THE I-D SCHOLASTIC TEST BATTERY AS A PREDICTOR OF GRADES IN ACADEMIC SUBJECTS IN THE WEST AFRICAN SCHOOL CERTIFICATE

BY

ARTHUR W. M. PARKE

A THESIS
SUBMITTED IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF EDUCATION

IN

GUIDANCE AND COUNSELLING

DEPARTMENT OF EDUCATION
AHMADU BELLO UNIVERSITY
ZARIA

1977

ABSTRACT

Eleven of the tests of the I-D Aptitude Test Battery are administered annually to second year pupils in the secondary schools of Kaduna State. They are used primarily to indicate aptitude for technical and commercial courses but could also be used to predict success in academic subjects.

This study investigates the validity of the four Scholastic Aptitude Tests using the West African School Certificate results as the criterion, and shows the efficiency of the Verbal Analogies Test and the Reading Comprehension Test as predictors of grades in the examinations taken three years later. For the sample of 150 of the pupils in Zaria schools who were tested in 1973 and who took the WASC examinations in 1976, these two tests gave a correlation coefficient of 0.69. This confirms previous research which shows the importance of the Intelligence and English Tests for success in secondary education, but leaves in doubt the role of a test of Arithmetic which has normally been considered an essential part of any battery predicting success in academic subjects.

APPROVAL PAGE

DEPARTMENT OF EDUCATION
AHMADU BELLO UNIVERSITY
ZARIA

The undersigned certify that they have read, and recommend to the Faculty Board of Education for acceptance, a thesis entitled The I-D Scholastic Test Battery as a Predictor of Grades in Academic Subjects in the West African School Certificate submitted by Arthur W. M. Parke in partial fulfilment of the requirements for the degree of Master of Education.

_____
Supervisor

_____

_____

_____

_____
Date

_____
External Examiner

DEDICATION

To my students past and present who stimulated my interest in Guidance and Counselling and provided the motivation to carry out this study.

ACKNOWLEDGEMENTS

It has been a satisfying experience to receive assistance and advice from many quarters, too numerous to mention.

My gratitude must go, in the first place, to the Ministry of Education, Kaduna State and to the West African Examinations Council for their cooperation and encouragement, and also to the staff of the New University of Ulster, in particular, Miss E. M. Tyrrell of the Library and her colleagues, Dr. D. James of the Department of Psychology and Mr. A. C. Todd of the Computer Centre who gave me so much of their time and full access to the facilities of the University during my period of study leave.

Above all I am grateful to Dr. E. W. Graber of the Department of Education (Guidance and Counselling) A.B.U., Zaria, without whose exhortation and practical help, I should never have completed this study. Finally I should like to thank my students for putting up with all the inconvenience of my study programme, and Miss H. McCollum for the quality of her typing and presentation of this thesis.

A. W. M. Parke

TABLE OF CONTENTS

	Page
CHAPTER I : THE PROBLEM	1
1.1 Introduction	1
1.2 Statement of the problem	2
1.3 Significance of the study	3
1.4 Assumptions	4
1.5 Hypotheses	5
1.6 Scope of the study	6
1.7 Definitions	8
CHAPTER II : REVIEW OF RELATED LITERATURE ..	11
2.1 Prediction studies in the U.K. ..	11
2.2 Prediction studies in the U.S.A. ..	24
2.3 Prediction studies in Africa ..	26
2.4 Summary	39
CHAPTER III : DESIGN AND PROCEDURE	41
3.1 The sample	41
3.2 Collection of data	44
3.3 The independent variables	45
3.4 The criterion variables	46
3.5 Processing of data	48
3.6 Hypothesis testing	50
CHAPTER IV : RESULTS AND ANALYSIS	52
4.1 Hypothesis 1	52
4.2 Hypothesis 2	61

4.3 Hypothesis 3 72
4.4 Hypothesis 4 74
4.5 Hypothesis 5 79

CHAPTER V : SUMMARY AND CONCLUSIONS 83

5.1 Summary 83
5.2 Generalisation 86
5.3 Recommendations 89
5.4 Suggestions for further study .. 92

REFERENCES 94

		Page
Table	2.1	16
	2.2	27
	4.1	54
	4.2	55
	4.3	57
	4.4	60
	4.5	64
	4.6	66
	4.7	68
	4.8	70
	4.9	73
	4.10	74
	4.11	80
	4.12	81
Figure	4.1	59
	4.2	62
	4.3	63
	4.4	76
	4.5	75

Chapter 1

THE PROBLEM

1.1 Introduction

Since 1973 Kaduna State has been administering a battery of I-D Aptitude Tests to second year pupils in its secondary schools. The main purpose behind the use of these tests has been to facilitate the guidance of pupils into different third year options which include commercial and technical courses as well as the purely academic. As the guidance programme takes into account the pupils' personal preferences and those of the parents, and as places in these specialised courses are very limited, the usefulness of the tests is being questioned. The expenditure of time and money is considerable, and with changes of senior personnel in the Ministry of Education the original enthusiasm for the programme could be lost.

