailed results for these pupils were obtained from the Department of Education's Examinations Division after Pass lists were published in January 1966.

Pupils' S.C.E. marks, Tasman test scores and Otis IQs were punched out on data cards and processed by I.B.M. 1620 computer using a programme for calculating correlation coefficients written by Wright (1962). Differences between obtained coefficients were then examined.

2. Measures Used.

(1) Predictor Tests

(a) Otis Self-Administering Test (Higher Examination) Form A

The Otis Self-Administering tests are group verbal intelligence tests, there being two levels of difficulty (Intermediate and Higher), each level having

two forms (A and B). They were first published in the United States in 1922, and adaptations for New Zealand conditions were published by the New Zealand Council for Educational Research in 1937, the first local manual "Otis Self-Administering Tests of Mental Ability: Instructions and Norms" appearing two years later.

These New Zealand adaptations were for two age groups.

1. Otis Intermediate, forms A and B, for ages 9 - 15 years.
2. Otis Higher, forms A and B, for ages 12 - 18 years.

Only the Intermediate Form A has been directly standardized in New Zealand. The raw data for this standardization was gathered in 1936, and Redmond and Davies (1940) in a full account of the procedures followed described a massive survey requiring the administration of this test to nearly 26,000 children attending 565 schools.

Both forms of the Otis Higher test contain 75 questions and are usually administered under group conditions, the time allowed for each form being 30 minutes. The Manual of Directions states that the Higher Examination is for use with children in Form II and secondary schools. The raw scores obtained are converted to age-based IQs having a mean of 100 and a standard deviation of 15 IQ points.

(b) Tasman Tests: Tasman Junior Reasoning Test (TJRT)
Tasman Verbal Ability Test (TVAT)

These two group tests were developed in the early 1960's by G. Shouksmith. Their purpose was to assist pupil classification at entry to

secondary school at the third form level. Of the two tests Shouksmith writes

> "together they measure overall general ability - and give us an IQ - and separately show differences between pupils whose ability is on the verbal side and those whose development can be expected to lie along the numerical, mathematical and possibly science side" (1963 a, p.504)

The TJRT (Appendix B) contains 50 items of increasing difficulty to be attempted in 25 minutes. Each item requires the testee to complete a series of numbers, letters, or words by finding the general rule governing the relationships between the given elements of the series ("inductive reasoning"). The TJRT is "a relatively difficult test" which "attempts to measure reasoning ability"; "its bias, which is deliberate, is towards the upper end of the scale of reasoning ability", and it "is not assessing the 'v:ed' side of the factor pattern" but "is concerned roughly with 'g' + 'n'" (Shouksmith, 1963 b).

The TVAT (Appendix C) is in two parts. Part A, Vocabulary, consists of 40 multiple choice word knowledge items to be attempted in 10 minutes. Part B, Analogies, comprises 25 items to be attempted in 5 minutes and requires the testee to complete a wide variety of analogies by drawing upon his accumulated general knowledge and information.

The connections between the various parts of the Tasman battery can be shown in equation form.

TVAT A (Vocabulary Score) + TVAT B (Analogies Score) = TVAT (Total Score)

TVAT (Total Score) + 2xTJRT Score = TOTAL Scaled Score.

It was intended that an IQ would be derived from the Total Scaled

Score.

Thus, when the two Tasman tests are taken together

> "an overall score which gives a global IQ can be obtained. Taking the scores individually, the pupil can be guided more appropriately into the correct course. Thus the two Tasman tests together can be used for general classification of all third form pupils on entry to post-primary schools" (Shouksmith, 1963 a, p.505)

(c) Limitations of the Otis.

Although Redmond and Davies observed that tests like the Otis "are playing an increasingly important part in both educational and vocational guidance", and are "an aid in school classification and in deciding upon the kind of education for which the child is most suited" (1940, p.2), the original Otis was not in fact designed for the specific purpose of making classifications in New Zealand secondary schools.

Because it is one of the few group intelligence tests readily available in New Zealand the Otis is used repeatedly and so "often loses its effectiveness as an independent measure" (Shouksmith, 1963 a, p.505).

"The single score Otis test ... cannot provide the further information which informed educators nowadays require, differential assessments of the major aspects of ability" (Shouksmith, 1964, p.178). The Tasman tests "represent an improvement on traditional and older testing methods, of which Otis is but one, in that ... the two tests present differential information ... which enables the teacher to make recommendations about the kind of school course, as well as the

level of course, that a particular pupil should undertake" (Shouksmith, 1964, p.178).

"Otis test claims to be a measure of 'general ability' yet a whole area of general ability is missed out" (Shouksmith, 1964, p.178). Lefever, writing of the Otis in The Fifth Mental Measurements Yearbook (Buros ed., 1959), warns that "the abilities of pupils whose interests and talents are distinctly along 'nonbookish' lines may be incorrectly appraised by this type of test" (p.498).

"The fact that a standardization sample of huge size has been made does not make the standardization valid for any longer period of time" (Shouksmith, 1964, p.179). At the time of the Otis Intermediate standardization in 1936 primary school education in New Zealand was compulsory only until either the age of 14 years or the completion of the primary course at standard six. (The Proficiency Examination was only abolished in 1936). Of all primary school leavers in 1936 only 59% proceeded to secondary schools. The educational provisions and requirements of 1936 differ appreciably from current trends and conditions when near-automatic "social promotion" from primary (or intermediate) to secondary schools prevails and when the minimal school leaving age is 15 years (raised in 1944). Because of these changes the 1936 standardization would almost certainly not be replicable today.

The Manual of Directions for the Otis Self-Administering Tests notes that "the Otis Higher has not been standardized in New Zealand".

Redmond and Davies (1940) observed re the Otis that "the 'ceiling' effect is noticeable ... the older very bright child (cannot) distinguish himself

from the fairly bright children of his age group" (p.106). In contrast to this Shouksmith (1962) emphasizes that in the Tasman tests, and particularly in the TJRT, "there is plenty of 'ceiling' ... which allows for differentiation at the top end of the third form stream. In many ways, and particularly from the point of view of picking out possible sixth form pupils of the future, this is more important than that the test should differentiate at the lower level" (p.1).

The Otis norms are presented graphically using monthly age divisions in a way which could suggest a precision which is not really present in the test. The scheme adopted for the Rogers New Zealand standardization of the P.M.A. (1956) in which yearly age group divisions are used seems more appropriate for the interpretation of scores from a speeded group test.

(d) Limitations of the Tasman.

No manual to accompany the tests has been published. Finalized forms of administration and scoring procedures, details of test construction and standardization, tables of norms, guides for the interpretation of scores, reliability and validity data are therefore lacking. Parallel forms of neither the TJRT nor TVAT are available, and so the Tasman tests are particularly prone to the effects of practice through repeated exposure and coaching.

An Interim Manual for the Canterbury Junior Reasoning Test (now the TJRT) was prepared in cyclostyled form, and in it test administration and scoring procedures as well as tables of norms were given for "a third form group of an approximate age range of 13-6 to 14-6. The number in the sample was

approximately 300 and the sample appears to be fully representative of an urban population" (Shouksmith, 1962, p.3).

A subsequent set of (also unpublished) norms based on a "larger" but unspecified sample was prepared under the heading Tentative IQ Tables (Shouksmith, 1963 b) By using these norms a Total Scaled Score (= TVAT + 2 x TJRT) is convertible to an IQ, and raw scores on the TVAT and TJRT are convertible to grades and percentile groups. There are considerable discrepancies between the 1962 and 1963 b norms for the TJRT and the quality of even the latter (last available) norms is brought into doubt by a footnote to the IQ conversion table: "it is emphasized that these norms are entirely tentative and will be revised later". Unfortunately, no later revision has been published.

Unlike Otis norms which provide an age-based IQ, there is no division of available Tasman norms by age groups. The only common factor of the samples used for the development of available tentative norms is that all are comprised of new entrant third form pupils.

The differences between boys and girls on the Otis (Intermediate) are, for some age groups statistically significant, but in no instance do these differences exceed 1.6 IQ points in favour of the girls (Redmond and Davies, 1940, p.69). No statement of this type of information appears to have been made by Shouksmith for the Tasman tests. However, an examination by this writer of TVAT and TJRT raw scores for a sample of over 3,000 third form pupils attending 15 secondary schools indicates that only on the TVAT B is a statistically significant sex difference found. This difference is small, the

average score for boys being .3 of a mark (about one-tenth of a standard deviation) more than for girls.

(2) Criterion Tests: The School Certificate Examination (S.C.E.)

(a) History of the S.C.E.

The S.C.E. was instituted in 1946. The Thomas Committee Report (1943) had recommended that the examination should follow a four year course at secondary school for the average pupil, but indicated that an able minority could sit it successfully after three years. The School Certificate was to be awarded to pupils who aggregated 200 or more marks for English and their three next best subjects, with a minimum of 30 marks in each subject.

After a decade of experience with the S.C.E. the various purposes which it appeared to fulfil were stated in the Wallis House Committee Report (1956).

> "For many pupils it represents a goal to be attained before their post-primary schooling is completed; and for those who intend to proceed to higher education it is at once a test of the progress they have made and an indication of the success they are likely to achieve in future study" (para.16).

Following the School Certificate Review Committee Report (1960) and the recommendations of the Commission on Education (1962) the S.C.E. is to be modified so that after 1967 candidates will be able to offer for examination in any one year any number of subjects from one to six rather than the four to six as previously.

(b) The S.C.E. as a Measure of Educational Goals.

Some educators question whether the S.C.E. really measures the degree to which teaching has helped pupils to achieve the wider goals of education. They suggest that the use of such an examination encourages a narrow focussing of attention on examinable, academic behaviour.