The "use of tests in counselling and guidance eliminates a great deal of subjective judgment on an individual's ability" (Durojaiye 1972:81). The all-

important factor is therefore their predictive validity. In addition to highlighting a particular bent for commercial or technical courses, the Tests could possibly be used to give an objective assessment of a pupil's suitability for an academic course leading to the West African School Certificate three years later.

1.2 Statement of the Problem

The investigation had as its objective to confirm that there is a correlation between the Aptitude Test scores and the West African School Certificate (WASC) results, and to suggest a minimum score for ultimate success. In addition, it was hoped to establish a correlation between certain tests and results in particular subjects or groups of subjects.

The questions asked were:

(a) Can the I-D Scholastic Aptitude Test Battery as used in Kaduna State predict grades in WASC taken three years later?

(b) Can a minimum score in one or a group of tests be established below which no candidate has more than a 5% chance of passing?

(c) Can certain test scores be used to guide students into taking certain academic subjects?

1.3 Significance of the Study

If it could be shown that the Aptitude Test scores do predict grades in either some subjects or the overall results of the West African School Certificate Examinations, then the Guidance Teacher would be able to help his students towards realising their potential and their limitations, the school administration would have a forecast of results to be expected in the School Certificate, and the Ministry of Education could take advantage of the Test scores to eliminate pupils who have no chance of success.

Students could be saved considerable personal frustration, over-crowding in classes could be reduced, and irregularities in the Common Entrance procedures could be counteracted. Twenty-five years ago McClelland (1942:96), drawing attention to the unsuitability of secondary grammar school education for many children even though they all had their individual educational claims, made the following comments which seem even

more relevant in our own day:

> The admission of pupils with no chance of success is harmful to the pupils themselves, and imposes a heavy burden upon the teachers, but there is [another] aspect which is sometimes overlooked. In these democratic days we think so much of the claims of the average and dull that we run the risk of forgetting those of the intellectual élite who will be the future torchbearers in cultural, social and industrial advance; and it is partly in the interests of this élite that improvement in our methods of selection and guidance is necessary.

The failure of students suited to secondary grammar school education to reach expected standards could be a means of drawing the attention of the Ministry of Education to the possible inadequacy of the facilities and instruction provided. In view of the foregoing, the continued administration of the Tests could then be justified.

Finally, it was hoped to demonstrate the application of simple statistical techniques in carrying out investigations within our educational system and point to areas requiring further research.

1.4 Assumptions

(a) No account need be taken for the possible sources of variance in test and examination

scores (outlined by Thorndike, 1949:73)

(b) Stanine Scores as used in the Tests can be regarded as on an interval scale for statistical purposes and parametric statistic techniques can be used.

1.5 Hypotheses

In the original validity studies on the I-D Scholastic Test Battery, Schwarz (1964b) used concurrent achievement ratings based on terminal grades and recorded a co-efficient of 0.56. As the most important criterion of achievement in secondary education is the West African School Certificate Examinations, it would not be surprising if a relationship should exist between the Test scores and the WASC grades. The first hypothesis was therefore:

(a) There is a significant relationship between scores in the I-D Scholastic Aptitude Test Battery and aggregate scores in the six best academic subjects in the West African School Certificate.

The Scholastic Aptitude Test Battery includes two tests which require a certain command of English and one which is arithmetic. Therefore it would be reasonable to anticipate a relationship between the tests and achievement in English

Language and Mathematics. Hence the second hypothesis:

(b) There is a significant relationship between the scores in the Scholastic Aptitude Test Battery and WASC grades in English Language and Mathematics.

Certain optional subjects taken by candidates in WASC are considered 'easy', e.g. Religious Knowledge, Nigerian Languages, Fine Art, and so it can be assumed that students put more effort into the core subjects reflecting a basic education and taken by every candidate. It would therefore be possible that a more significant relationship could exist between the tests and the core subjects than with an aggregate score including some of the optional subjects. This was tested by the third hypothesis:

(c) Grades in the four subjects taken by all students in WASC - English Language, Mathematics, Geography and Biology - may be predicted from the I-D Scholastic Test Battery with more accuracy than the aggregate of the six best subjects.

Though the tests are administered to pupils already selected for secondary education, it is evident that many do not reach the anticipated standard of achievement. The investigator was interested in testing whether a cut-off point in the Aptitude Test scores could be ascertained so

that pupils with virtually no chance of success could be eliminated:

(d) A cut-off point in the Aptitude Test scores can be fixed below which a pupil will only have a 5% chance of success in WASC.

The final hypothesis arose from the experience of the researcher who has been confronted by students requesting permission to change from Science subjects to Arts subjects in order to enhance their grades in WASC. If the Aptitude test scores could be used to control for differences in ability, it might be feasible to test for a significant difference in grades obtained by Science Students and those obtained by Arts students:

(e) Students who take Science subjects in WASC score higher grades than students who take Arts subjects when the effect of differences in Aptitude scores has been controlled.

1.6 Scope of the Study

The Kaduna State Aptitude Testing programme was begun in 1973, and the first set of pupils tested completed their secondary education and sat for the WASC examinations in 1976.

This prediction study is of necessity limited to that

one set. It was, however, not feasible to consider the entire set, because some failed to complete their courses, and many were transferred to other schools as part of the programme of placement into technical and commercial courses or by their own arrangement. The implications of these limitations are discussed by Cronbach (1961:115) and Schwarz and Krug (1972:179-80) and will be mentioned when interpreting the results.

It was therefore necessary to take a sample of those who remained in the same schools and took academic courses as distinct from courses with a technical or commercial bias. Ndahi (1975) has already considered the success of the placement scheme so this study is limited to the extent with which WASC results in general subjects can be predicted within a sample and how far inferences about the entire population can be made. For a really accurate prediction study each institution should be taken separately and expectancy tables applied to that school alone. This point is made clear by Cronbach (1961:119):

> In most predictive uses of tests the published validity co-efficient is no more than a hint as to whether the test is relevant to the tester's decision. He must validate the test in his own school, and even then he can expect the co-efficient to fluctuate. For this reason testers are usually forced back upon a psychological rather than a purely statistical use of scores.

Nevertheless it was preferred to give a general overview of the situation in the State, and it is hoped that such a study will stimulate others to devise tables applicable to their own situation. Other factors which may affect the pupils' progress and results in WASC mentioned in the next chapter were not taken into consideration although they could contribute to a more accurate prediction because they vary from individual to individual and school to school. The study was restricted to showing associations between scores and it should be kept in mind that the existence of a correlation between two variables must not be assumed to prove causality (Wiseman 1966:vi).

1.7 Definitions

I-D Aptitude Tests are the titles given to a set of 21 ability tests developed in Nigeria by the American Institutes for Research (I-D = Internationally Developed) between 1960 and 1963 and administered since 1966 by the Test Development and Research Division, a part of the West African Examinations Council. Eleven of these tests have been selected for the battery used in Kaduna State, and these scores are grouped variously to give indications of Scholastic, Technical and Commercial aptitudes. Scores are standardised

according to the Stanine system, and the Tests are administered to all Form 2 pupils in the State's secondary schools.

Aptitude is used indiscriminantly in both the British and American senses. As Ord (1972:141) explains:

> Aptitude in the American sense covers a broad range of cognitive abilities measured in such batteries, which include tests of intelligence, attainment and skills as well as aptitudes in the narrower sense in which the British use the term.

West African School Certificate (WASC) is used either to refer to the certificate granted by the West African Examinations Council as a result of the examinations taken at the end of a five year course of secondary education in at least six subjects, or to refer to the examinations themselves. Candidates normally take eight subjects including English and Mathematics, but only their best six subjects count towards their aggregate score. Scores in individual subjects are not standardised and the system of awarding grades seems to vary from subject to subject and from year to year. (Ohuche & Akeju 1975:106).

Academic subjects are considered in this study to comprise the normal range of Arts and Science subjects taken in secondary schools in Kaduna State. Students who took two or more Technical or Commercial subjects were deemed to have chosen the Technical or Commercial options and were

excluded even though their courses were still basically academic.

Predictor: instrument permitting estimation of the scores that can be expected by each candidate in an examination yet to be taken.