However, the S.C.E. certainly provides a nationwide standard against which the individual pupil's competence in his chosen subjects can be assessed. In this study the S.C.E. is viewed simply as a measure of scholastic achievement with no suggestion that it measures educational and teaching outcomes in a general sense. The discussion by Adams (1964) of the relationship between "intermediate" and "ultimate" criteria is of relevance in this connection.

(c) The S.C.E. and the Community.

The School Certificate Review Committee Report (1960) noted that since the S.C.E. is "the only national measure of educational attainment available to the community before the sixth form ... nearly all parents want their children to embark upon School Certificate courses, often regardless of ability and of the advice of school authorities" (p.20).

The Report also noted that parents have further influenced school policy by often demanding that their children sit the S.C.E. after three years rather than the four envisaged by the Thomas Committee.

The observation in the 1960 Report that "In the community at large the actual certificate is accepted as a measured standard of attainment in post-primary studies" (p.13) remains true. Further, both public and private employers

frequently use the attainment of the S.C.E. as a major screening tool in selecting trainees and employees for the salaried professions and skilled trades.

Tables 2 and 3 indicate quite clearly community reaction to the examination itself. In 1965 about 90% of all (first and later year) fifth formers were S.C.E. candidates; about 84% of first-year fifth formers were candidates. Over 95% of the first-time candidates were sitting after only three years of secondary school attendance — very few now wait four years before first sitting as was recommended by the Thomas Committee Report.

Table 2 S.C.E. Candidates as a Percentage of All Fifth Form Pupils (1946 - 65)

Year	%
1946	65
1952	74
1960	83
1965	90

Table 3 Percentage of First-time Candidates Sitting the S.C.E. in Third Year — (1949 - 65)

Year	% Boys	% Girls
1949	76	66
1952	87	74
1960	96	92
1965	97	94

(d) The S.C.E. : 1965.

Candidates were able to select from 37 subjects, English being compulsory. Optional papers were available in General Science (three options),

Mathematics, Physics and Technical Drawing (two options in each). The subjects most frequently offered for examination were, in order, Geography, Mathematics, General Science, History, French and Biology, i.e. "general academic" subjects, rather than the more specialist academic, e.g. Physics, cultural, e.g. Music, and technical, e.g. Electricity, subjects.

(e) 1965 S.C.E. Marks Used as Criteria.

For each candidate the following marks were used as criterion measures.

(i) English.

(ii) The three highest marks obtained in other single subjects offered for examination.

(iii) Total, calculated by aggregating all single subject marks used as criteria, i.e. (i) + (ii).

Two of the Department of Education's qualifications, namely that a candidate must (1) score 30 or more marks out of the 100 possible in a subject before such marks can be included in the aggregate or total and (2) score 30 or more marks in English if his aggregate is to be other than zero, were not observed.

Thus the total mark used in this study was derived from the same number of subjects (four) for all candidates, this probably being more representative of "overall" scholastic achievement than the total mark as calculated by the Department's formula.

Although most candidates offered English plus four other subjects for examination the mark in the fifth subject was omitted from consideration in this study since many candidates offer it in the knowledge that it is their weakest and therefore unlikely to be included in the total mark, and in many cases make significantly less preparation for it than for the four others.

3. The Sample.

The S.C.E. candidates sample comprised 1,324 first-year fifth form pupils who were drawn from a sample of 2,180 third formers who, in 1963, were new entrants at 10 urban state secondary schools (4 boys schools, 2 girls schools, 4 co-educational schools). The median age of all new entrants to New Zealand secondary schools in 1963 was 13 years 4 months, and although it is not known the median age for the sample would probably be similar to this.

Table 4 shows characteristics of the sample in relation to the new entrant population. One feature is that a higher percentage of the new entrant sample attempted the S.C.E. than was the case in the population as a whole. Another feature is that the sex distribution in the sample does not correspond to that in the population. Possible effects of these characteristics are taken account of in the results (Chapter 5) and discussion thereof (Chapter 6). The sample can be seen to comprise nearly 5% of the population of first-year fifth form candidates.

Table 4 Composition of Sample and Population

	Sample	Population
1963 New Entrants	1325 Boys 60.7% 855 Girls 39.3% 2180 Total	25574 Boys 51.5% 24009 Girls 48.5% 49584 Total
1965 First-Year Fifth Form S.C.E. Candidates	819 Boys 61.8% 505 Girls 38.2% 1324 Total	14786 Boys 53.0% 13110 Girls 47.0% 27896 Total
Candidates as a % of New Entrants	61.8% Boys 59.0% Girls 60.7% Total	57.8% Boys 54.6% Girls 56.3% Total

Table 5 below gives means (M) and standard deviations (S.D.) for the scores of the S.C.E. sample on the predictor tests. Whereas the Otis has been standardized to yield a distribution of IQs with M=100 and S.D.=15 no such information has been formally compiled for the Tasman. However, from a normally distributed unselected sample of Tasman total scaled scores this writer obtained values of M=68.1, S.D.=21.5 (N=3,054; 1428 boys, 1626 girls).

Both Tasman and Otis mean values are significantly greater for the sample than for unselected groups. This would follow from the fact that a substantial proportion of pupils who leave secondary school during the first three years would fall in the below-average intelligence group. This would affect the distribution of scores at the lower levels particularly, with a shrinkage of S.D.s as occurred in this sample.

The only significant difference in M values between the sexes occurs with TVAT B scores (at the 1% level of confidence), and is in the same direction as that found in the unselected sample refered to above (significant at the

5%level). However, the actual difference of .4 of a mark in favour of boys is small (about one-eighth of a S.D.).

It is clear that the S.C.E. sample is a selected one, and the effects of such selection will be examined in the discussion of the results (Chapter 6).

Table 5 Means and Standard Deviations of Predictor Tests

Test	N=1324		819 Boys		505 Girls	
	M	S.D.	M	S.D.	M	S.D.
TJRT	20.0	6.8	19.7	6.9	20.4	6.7
TVAT A	22.5	5.9	22.4	6.1	22.6	5.6
TVAT B	15.1	3.4	15.2	3.5	14.8	3.3
TVAT	37.5	8.5	37.6	8.7	37.3	8.1
Tasman	77.4	19.4	77.0	19.5	78.1	19.1
Otis	106.4	13.0	106.0	12.9	107.2	13.1

Table 6 below gives the percentages of S.C.E. candidates in the sample and population for the more popular subjects. While the trend in the sample is basically similar to that in the population the sample does have a higher proportion of candidates offering the "academic" subjects (MA, HI, FR, CH, PH) with somewhat lower proportions in the more "general" subjects (GE, GS, BI).

Table 6. Percentages of Candidates Offering S.C.E. Subjects

Subject	Abbreviation	Sample	Population
English	EN	100.0	100.0
Geography	GE	58.0	68.0
Mathematics	MA	66.1	54.6
General Science	GS	41.9	44.5
History	HI	45.0	40.6
French	FR	37.7	32.0
Biology	BI	22.1	31.8
Commercial Practice	CP	17.2	18.8
Book-keeping	BK	16.9	15.7
Chemistry	CH	19.7	13.9
Physics	PH	16.6	9.2
Drawing + Design	DD	9.0	8.0

Table 7 gives statistics for the more popular S.C.E subjects as contained in the sample. Marks obtained by the population of candidates in each subject are scaled by the examiners to a mean of 50, leading to a mean total mark of 200. It can be seen that means for MA, FR, and CH samples are substantially above the population average of 50 while the BI mean is well below average. A particularly wide spread of scores occurred in the GE and MA samples.

Table 7 Statistics for S C E — Subject Samples

Mark	N	M	S.D
Total	1324	203.8	62.1
EN	1324	48.7	16.1
GE	642	51.7	19.9
MA	581	56.1	20.9
GS	456	48.8	18.2
HI	484	52.1	17.7
FR	312	58.4	16.6
BI	217	45.3	16.3
CP	202	47.0	16.2
BK	188	49.0	17.6
CH	165	58.6	17.3
PH	148	51.2	16.0
DD	98	51.1	15.8

Table 8 below gives r values for pairs of predictor tests for cases in which there is no item overlap, i.e. one test of the pair is not a component of the other. With one exception (TVAT A with TVAT B) correlations between the tests are consistently higher for girls than for boys. None of the sex differences are significant, however.

Other major points shown by Table 8 are as follows.

(i) Although there is a moderate amount of variance shared by TVAT and TJRT (r=.503), this is clearly less than that shared by, for example, TVAT and Otis (r=.700). This indicates that TVAT and TJRT are measuring somewhat different aspects of mental ability, as was intended by their author.

(ii) This first point is emphasized by the relationship which Otis bears to TJRT (r=.600) and TVAT (r=.700). This shows that, if Otis is accepted as a typical test of verbal-educational intelligence, the TJRT has a much smaller v:ed component than the TVAT, as was also intended by the test's author.

(iii) There is relatively little difference between Otis correlations with TVAT and with the full Tasman. This follows from the heavy weighting given the TJRT in arriving at the Tasman Total Scaled Score and the relatively low (Otis, TJRT) correlation.

Table 8 Correlations Between Predictor Tests

Tests Correlated	N=1324	819 Boys	505 Girls
Otis, Tasman	.731	.723	.745
Otis, TVAT	.700	.698	.710
Otis, TVAT A	.653	.648	.663
Otis, TVAT B	.610	.612	.621
Otis, TJRT	.600	.582	.627
TJRT, TVAT	.503	.484	.541
TJRT, TVAT A	.444	.420	.488
TJRT, TVAT B	.484	.477	.514
TVAT A, TVAT B	.626	630	.622

This chapter has outlined procedures used in the present study, discussed features of the predictor tests and criterion measures used, and examined characteristics of the sample. Chapter 5 presents the main results and a discussion of these is offered in the first part of Chapter 6.

CHAPTER V

RESULTS

The tables and results presented here summarize the main findings of the investigation. All correlation coefficients reported are Pearson's product-moment values. Throughout the remainder of this report differences significant at above the 5% level of confidence but which do not reach significance at the 1% level are marked thus: *. Differences significant at or above the 1% level are marked thus: **.

1. Differences between Tasman and Otis in Predicting S.C.E. Marks.

Table 9 below gives (positive) differences between Tasman and Otis correlation coefficients. Significance or non-significance of a difference ($r_{12} - r_{13}$) was estimated by the formula

$$t = \frac{(r_{12} - r_{13}) \sqrt{(N-3)(1+r_{23})}}{\sqrt{2(1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2 r_{12} r_{13} r_{23})}}$$

suggested by McNemar (1955, p.148) for the case when both r_{12} and r_{13} are based on the same sample N, and r_{23} is known.

Correlation coefficients from which the differences in Table 9 are derived are detailed in Appendix D.

Table 9 Differences between Otis and Tasman Tests in Predicting S.C.E. Marks

r difference	Total	EN	GE	MA	GS	HI	FR	BI	CP	BK	CH	PH	DD	TD	ES	WW	ST	CL	HC
r(Otis) - r(TJRT)	.072**	.159**	.097**		.006	.048	.158**	.094			.034	.168*	.043		.145				.037
r(TJRT) - r(Otis)				.013					.004	.057				.155		.020	.050	.033	
r(Otis) - r(TVAT A)	.028		.015	.083**	.015		.113*	.017		.039	.017	.188*		.155	.171	.159			.264
r(TVAT A) - r(Otis)		.021				.036			.023				.033				.065	.127	
r(Otis) - r(TVAT B)	.029	.043**		.058	.018		.084		.024	.221**	.028	.149*	.062		.020		.042		.083
r(TVAT B) - r(Otis)			.039			.012		.036						.073		.107		.179	
r(Otis) - r(TVAT)				.027			.026			.056		.131*		.036	.065	.028			.180
r(TVAT) - r(Otis)	.031*	.067**	.057*		.038	.086**		.048	.049		.024		.031				.089	.126	
r(Otis) - r(Tasman)							.040					.089			.057				.062
r(Tasman) - r(Otis)	.053**	.022	.031	.086**	.086**	.066*		.033	.095	.092	.060		.055	.151		.082	.124	.302	

2. Differences between TVAT and TJRT in Predicting S.C.E. Marks.

Table 10 is derived from data in Appendix D. The significance of differences between r's was determined by the formula given on p.33. Where a difference is significant the larger r of the pair is asterisked in the usual way.

Table 10 Comparison of TVAT and TJRT r's in Predicting S.C.E. Marks

Mark	TVAT	TJRT	Mark	TVAT	TJRT
EN	.771**	.545	CH	.427	.369
GE	.501**	.347	PH	.341	.304
MA	.525	.555	DD	.472	.398
GS	.526	.482	TD	.267	.458*
HI	.527**	.393	ES	.261	.181
FR	.511*	.379	WW	.246	.294
BI	.465*	.323	ST	.577	.538
CP	.442	.397	CL	.282	.375
BK	.318	.431	HC	.187	.330

3. Sex Differences in Prediction.

Correlation Coefficients for boys' and girls' subgroups are given in Appendix E.

Only three of the 66 differences are significant, all at the 5% level of confidence. Book-keeping and Commercial Practice marks correlated more highly with TVAT B scores for girls than for boys, while the (BK, TVAT) correlation was also higher for the girls' subgroup.

4. Relationship to Clark's (1958) Study.

Table 11 below compares r values reported by Clark (1958) with

those obtained in this study.

In both studies the Total and EN samples were selected on the same basis In all six cases for these two marks the Patchett r value exceeded the corresponding Clark value, the differences being significant in EN prediction.

For the six other S C E subjects Clark's samples comprised all candidates who offered that subject, i.e. irrespective of whether the mark was amongst the best four and included in the Total mark. The Patchett sample in each of these subjects comprised those candidates for whom the subject mark was amongst their four best and therefore included in the Total mark. Since these samples were selected differently a statistical examination of differences in obtained r's was not considered advisable. In 12 of the 17 cases the r value for the Patchett sample exceeded that for the Clark sample.

Table 11 Otis Correlations with S.C.E. Marks: a Comparison with Clark's (1958) Findings

S.C.E. Mark	Sample Size						Correlation Coefficients					
	Clark			Patchett			N		Boys		Girls	
	N	B	G	N	B	G	Clark	Patchett	Clark	Patchett	Clark	Patchett
Total	670	377	293	1324	819	505	.577	.611	.576	.595	.584	.638
EN	670	377	293	1324	819	505	.544	.704**	.511	.695**	.625	.722*
GE	651	376	275	642	402	240	.373	.444	.331	.423	.418	.479
MA	336	248	88	581	445	136	.546	.542	.563	.543	.533	.585
HI	559	304	255	484	285	199	.394	.441	.384	.402	.409	.484
FR	151	79	72	312	132	180	.507	.537	.463	.496	.656	.565
BK	166	51	115	188	111	77	.392	.374	.408	.285	.430	.500
CH	427	373	54	165	156	9	.370	.403	.375	.411	.404	-

CHAPTER VI

DISCUSSION

1. The Results.

(1) Differences between Tasman and Otis in Predicting S.C.E. Marks

Table 12 below (derived from data in Table 9) summarizes the distribution of obtained differences between correlation coefficients for the various pairs of tests.

The most obvious trend shown in Table 12 is for Tasman scores to be slightly better predictors than Otis IQs (in 15 of 19 instances). This was despite the fact that the Tasman score is heavily weighted with the TJRT component which was a significantly less successful predictor than Otis in a number of cases.

On balance TVAT also has somewhat greater predictive power than Otis, but its components TVAT A and TVAT B, taken separately, are less adequate predictors than is Otis.

Actual size of the difference in either direction between Tasman and Otis r's exceeds .100 in only 3 of 19 instances. Even in the case of the significant difference (1% level of confidence) of .053 in favour of Tasman in predicting S.C.E. Total mark (r with Tasman = .664, with Otis = .611) there must be considerable caution in making inferences about the relative predictive efficiencies of the two tests.

It has been shown by Guilford (1946) that predictive efficiency can be represented by $E = 100 (1 - \sqrt{1 - r^2})$, where for a given correlation coefficient r, E (expressed as a percentage) gives the reduction in errors of prediction compared with the case of prediction made without knowledge of r.

For $r = .664$ $E = 25.2\%$

and for $r = .611$ $E = 21.0\%$.

In the case of an individual pupil the chances of correct placement are only slightly greater when the Tasman rather than the Otis score is used.

It should be noted that because the sample comprises only S.C.E. candidates it represents a selected group having higher means and smaller standard deviations on the predictor tests than is the case for the unselected new entrant population. Validity coefficients reported in Appendices D and E are therefore lower than those which would have been obtained had a representative rather than a selected sample of new entrants sat the S.C.E. The correlations obtained are reported in uncorrected form to allow ready comparison with other studies which typically present such correlations in this way. Further, correction for selection is not necessary for the purpose of comparing Tasman and Otis values.

Table 12 Distribution of Differences between Tasman and Otis Tests as Predictors of S.C.E. Marks

r difference	1% level	5% level	Non-sig.	Total
r(Otis) - r(TJRT)	4	1	7	12
r(TJRT) - r(Otis)			7	7
				19
r(Otis) - r(TVAT A)	2	1	10	13
r(TVAT A) - r(Otis)			6	6
				19
r(Otis) - r(TVAT B)	2	1	10	13
r(TVAT B) - r(Otis)			6	6
				19
r(Otis) - r(TVAT)		1	7	8
r(TVAT) - r(Otis)	2	2	7	11
				19
r(Otis) - r(Tasman)			4	4
r(Tasman) - r(Otis)	3	1	11	15
				19

The relative predictive efficacy of Tasman and Otis tests can also be examined by comparing the percentages of pass and fail candidates in corresponding parts of the tests' score ranges.

Table 13 below shows these percentages for three levels of test scores:

(i) below average (Tasman scores 67 or less; Otis IQs 99 or less)

(ii) between M and +1 S.D. (Tasman scores 68 - 89; Otis IQs 100 - 114)

(iii) above +1 S.D. (Tasman scores 90 or more; Otis IQs 115 or more).

Table 13 Percentages of Pass/Fail Candidates for Three Levels of Test Scores

S.C.E.		Level (i)		Level (ii)		Level (iii)	
		Tasman	Otis	Tasman	Otis	Tasman	Otis
Total mark	Pass %	23.0	26.5	58.0	56.9	88.6	85.2
	Fail %	77.0	73.5	42.0	43.1	11.4	14.8
N = 1324	N	414	409	541	536	369	379
EN	Pass %	12.1	13.7	49.0	48.1	88.9	86.8
	Fail %	87.9	86.3	51.0	51.9	11.1	13.2
N = 1324	N	414	409	541	536	369	379
GE	Pass %	33.9	34.0	56.5	59.5	83.1	73.7
	Fail %	66.1	66.0	43.5	40.5	16.9	26.3
N = 642	N	230	221	273	269	142	152
MA	Pass %	15.3	27.2	65.4	65.8	87.3	84.4
	Fail %	84.7	72.8	34.6	34.2	12.7	15.6
N = 581	N	98	110	231	228	252	243
GS	Pass %	27.0	33.6	49.2	48.7	84.6	76.9
	Fail %	73.0	66.3	50.8	51.3	15.4	23.1
N = 456	N	152	146	187	193	117	117
HI	Pass %	34.0	38.0	57.8	57.2	79.7	78.1
	Fail %	66.0	62.0	42.2	42.8	20.3	21.9
N = 484	N	135	137	206	210	143	137
FR	Pass %	44.0	33.3	55.9	60.8	84.1	84.2
	Fail %	56.0	66.7	44.1	39.2	15.9	15.8
N = 312	N	25	27	111	115	176	170
BI	Pass %	26.5	25.4	50.0	45.0	65.8	67.8
	Fail %	73.5	74.6	50.0	55.0	34.2	32.2
N = 217	N	68	59	108	102	41	56
CP	Pass %	30.0	31.2	59.5	56.5	65.2	60.6
	Fail %	70.0	68.8	40.5	43.5	34.8	39.4
N = 202	N	100	93	79	76	23	33

At level (i), differences between Tasman and Otis exceeded 5% in three cases: Tasman scores related more closely to failure in MA (11.9% lower pass rate) and GS (6.6% lower pass rate); Otis IQs related more closely to failure in FR (10.7% lower pass rate).

At level (i), Tasman scores were particularly good indicators of failure in EN (only 12.1% pass rate) and MA (15.3% pass rate). Otis IQs at level (i) also gave a good indication of EN failure (13.7% pass rate).

At level (ii), only differences of note were: Tasman scores related more closely to passing in BI (5.0% higher pass rate); Otis IQs related more closely to passing in FR (4.9% higher pass rate).

At level (ii) both Tasman and Otis were reasonable indicators of success in MA (65.4% and 65.8% pass rates respectively). Otis also gave a fair indication of success in FR (60.8% pass rate).

At level (iii) notable differences were in favour of Tasman scores which related more closely to success in GE (9.4% higher pass rate) and GS (7.7% higher pass rate) than did Otis scores.

At level (iii) both tests were good indicators of success, yielding above 75% pass rates in Total mark and in all single subjects except GE, BI, and CP.

Table 14a below summarizes scattergram data for S.C.E. candidates. The "drop-out" group of new entrants is therefore excluded. If all those new entrants who were not S.C.E. candidates in their third year at secondary school are assumed to be failures in the examination as a whole (Total mark) and in EN (compulsory), then data in Table 14b below are generated. This is of course an extreme assumption, and if "drop-outs" had sat the S.C.E. then data which would have been obtained would lie between that reported in Table 14a and that postulated in Table 14b.

Table 14a Four Scattergram Distributions of 1324 S.C.E. Candidates

		Tasman		Otis	
		Below Average	Above Average	Below Average	Above Average
S.C.E. Total mark	Pass	7.3%	48.4%	8.2%	47.4%
	Fail	24%	20.3%	22.8%	21.6%
EN	Pass	3.9%	44.8%	4.3%	44.3%
	Fail	27.4%	23.9%	26.6%	24.8%

Table 14b Four Scattergram Distributions of 2180 New Entrants

		Tasman		Otis	
		Below Average	Above Average	Below Average	Above Average
S.C.E. Total mark	Pass	4.4%	29.4%	5.1%	28.8%
	Fail	43.6%	22.2%	45.5%	20.6%
EN	Pass	2.3%	27.2%	2.6%	26.9%
	Fail	45.6%	24.9%	47.9%	22.6%

Differences between Tasman and Otis can be seen to be slight. Both tests yielded low percentages of "false positives", i.e. new entrants who were below average on the tests (and therefore predicted as likely to fail) but who in fact passed on the criterion measures. This low yield is desirable from the individual pupil's point of view, since it shows that there is relatively little chance of a pupil being refused the opportunity to undertake a course of study

which is within his capacity.

The percentages of "misses", i.e. new entrants who were above average on the tests (and therefore identified as likely to pass) but who in fact failed on the criterion measures, are much higher however. From an institutional decision point of view this is undesirable because "misses", in the present context, can be loosely interpreted as failure by educational personnel to assign pupils to courses and/or teach them adequately for the purpose of passing the S.C.E. Another interpretation could be that the tests are not adequate for the particular task of prediction; in this case an examination of alternative testing strategies, perhaps within the framework provided by decision theory, is indicated as desirable.

(2) Differences between TVAT and TJRT in Predicting S.C.E. Marks

Table 10 in Chapter 5 shows a trend for TVAT to be a superior predictor to TJRT (for 12 of the 18 S.C.E. subjects considered, in 5 cases at a significant level). This is particularly so for the popular general-academic subjects EN, GE, GS, HI, FR, and BI, with MA the only exception to the general trend.

Shouksmith's argument as reviewed in Chapter 4 was that the TJRT would give a better indication of the ability of pupils with non-bookish aptitudes than would verbal tests such as the Otis and TVAT. He considered that TJRT would measure 'g' + 'n' rather than the 'v:ed' part of the factor pattern. Six S.C.E. subjects correlated more highly with TJRT than with TVAT, with TD

providing the only significant difference in this direction. However, MA and BK have substantially although not significantly greater r's with TJRT, and the content of school courses and the S.C.E. in these three subjects indicates a degree of support for Shouksmith's argument.

In the case of secondary school new entrants who are intending S.C.E. candidates if there should be any doubt as to the suitability of a pupil including MA, BK or TD in his course more weight should be given to his TJRT score than to his TVAT one in making a decision.

Correlations of TVAT with the optional general-academic subjects GE, MA, GS, HI, FR, and BI are very similar (in the range .465 - .527). TJRT values on the other hand cover a much wider range with r = .555 for MA and .482 for GS being substantially greater than r's for GE (.347), HI (.393), FR (.379), and BI (.323).

In the case of a new entrant who has sufficient general ability (Tasman score) or 'v:ed' ability (TVAT score) to cope with the levels of work required in all these general-academic subjects but who is in doubt as to which particular ones he should study some reasonable weight could be given his TJRT score if a decision is required about the inclusion of MA and/or GS in his course. A high/low TJRT score would be a better predictor of success/failure in these two subjects than in the other general-academic ones.

(3) Sex Differences in Prediction

Of 66 differences examined in Appendix E only 3 were significant,

all 3 being in favour of girls at the 5% level of confidence. The relatively small numbers in the sex subgroups for the two subjects concerned, CP and BK, and the fact that the correlation coefficients concerned for boys sitting BK failed to be significantly greater than zero and just reached significance at the 5% level in CP would suggest caution in claiming such findings as representative.

This general agreement in sizes of r for the sexes taken separately indicates that values of r for the sexes combined (as reported in Appendix D) do not conceal any substantial differences between boys and girls groups.

There is a pronounced trend however for correlation coefficients for all predictor tests to be greater for girls than for boys (Table 15 below). In particular, all 11 Otis r's are greater for girls. The general trend possibly reflects the earlier maturation of girls, while this developmental factor will be accentuated using an age-based test like the Otis.

Table 15 Distribution of Larger r: by Sex

Predictor Test	Boys r larger	Girls r larger
TJRT	3	8
TVAT A	1	10
TVAT B	2	9
TVAT	1	10
Tasman	2	9
Otis	0	11
	9	57

(4) Relationship to Clark's (1958) Study

In 18 of the 23 cases shown in Table 11 the correlation coefficients obtained in the present study were greater than those reported by Clark. This

observation is rather surprising because the basis of selection for the Clark samples was somewhat broader than in this study, but could reflect the fact that Clark's sampling took place over a six year period during which criterion measures would have changed annually (different S.C.E. questions) and periodically (revised curricula).

A further possible explanation is that by including candidates' fifth/lowest marks in his samples Clark introduced a factor which would depress the size of correlations. This would follow since, before even sitting the S.C.E., many candidates regard their mark in one of the four optional subjects as expendable and they make correspondingly little effort to prepare it adequately for examination. On the other hand other candidates may make particularly thorough preparations for examination in a fifth subject should there be a chance that two or more subjects are sufficiently weak to be likely to fall below the 30% cut-off mark required by the Department of Education. In either case the S.C.E. mark in the fifth subject would be atypical, depressed in the former instance and inflated in the latter, and not as representative of educational status in that subject as would be marks in other subjects.

2. Intervening and Uncontrolled Variables.

In this study, as in many of prediction in the applied field, there was a period (in this case of nearly three years) during which no control was exercised over events and circumstances which could have been relevant to its outcome. Travers (1964) feels that prediction is possible because of relatively enduring qualities within the individual, and that improvement in prediction will

stem from greater control of external conditions. One way of making such control possible of course is to shorten the period of time intervening between the measuring of behaviour on the predictor and criterion tests/performances.

Although not as accessible to control as most external factors there are also internal differences which, if taken into account, could be relevant to improved prediction. Nisbet and Illsley (1963) and Nisbet et al (1964) for example have shown that girls who reached puberty earlier than their age peers had slightly but consistently higher intelligence test scores through ages 7 to 13 years, than did later maturing girls. This difference however was smaller at age 16 years, and so the possibility of this maturational rate factor contributing, albeit in a small way, to misclassification is indicated.

The representativeness of the S.C.E. candidate sample used in this study could be questioned on three points.

(i) The boy:girl percentage is approximately 62:38 as against 53:47 in the population of S.C.E. candidates.

(ii) The boys schools:girls schools:co-educational schools ratio is 4:2:4 as against 1:1:5 in the population of state secondary schools.

(iii) All are city schools.

In relation to the first point Dale (1960) cites an unpublished manuscript by Sutherland (1959) as the basis for urging separate prediction for the sexes. Sutherland found a lower correlation coefficient for a combined group than for either sex separately. The present study does not confirm this finding, and in fact found statistically significant sex differences in prediction to be

almost negligible. Thus the correlation coefficients for the combined samples (Appendix D) adequately represent those for the sexes separately. There was however a general trend for correlations to be higher for girls than for boys.

The second point could be used as the basis for criticism of the sample used in this study if there could be shown to be differences in the attainment of boys and girls attending single-sex and co-educational schools and if these differences were not accountable for by IQ. In the United Kingdom Sutherland (1961) found 17 year old segregated girls to have more success than co-educational girls at public examinations with the reverse holding for boys, the differences not being accountable for by IQ. However, Dale (1964) found only slight differences between single-sex and co-educational grammar school pupils attainment in English at the School Certificate Ordinary Level, with attainment differences between schools of different types often being less than differences between schools of the same type. Hence there is no consistent indication that the type of school attended significantly affects later attainment, although data concerning this point have not been reported for New Zealand Schools.

The third point shows the presence of a distinct bias in the sample with only large city schools being included. Obtained correlations could have been affected if the substantial proportion of pupils attending small town and country/rural secondary schools was represented. Evidence which points this way is reported by Armstrong (1964) who found that 11 year old children from rural schools are to some extent handicapped by an assessment of their academic potential compared with urban children if objective, speeded test procedures are

used. He found that 7% of the rural children who would not be selected by the tests would justify a grammar school placement by their later performance, ie. less weight can be given the IQs of such children than those of city children when making classification decisions.

A wide variety of influences could have operated between the two times at which measurements were made in this study. Such influences operate at the individual level, e.g. prolonged illness leading to school absence, personality clash with teacher, family circumstances; at the class/group level, e.g. frequent teacher changes in one subject, a very capable teacher in another subject; and at the school level, e.g. subjects and courses available, timetable arrangements, policy re examination entries. Any one S.C.E. candidate could have been subjected to factors such as these to a greater or less degree than his fellow candidates, and the importance of both adventitious and planned experiences is therefore clearly quite considerable over a three year period in relation to an event such as the S.C.E.

While influences at the individual and class/group levels are largely adventitious (but none-the-less important in their effects) those cited above as possibly operating at the school level are more obviously open to control by school administrators. A combination of subjects ("course") available to pupils at one school may not be available at another; at one school a particular subject may be timetabled so as to allow less effective teaching time than is given at another school; one school may arrange its overall programme so as to encourage first-year fifth form pupils to attempt the S.C.E., while another may regard a

four-year preparation as necessary for most pupils and actively discourage them from sitting the examination earlier.

School policy differences such as these could clearly have substantial effects upon the characteristics of the S.C.E. candidature, and so any general statement about prediction of S.C.E. marks based on a sample of ten schools could conceal differences in prediction between separate schools. This impinges on the question of national versus local norms as discussed by Bloom and Peters (1961) in relation to educational classification and guidance practices. They considered both types of norm to have value, the nature of the norm required or used bearing relevance to the type of decision to be made or guidance offered. In the context of the present study the comparison of correlation coefficients obtained by a school with those for a national group could indicate whether the school should review its classification and/or teaching provisions.

Although the standardized tests administered at the beginning of this study remain constant in item content, the S.C.E. papers vary from year to year. This is so even when syllabuses are unchanged. The particular questions included in any one examination paper are simply an arbitrary sample of all possible questions which could be designed to measure the candidates' grasp of the syllabus. Although the S.C.E questions are devised by experts there is always a proportion of teachers (expert in their own right in this matter) who indicate dissatisfaction with the representativeness of syllabus sampling.

Thus a further problem arises in the evaluation of studies of this type in which the criterion is not of a stable and standardized form. Exact replication

is not possible because of this uncontrolled factor, even if it could be assumed that the effects of other (intervening) variables discussed earlier are relatively stable from year to year within large samples.

The substitution of one standardized factorially "pure" test in each subject in place of the various annual examination papers might satisfy the experimental design requirement of replication (of criterion) but would be open to objection on grounds of coaching and practice effects, as well as defeating the educators' aim of a full and even coverage of a wide syllabus.

Astin (1964) in a discussion of criterion-centred research argues that "criterion" should be used "only with reference to variables or events which are judged to have immediate social relevance or importance" (footnote, p.808). He defines a "conceptual criterion" as "a verbal statement of important or socially relevant outcomes based on the more general purposes or aims of the sponsor" (p.809). A "criterion performance" must be stated in operational terms, and is "any observable event which is judged to be relevant to the conceptual criterion." (p.810), while "criterion measures" are derived from the operationally defined criterion performance. The general aims of the New Zealand education system and the S.C.E. as reviewed in this study fit coherently into Astin's scheme and his discussion provides an appropriate justification of practical, criterion-centred studies.

3. Further Research Indicated.

Closer examination of the construction of the Tasman tests and their

relationships is needed. The possibility of shortening both the TJRT and TVAT by internal consistency techniques should be considered, in the former by reducing the number of very hard items and in the latter by a reduction in easy items (although a "ceiling" effect should be avoided).

There is also a need to examine the weightings which are given the TJRT and TVAT components in arriving at the Tasman Total Scaled Score from which an IQ was to have been derived. This follows from the finding of this study that TJRT correlations were typically much lower than TVAT correlations with S.C.E. marks, yet the TJRT score is twice as heavily weighted in the Total Scaled Score.

In the event of a restandardization of the Tasman the size and complexity of the national new entrant group would suggest that randomness in drawing a sample is probably unattainable simply by testing all new entrants from a number of accessible secondary schools. It could be better, as pointed out by Loevinger (1965), to define and weight relevant characteristics of the population and select a representative sample on this basis, rather than to rely on a sample being representative simply because it is large.

Restandardization of the Otis at both Intermediate and Higher levels is also indicated, particularly in view of the weaknesses attributed to it in Chapter 4.

The lack of an adequate presentation in manual form of finalized administration and scoring procedures, standardization data, age norms, sex differences, reliability and validity data, and suggestions as to how the test

results can be interpreted and used must seriously limit the use which is made of the Tasman tests. Yet this study has shown the Tasman even in its present form to be at least as adequate as the Otis for general scholastic prediction purposes (and hence for "ability streaming") as well as having some potential value in differential prediction (classification). If the limitations could be overcome and parallel forms constructed the Tasman tests could become important additions to the small range of adequate tests suitable for use at the secondary school new entrant level.

In view of the relatively low (TJRT, TVAT) correlation there could be merit in setting up multiple R regression equations for differential prediction purposes, and following such work using one sample with a cross-validation study to establish the "shrinkage" of R using another sample. There could be value in clustering criterion measures with a view to establishing whether prediction is improved. Such clustering of school subjects has been studied factorially, e.g. Lewis (1961), and further work in this direction has been urged by Travers (1964, p.369). Such work could have implications for the composition of secondary school courses and for classification systems.

Although few significant differences in prediction for the sexes separately were found in this study, evidence of differential attainment levels has been reported in the literature and further research with the Tasman using fractionated populations and samples is indicated before it can be confidently asserted that prediction is unaffected by sex. As well there could be an investigation of prediction for high (and low) scoring groups on criterion measures,

since one of Shouksmith's aims in designing the Tasman tests, and the TJRT in particular, was to have a very high ceiling and so spread the scores of the more able pupils to a greater extent than occurs at the upper end of the Otis scale.

A final area in which further research could be of interest would be in the relationship between the Vocabulary section (Part A) of the TVAT and the Total Tasman score (or an independent IQ). Millman and Glock (1965) in a review of trends in mental testing state: "Investigators have found substantial relationships between scores on short vocabulary tests and scores on more general measures of intelligence" (p.21). Miner (1957) summarized twenty-one studies yielding estimates of such relationships and obtained a median value of r = .83. Using two forms of a short (20 item) multiple choice vocabulary test with a selected adult sample Miner (1961) concluded: "In the population as a whole the correlation between the short test and tests of general intelligence appears to be at least .75" (p.159).

Cronbach and Gleser (1965) have presented a mathematical treatment of decision theory in relation to institutional decisions, the purpose being to find the strategy of testing and consequent courses of action which will yield most benefit to the institution. In their scheme, which differs from the traditional psychometric approach adopted in this study, two major elements are the cost of using certain test procedures and the benefits/utility of the various possible courses of action suggested. The purpose of decision theory is to maximize the average gain from decisions made on the basis of testing and to minimize the average loss. Research within such a framework might indicate how the relatively high

proportion of "misses" (pupils with above average test scores who failed S.C.E.) could be reduced.

CHAPTER VII

SUMMARY AND CONCLUSIONS

1. Summary.

During the early 1960's the Tasman tests were developed with a view to providing a wider basis for secondary school classification decisions than was considered possible by using traditional, single-score, verbal tests of intelligence like the Otis. The main purpose of this study was to see whether the new test compared sufficiently favourably with the established Otis as a predictor of long-term scholastic attainment to justify its use in the testing of secondary school new entrants when such test results may be used to assign pupils to various courses of study.

2180 new entrants whose Tasman scores and Otis IQs were known comprised the third form sample of which 1324 pupils became candidates for the School Certificate by sitting the nationwide examination after three years at secondary school.

The results showed the Tasman to compare satisfactorily with the Otis as a predictor of total and single subject marks in the S.C.E. Tasman correlation coefficients ranged from .726 to .383 in the case of general-academic subjects, and from .612 to .269 for the technical-practical ones.

Significant sex differences were negligible, but there was an overall trend for prediction to be better for girls. In general the Otis correlation coefficients obtained in this study were greater than those reported by Clark (1958)

although rarely significantly so. The comparatively low correlation between concurrent TVAT and TJRT scores and the finding of a number of significant or substantial differences between them in predicting marks in single subjects indicated that further investigation of the use of these tests as differential predictors could be worthwhile.

2. Conclusions.

Prediction of the total or aggregated S.C.E. mark by the Tasman Total Scaled Score was at a statistically significant and satisfactory level ($r = .664$) considering the many different subjects and their combinations which were offered by the S.C.E. candidates. Values of r ranged from .726 (EN) to .269 (ES), being typically larger for the general-academic subjects (r range approximately .4 to .6) than for the technical-practical ones (r range approximately .3 to .5).

These general findings are consistent with Wallen's (1962) review statement concerning more recently reported correlations between group tests (usually with a substantial "educational loading") and scholastic content areas: "In general the correlations reported are consistent with previous studies, the range being approximately .30 to .60, with prediction best in the more academic subjects" (p.16).

The importance of the 'v:ed' type of test in predicting general-academic subject marks is clearly shown by this study, the TVAT component correlating more highly than the Tasman total in a number of subjects. In

Geography, for instance, TVAT r = .501 while Tasman r = .475, this observation fitting with Taylor's (1960) factor-analytic finding that "attainment in Geography was highly dependent on verbal-educational ability" (p.270).

The Tasman test compared quite favourably with the Otis in predicting overall and subject achievement levels, and in a number of cases predicted better at a statistically significant level. The TVAT component also compared favourably with the Otis, particularly in the general-academic subjects, but the Otis was usually a much better predictor than the TJRT.

The TVAT was a superior predictor to TJRT in general-academic subjects except for MA, but the TJRT was superior in a number of technical-practical subjects. This has some correspondence to the findings of Tempero and Ivanoff (1960) who found the Verbal score on the standardized test which they used to be a better predictor of English attainment than the Quantitative score, with the reverse holding for attainment in Mathematics and Physics. Blumenfeld (1965) found the Language subtest of a standardized battery to be a superior predictor of overall scholastic attainment than were the Reasoning and Quantitative subtests. This is consistent with the findings of the present study.

The general superiority of TVAT to TJRT may be due in part to the item type of the latter. Campbell (1964) says, of test items of the series type in which a rule or method has to be supplied by the testee, that "success depends to too great an extent on whether S has come across the correct (or closely correlated) rule before, and incorporated it into his 'personal universe' of rule forms" (p.103). It also depends on the point in his repertoire of rules

at which the testee begins his attempt to solve the item. Campbell suggests that scores from such tests are suspect because "a S of superior intelligence may score below average just because he happens to start off in the wrong places" (p.106).

Although very few significant differences in prediction for the sexes considered separately were found there was a trend for r values for girls to be greater than for boys, a trend also observed by Clark (1958). Earlier female physical maturation together with correspondingly earlier stability of motivational and behavioural characteristics could be determining factors in this trend.

Predictors in this study were Otis IQs and Tasman raw scores. An individual's Otis IQ shows how he compares with others of a specified age group while the Tasman raw score shows how he compares with other new entrants without reference to age. Once they enter social institutions, e.g. school, children are usually compared amongst themselves as members of a socially defined membership group, e.g. new entrants at secondary school, rather than as members of particular chronological age (C.A.) groups. At school they are compared with their educationally equivalent peers, i.e. those in the same class, and any decisions are made by comparing the individual pupil with others of his educational level, with C.A. as a secondary consideration (although its relevance may be considerable where an individual's C.A. deviates grossly from the median for his educational group).

The S.C.E., recognized by society generally as an index of educational status, compares the performances of the candidates without reference

to age. It would appear therefore from the results of this investigation that for educational classification and evaluation purposes a pupil's test performance may be more meaningful if related to his current educational status/level alone rather than to C.A. alone. Data obtained for this study did not permit an examination of prediction using Otis raw scores, but future studies in this area could examine this factor in relation to prediction.

This study has been concerned with a practical problem rather than with fundamental considerations. If a defence of such a choice is necessary Astin (1964) has provided a thoughtful discussion of considerable relevance. He evaluates the advantages and disadvantages of the criterion-centred and construct-centred types of validity study and sees them as similar to the extent that both are attempts to account for variance in some variable. In the former this variable is the criterion, and in the latter the test itself.

Astin is of course referring to a stable and operational criterion, and in this respect the S.C.E. papers are neither the same from year to year nor clearly differentiable on the basis of aptitudes. The former limitation impinges upon attempts to compare replication type studies in this area, while the latter will have implications for attempts to design differential test batteries for prediction.

In conclusion, it appears from the findings reported above that the Tasman tests may be used in secondary schools with a modest degree of confidence and a great deal of caution. While Tasman has some merit in its present form basic improvements are necessary and refinements desirable.

Representative standardization data and a published manual are urgently needed. Further research could be valuable with a view to establishing multiple regression equations, providing validity data for selected groups, e.g. testees who score at high levels on Tasman and criteria measures, and examining the test from a decision-theory viewpoint rather than from a psychometric one.

The Tasman is a test designed for a particular purpose: aiding correct classification of new entrants in New Zealand secondary schools. While this study clearly shows its main strength to be that of predicting overall rather than differential scholastic aptitude, there are signs that it may have some potential for development as a classificatory tool. For the purpose of general scholastic prediction it appears to be at least as adequate as the Otis.

BIBLIOGRAPHY

ADAMS, G. S. Measurement and Evaluation: in Education, Psychology, and Guidance. New York: Holt, Rinehart and Winston. 1964.

ANASTASI, A. Psychological Testing. New York: Macmillan. 1961.

ARMSTRONG, H. G. "A Comparison of Rural Schools' Performance in a Test-Controlled and Test-Free Selection Procedure at Eleven." Br. J. Ed. Psych. 1964, 34, 237 - 46.

ARVIDSON, G. L. The Hutt Valley Form II Classification Scheme, 1958. Unpub. Cyclostyled Notes: Victoria University College. 1959.

ASTIN, A. W. "Criterion-Centred Research." Educ. psychol. Measmt., 1964, 24, 807 - 22.

BLOOM, B. S., and PETERS, F. R. The Use of Academic Prediction Scales. New York: Free Press of Glencoe. 1961.

BLUMENFELD, W. S. "Predicting Grade Point Average with the S.R.A. Tests of Educational Ability." Educ. psychol. Measmt., 1965, 25, 555 - 6.

BOAG, P. W. Classification of New Entrants to Secondary Schools. Unpub. Dip. Ed. Thesis: Univ., of Auckland. 1960.

BUROS, O. K. The Fifth Mental Measurements Yearbook. New Jersey: Grypon Press. 1959.

CAMPBELL, A. C. "Intelligence Test Items and Serendipity: Further Considerations." Aust. J. Psych., 1964, 16, 102 - 6.

CLARK, G. H. A Statistical Investigation of the Predictive Value of the Subjects of a Post-Primary School Entrance Examination. Unpub. Dip. Ed. Thesis: Univ. of Canterbury. 1958.

CRONBACH, L. J. and GLESER, G. C. Psychological Tests and Personnel Decisions. University of Illinois Press, Urbana. 1965.

COMMISSION ON EDUCATION: Report. Wellington: Department of Education. 1962.

DALE, R. R. "Reflections on Research on Allocation to Secondary Education." Br. J. Ed. Psych., 1960, 30, 220 - 4.

DALE, R. R. "Research on Comparative Attainment in English in Single-Sex and Co-Educational Grammar Schools." Educ. Res., 1964, 6, 170 - 8.

EYSENCK, H. J. "Intelligence Assessment: A Theoretical and Experimental Approach." Br. J. Ed. Psych., 1967, 37, 81 - 98.

GUILFORD, J. P. "New Standards for Test Evaluation." Educ. psychol. Measmt., 1946, 6, 427 - 38.

GUILFORD, J. P. "The Three Faces of Intellect." Amer. Psychologist: 1959, 14, 469 - 79.

KETTNER, N. W., GUILFORD, J. P., and CHRISTENSEN, P. R. "A Factor-Analytic Investigation of the Factor Called General Reasoning." Educ. psychol. Measmt., 1956, 16, 438 - 53.

KETTNER, N. W., GUILFORD, J. P., and CHRISTENSEN, P. R. "A Factor-Analytic Investigation Across the Domains of Reasoning, Creativity and Evaluation." Psych. Monographs, 73, No.9, 1959.

LEWIS, D. G. "Group Factors of Attainment in Grammar School Subjects." Br. J. Ed. Psych., 1961, 31, 241 - 8.

LOEVINGER, J. "Persons and Populations as Psychometric Concepts." Psych. Rev., 1965, 72, 143 - 55.

McGARVEY, R. D. A Test Battery for Third Form New Entrants. Unpub. Dip. Ed. Thesis: Univ. of Auckland. 1962.

McNEMAR, Q. Psychological Statistics. New York: Wiley. 1955.

McNEMAR, Q. "Lost: Our Intelligence? Why?" Amer. Psychologist, 1964, 19, 871 - 882.

MANN, L. and PHILLIPS, W. A. "Fractional Practices in Special Education: A Critique." Exceptional Children, 1967, 33, 311 - 317.

MILLMAN, J. and GLOCK, M. D. "Trends in the Measurement of General Mental Ability." Rev. Educ. Res., 1965, 35, 17 - 34.

MINER, J. B. Intelligence in the United States. New York: Springer. 1957.

MINER, J. B. "On the Use of a Short Vocabulary Test to Measure General Intelligence." J. Ed. Psych., 1961, 52, 157 - 60.

NISBET, J. and BUCHAN, J. "The Long-Term Follow-Up of Assessment at Age Eleven." Br. J. Ed. Psych., 1959, 29, 1 - 8.

NISBET, J. D. and ILLSLEY, R. "The Influence of Early Puberty on Test Performance at Age Eleven." Br. J. Ed. Psych., 1963, 33, 169 - 76.

NISBET, J. D., ILLSLEY, R., SUTHERLAND, A. E. and DOUSE, M. J. "Puberty and Test Performance: a Further Report." Br. J. Ed. Psych., 1964, 34, 202 - 3.

REDMOND, M. and DAVIES, F.R.J. The Standardization of Two Intelligence Tests. Wellington: New Zealand Council for Educational Research. 1940.

RISKE, M. "The Relationship between I.Q. and Examination Success." Education, 1948, 1, 67 - 73.

ROGERS, C. A. Measuring Intelligence in New Zealand. Auckland: Auckland University College. 1956.

SCHOOL CERTIFICATE REVIEW COMMITTEE: Report. Wellington: Department of Education. 1960.

SHOUKSMITH, G. The Canterbury Junior Reasoning Test: an Interim Manual. Unpub. Cyclostyled Notes: University of Canterbury. 1962.

SHOUKSMITH, G. "Aptitude Tests in New Zealand Schools." National Education, 1963(a), 45, 503 - 5.

SHOUKSMITH, G. Tentative I.Q. Tables. Unpub. Cyclostyled Notes: University of Canterbury. 1963(b).

SHOUKSMITH, G. "Testing the Tests." National Education, 1964, 46, 174 - 9.

SPEARMAN, C. E. The Abilities of Man. London: Macmillan. 1927.

SUTHERLAND, M. B. "Co-Education and School Attainment." Br. J. Ed. Psych., 1961, 31, 158 - 69.

TAYLOR, C. C. "A Study of the Nature of Spatial Ability and Its Relationship to Attainment in Geography." Br. J. Ed. Psych., 1960, 30, 266 - 70.

TEMPERO, H. E. and IVANOFF, J. M. "The S.C.A.T. as a Predictor of Achievement in Selected High School Subjects." Educ. psychol. Measmt., 1960, 20, 835 - 38.

THOMAS COMMITTEE: Report. Wellington: Department of Education. 1943.

THURSTONE, L. L. The Vectors of the Mind. Chicago: University of Chicago Press. 1935.

THURSTONE, L. L. "Primary Mental Abilities." Psychometr. Monogr., 1, 1938.

TRAVERS, R.M.W. An Introduction to Educational Research. New York: Macmillan. 1964.

VERNON, P. E. Secondary School Selection. London: Methuen. 1957.

VERNON, P. E. "A New Look at Intelligence Testing." Educ. Res., 1958, 1, 3 - 11.

VERNON, P. E. The Structure of Human Abilities. London: Methuen. 1961.

WALLEN, N. E. "Development and Application of Tests of General Mental Ability." Rev. Educ. Res., 1962, 32, 15 - 24.

WALLIS HOUSE COMMITTEE: Report. Wellington: Department of Education. 1956.

WRIGHT, W. E. Missing Data Correlation Program for Symmetric and Asymmetric Matrices for the I.B.M. 1620. Southern Illinois University Data Processing and Computing Centre. 1962.

YATES, A. and PIDGEON, D. A. Admission to Grammar Schools. London: Newnes. 1957.

APPENDICES

APPENDIX A

Symbols and Abbreviations

Predictor Tests

TJRT	Tasman Junior Reasoning Test
TVAT	Tasman Verbal Ability Test
TVAT A	Part A of the TVAT
TVAT B	Part B of the TVAT

School Certificate Examination

S.C.E.	School Certificate Examination
EN	English
GE	Geography
MA	Mathematics
GS	General Science
HI	History
FR	French
BI	Biology
CP	Commercial Practice
BK	Book-keeping
CH	Chemistry
PH	Physics
DD	Drawing and Design
TD	Technical Drawing
ES	Engineering Shopwork
WW	Woodwork
ST	Shorthand Typing
CL	Clothing
HC	Homecraft

Other

IQ	Intelligence Quotient
N	Sample size
M	Sample mean
S.D.	Standard Deviation
r	Pearson product moment correlation coefficient
'g'	General Intelligence factor
's'	Specific factor
'v:ed'	Verbal-educational group factor
'k:m'	Practical-mechanical group factor
'v'	Verbal subfactor of 'v:ed'
'n'	Number subfactor of 'v:ed'
P.M.A.	Primary Mental Abilities
D.A.T.	Differential Aptitude Tests

APPENDIX B

THE CANTERBURY JUNIOR REASONING TEST

PREPARED BY G. SHOUKSMITH

Name ..... Today's date .....

Age .....years. .....mths. Date of birth .....

.....

.....

Instructions

There are fifty questions in this test and for each one you will see a dot or a number of dots at the right hand side. Wherever there is a dot, that means that there is a letter or a figure missing. Your job is to write the correct letters, words or numbers just over the dots, in order to answer the question. Here is an example to show how it is done:—

Ex. 1 A B C D E F G

In this example the letters on the left are in alphabetical order. The next two letters in the alphabet after E, are 'F' and 'G'. Therefore 'F' and 'G' have been filled in above the dots on the right.

Here are two more examples:

Ex. 2. 1 3 5 7 9 11

Here the rule is to add 2 to each number to get the next.

Ex. 3. SMALL BIG UP DOWN IN OUT

Each second word means the opposite of the one just before it.

Now try this one for yourself:

Ex. 4. ABC DEF GHI . . .

Note that the number of dots on the right corresponds to the number of missing pieces.

Work as quickly as you can. You are not expected to answer every question but if you do finish before time is called, go back and check your work.

Are there any questions?

DO NOT TURN OVER UNTIL YOU ARE TOLD
YOU WILL HAVE 25 MINUTES FOR THIS TEST.

CANTERBURY JUNIOR REASONING TEST

Question No.		
1	20 19 18 17	. . . .
2	P Q R S	. .
3	AAT BAT	. . .
4	14 17 20	. .
5	ON OFF IN	. . .
6	2 3 5 8 12	. .
7	Z X V	.
8	PAT SAT	. . .
9	FATE ATE	. .
10	DAY DEY DIY	. . .
11	G K O	.
12	A Z B Y	. .
13	ANIN ENON	. . . . .
14	3 2 6 4 9	. . . .
15	100 1 80 2 60 4	. . . .
16	10 A 20 C 15	.
17	ABC BDE CFG	. . .
18	M N A B O	. .
19	BUSY BUS	. .
20	1d 3d 9d	. .
21	1 3 3 9 9 18	. .
22	49 36 25	. .
23	2301 3522 4743	. . . . .
24	3 1 5 1 7 3 9 3	. .
25	A C B D C	. .

GO ON TO THE NEXT PAGE

CANTERBURY JUNIOR REASONING TEST

Question No.
No.

26	ABC BEG CHK	. . .
27	0.2 0.4 0.6	. .
28	FINISH END BEGINNING ULTIMATE LAST	. . . . .
29	6″ 1′ 6″ 4′ 6″	. . .
30	N Q O R P	. .
31	ACE BEH CGK	. . .
32	2 3 5 8	. .
33	ABY BCZ CDY	. . .
34	1 $\frac{1}{4}$ $1\frac{3}{4}$ $-\frac{1}{2}$	. .
35	0.3 0.6 0.9	. .
36	$\frac{1}{2}$ 1 3 12	. .
37	CHARTER ART MISS	. .
38	5 TENDE 3 END 1	.
39	SOFT SOOT	. . . .
40	12345678 DAUGHTER 18 'x' 4 DRAG 6 'x' 8	. . .
41	6-1 BORDER 6-2 REDROB 3-1 ORD 3-2	. . .
42	1 6 30 120	. . .
43	20 18 21 16 22	. . . .
44	GERALD 6243 DEAR 5246	. . . .
45	10 100 9 81	. . .
46	1 1 2 6	. .
47	2 + 1 5 4 + 3	. . . .
48	A25 C16 E	.
49	A3 C6 F	. .
50	A 1 B 1 C 3 F 5	. .

APPENDIX C

CANTERBURY VERBAL ABILITY TEST

PREPARED BY G. SHOUKSMITH

Name ..... Today's date.....

Age .....yrs.....mths. Date of birth.....

.....

.....

PART A — VOCABULARY

Instructions

This test is designed to find out how many words you know. On the left-hand side of the page, alongside each question number, you will see a word printed in capital letters. To the right are five other words printed ordinarily and lettered A, B, C, D and E. Your job is to choose *one* of the words A, B, C, D or E which means the same as, or most nearly the same as the word in capital letters. When you have picked the word which you think means the same as the word in capitals, place that word's letter in the space provided at the right.

Look at Example 1, which shows how this is done:

Eg. 1 BATTLE A Afraid B Machine C Fight D History E Bar **C**

"Fight" means the same as "Battle" so its letter, the letter C, has been placed in the space on the right.

Here is another example:

Eg. 2 ABROAD A Beyond B Overseas C Author D Deal E Outline **B**

The one word which means most nearly the same as "Abroad" is the word "Overseas" so its letter, letter B, has been written in the space on the right.

Now try this one:

Eg. 3 QUICK A Bright B Cut C Thick D Fast E Tread .....

Remember—from the five words on the right, you pick the one which means the same as, or most nearly the same as the one on the left in capitals. Then write its letter, *not* the word itself, in the space on the right.

Are there any questions?

YOU WILL HAVE 10 MINUTES FOR THIS TEST.

DO NOT TURN OVER TILL YOU ARE TOLD.

1	TREMENDOUS	A Huge	B Volume	C Pretend	D Bargain	E Hasten	.....
2	WORTH	A Useless	B Volume	C Appear	D Value	E Wisdom	.....
3	SCHEME	A Verse	B Astonish	C Humour	D Mercy	E Plan	.....
4	DYE	A Chemical	B Colouring	C Yonder	D Sacred	E Gasp	.....
5	JACKET	A Blaze	B Coat	C Carriage	D Terror	E Disguise	.....
6	ABSOLUTE	A Spread	B Beggar	C Certain	D Measure	E Observer	.....
7	BRILLIANT	A Collection	B Twinkle	C Shining	D Fellow	E Bushell	.....
8	GOVERN	A Warrior	B Guide	C Inquire	D Rule	E Bloom	.....
9	CHASE	A Sharply	B Voyage	C Appetite	D Pursue	E Lame	.....
10	GRIEF	A Alas	B Sorrow	C Pleasant	D Jealous	E Grateful	.....
11	MODIFY	A Alter	B Bruise	C Minority	D Yarn	E Hazard	.....
12	BOUNDLESS	A Adjacent	B Combustion	C Unlimited	D Embody	E Pyramid	.....
13	CONCEAL	A Advise	B Hide	C Repair	D Trap	E Seal	.....
14	YEARNING	A Flaming	B Tidings	C Longing	D Kindred	E Moulding	.....
15	UPROAR	A Testify	B Wondrous	C Phase	D Tumult	E Tease	.....
16	DISMAL	A Sample	B Embarrass	C Subdue	D Weariness	E Dreary	.....
17	INTERPRET	A Allege	B Translate	C Ignore	D Royalty	E Sermon	.....
18	FONDLY	A Forenoon	B Solemnly	C Shanty	D Tenderly	E Convincing	.....
19	BEWITCH	A Acquit	B Enchant	C Demon	D Fabulous	E Harry	.....
20	JOSTLE	A Consecrate	B Hustle	C Strenuous	D Lapse	E Deplore	.....
21	WRANGLE	A Schism	B Cede	C Brawl	D Artichoke	E Caisson	.....
22	KNIGHTLY	A Huntsman	B Chivalrous	C Composed	D Clearly	E Gorge	.....
23	AERIAL	A Beguile	B Ethereal	C Wield	D Groove	E Junction	.....
24	ERE	A Ago	B Before	C Vine	D Poison	E Keep	.....
25	COUNTERMAND	A Superscribe	B Revoke	C Amortize	D Lacerate	E Embassage	.....
26	VERNACULAR	A Aborigines	B Kirtle	C Mandatory	D Dialect	E Axiom	.....
27	CHISEL	A Chaos	B Defraud	C Mortar	D Glint	E Nook	.....
28	ZEALOT	A Itinerant	B Ocher	C Linnet	D Junket	E Fanatic	.....
29	SOLICITOUS	A Stipulate	B Hazardous	C Desirous	D Tempestuous	E Heretical	.....
30	CELESTIAL	A Idol	B Temporary	C Ultimate	D Lasting	E Heavenly	.....

31	FETTER	A Impede	B Layman	C Keenly	D Deceit	E Data	.....
32	VENERATE	A Leaven	B Revere	C Blatant	D Cilia	E Override	.....
33	APPEASE	A Pacify	B Ordinance	C Incur	D Believer	E Loathe	.....
34	GRANDEUR	A Explosive	B Plebian	C Eminence	D Brutal	E Manoeuvre	.....
35	FOREORDAIN	A Freeholder	B Unburden	C Predetermine	D Thraldom	E Equivocate	.....
36	REFULGENT	A Scintillating	B Pestilent	C Bubonic	D Vibratory	E Deflate	.....
37	NEMESIS	A Adherent	B Zephyr	C Equilibrium	D Retribution	E Fiduciary	.....
38	PRESCIENT	A Galvanic	B Foresight	C Malted	D Excrete	E Allocate	.....
39	AMBIENT	A Datine	B Umbrage	C Sophistry	D Phylactery	E Circumfused	.....
40	TRUISM	A Platitude	B Baseness	C Kine	D Redoubt	E Pittance	.....

STOP put your pen or pencil down

PART B — ANALOGIES

Instructions

In this test, the questions are in the form of a sentence in which the last word is missing. Your job is to complete the sentence by filling in the last word. Look at the following example.

Ex. 1 Cat is to Kitten as Cow is to ..... **CALF**

The word needed to complete this sentence is Calf, since the word calf has the same relationship to "cow" as "kitten" has to "cat", that is, kitten is a young cat; a calf is a young cow.

Here is another example:

Ex. 2 Food is to Hunger as Drink is to ..... **THIRST**

Again a word has to be found which has the same relationship to "drink" as the word "hunger" has to "food". When you are hungry you take food; when you are *thirsty* you take drink. Therefore the complete sentence reads: "Food is to Hunger as Drink is to Thirst".

Now do this one for yourself:

Ex. 3 Hot is to Cold as Black is to .....

Are there any questions?

YOU WILL HAVE 5 MINUTES FOR THIS TEST

DO NOT TURN OVER UNTIL YOU ARE TOLD.

Question No.

(1) Big is to Bigger as Small is to .....

(2) Green is to Go as Red is to .....

(3) Summer is to Winter as Spring is to .....

(4) Arm is to Hand as Leg is to .....

(5) Cup is to Saucer as Knife is to .....

(6) The Wind is to The Willows as Alice is to .....

(7) Jason is to Fleece as Noah is to .....

(8) Outside is to Exterior as Inside is to .....

(9) First is to Last as Beginning is to .....

(10) Golf is to Club as Cricket is to .....

(11) Pianist is to Piano as Cellist is to .....

(12) Sunday is to Tomorrow as Friday is to .....

(13) Good is to Best as Bad is to .....

(14) Cook is to New Zealand as Columbus is to .....

(15) Duck is to Drake as Mare is to .....

(16) Hamlet is to Shakespeare as Copperfield is to .....

(17) Wheels are to Car as Wings are to .....

(18) The Union Jack is to the U.K. as the Stars and Stripes is to .....

(19) London is to Britain as Athens is to .....

(20) Dawn is to Night as Dusk is to .....

(21) North Pole is to Nansen as South Pole is to .....

(22) Joy is to Laughter as Sorrow is to .....

(23) Emperor is to Empress as Earl is to .....

(24) Space is to Spatial as Tempo is to .....

(25) H_2O is to Water as H_2SO_4 is to .....

APPENDIX D

Validity Coefficients: Combined Samples

S.C.E. Mark	Predictor Test						Sample size		
	TJRT	TVAT A	TVAT B	TVAT	Tasman	Otis	N	Boys	Girls
Total	.539	.583	.582	.642	.664	.611	1324	819	505
EN	.545	.725	.661	.771	.726	.704	1324	819	505
GE	.347	.429	.483	.501	.475	.444	642	402	240
MA	.555	.459	.484	.525	.628	.542	581	445	136
GS	.482	.473	.470	.526	.574	.488	456	307	149
HI	.393	.477	.453	.527	.507	.441	484	285	199
FR	.379	.424	.453	.511	.497	.537	312	132	180
BI	.323	.400	.453	.465	.450	.417	217	55	162
CP	.397	.416	.369	.442	.488	.393	202	93	109
BK	.431	.334	.163	.318	.466	.374	188	111	77
CH	.369	.386	.375	.427	.463	.403	165	156	9
PH	.304	.284	.323	.341	.383	.472	148	140	8
DD	.398	.474	.379	.472	.496	.441	98	55	43
TD	.458	.148	.376	.267	.454	.303	124	124	-
ES	.181	.155	.305	.261	.269	.326	66	66	-
WW	.294	.115	.381	.246	.356	.274	37	37	-
ST	.538	.553	.446	.577	.612	.488	60	-	60
CL	.375	.281	.229	.282	.408	.106	41	-	41
HC	.330	.103	.284	.187	.305	.367	36	-	36

APPENDIX E

Validity Coefficients: Boys and Girls

Predictor / Test	S.C. Total		S.C. English		S.C. Geography		S.C. Maths		S.C. Gen. Science		S.C. History	
	819 Boys	505 Girls	819 Boys	505 Girls	402 Boys	240 Girls	445 Boys	136 Girls	307 Boys	149 Girls	285 Boys	199 Girls
TJRT	.525	.562	.530	.570	.300	.423	.562	.579	.469	.459	.361	.426
TVAT A	.561	.624	.737	.714	.395	.475	.442	.552	.418	.537	.469	.481
TVAT B	.575	.604	.684	.656	.466	.477	.475	.525	.430	.517	.463	.451
TVAT	.625	.678	.790	.760	.472	.526	.508	.605	.475	.588	.524	.531
Tasman	.650	.687	.729	.728	.425	.544	.627	.679	.549	.576	.486	.528
Otis IQ	.595	.638	.695	.722	.423	.479	.543	.585	.453	.518	.402	.484

Predictor / Test	S.C. French		S.C. Biology		S.C. Comm. Pract.		S.C. Book-keep.		S.C. Drawing & Design	
	132 Boys	180 Girls	55 Boys	162 Girls	93 Boys	109 Girls	111 Boys	77 Girls	55 Boys	43 Girls
TJRT	.386	.374	.171	.372	.277	.458	.430	.435	.424	.367
TVAT A	.403	.459	.283	.455	.410	.456	.219	.474	.449	.508
TVAT B	.452	.464	.335	.495	.251	.527*	.013	.343*	.349	.434
TVAT	.487	.557	.353	.514	.396	.527	.181	.476*	.447	.507
Tasman	.496	.501	.273	.508	.383	.562	.415	.529	.505	.481
Otis IQ	.496	.565	.335	.444	.373	.396	.285	.500	.415	.469

* Significantly greater than for the other sex at 5% level of confidence (estimated using Fisher's z transformation).