Chapter 2

REVIEW OF RELATED LITERATURE

This review of literature is intended to place the present study in the context of the wider field of prediction studies in education which have to a considerable extent been concerned with the processes of selection - in the U.K. for secondary education and in the U.S.A. for College courses. Other factors contributing to accuracy of prediction, though outside the scope of this study, will be mentioned and an outline of the development of aptitude testing in Africa will help to show the contribution made by this piece of research.

2.1 Prediction studies in the United Kingdom

Selection for Secondary Grammar schools. The majority of prediction studies in the U.K. are preoccupied with finding the most accurate means of predicting success in secondary schools.

McClelland (1942:1-2) reported on a very detailed inquiry which set out to "find the system of assessment

of ability and attainment which will give the most reliable forecast of the type of post-primary courses for which each pupil is fitted". The questions asked concerned the relative predictive values of

(a) traditional examinations

(b) teachers' estimates

(c) group intelligence tests and

(d) standardised scholastic tests

and the best combinations of these measures in selecting pupils for secondary courses were sought. The study comprised all the 3,229 pupils involved in the selection procedures in a Scottish city in 1935-36 and some 1,000 correlation coefficients were computed manually by the team, and adjusted to take into account the population elimination effect of selection (Cronbach, 1961:351). Owing to the disruption caused by the Second World War, the General Certificate of Education (G.C.E.) could not be used as the final follow-up criterion and so a much less satisfactory general estimate of the pupils' progress at the conclusion of the first three years of the post-primary course was used. Taken singly, the Qualifying examination (Q, a non-

standardised test in English and Arithmetic) correlated more highly with the criterion, followed by Teachers' estimates of ability in English and Arithmetic (Ts), standardised tests in Arithmetic and English (S) and then Intelligence Quotient (IQ). The battery giving the highest multiple R was IQ + Q + Ts. McClelland does not draw attention to the fact that his data show that the Intelligence Test was the best single predictor of success after three years in the senior secondary (grammar) school.

Rutter (1950:3-11) showed that a single standardised intelligence test used during the war years 1940-43 for selection to secondary grammar schools gave predictions of .63, .43, .62 and .68 respectively. Schonell (1949:24) asserted that previous research had shown that an intelligence test is the best single prognostic element in an examination test battery, but Stubbins (1940-16) had already demonstrated that in the schools and examinations he studied the entrance examination English results would have been the most reliable of the four sets of results (the others being Arithmetic, IQ and Teachers' Estimates) in predicting marks

in the majority of school certificate subjects. For Arithmetic to be of use in predicting Mathematics results it had to be combined with IQ.

Schonell did however add (1949:24) that "when combined with the results for English and Arithmetic the [Intelligence Test] materially improves the examination as an instrument of selection". Peel and Rutter (1951: 30-35) concluded that the standardised intelligence test was the best single predictor of achievement at school certificate level, confirming perhaps what McClelland (1942:78) surmised that it is "very probable that in the later years of the course IQ rises in predictive value relative to the other Qualifying measures". It would appear in any case that there is a rise in correlation between the selection examinations and school success with passage of time due in part, according to Emmett (1945:33), to a more accurate criterion in later years and in part to better efforts from the children as they become more mature. In a later study by Emmett and Wilmut (1952:52-62) the two samples consisted of pupils admitted to four grammar schools in the years 1941 and 1942 who took the School

Certificate five years later. Their scores in the Moray House Tests used for the entrance examination were correlated with an aggregate of equally weighted scores in School Certificate English Language, English Literature, History, French and Mathematics and with each subject individually. For the purposes of comparison between studies they recommend correction of correlations derived from their sample to population values. Their raw correlations however are of interest for this study as the situations are comparable and a selection of relevant results is given in Table 2.1 (on next page). Correlations of test scores with individual subjects ranged from .224 (English with Physics) to .514 (Intelligence with Mathematics) excluding Geography and Biology where correlations were not significant.

An analysis of variance was carried out to test the effects of the four different grammar schools on their pupils, controlling for varying performance in the entrance test. The difference was highly significant ($p < .001$).

Wrigley (1955:107-116) attempted to relate the

Table 2.1

Correlations & Multiple Correlations for five S.C. subjects pooled*

Year		1941/46	1942/47
No. of pupils in sample		128	153
Moray House Tests	Intelligence	.499	.583
	English	.476	.488
	Arithmetic	.424	.408
Aggregate of tests		.569	.570
Multiple R		.577	.578

*Adapted from Emmett & Wilmut (1952:61)

different methods of presenting correlation coefficients in prediction studies, and concluded that a "survey of published research showed that the relative order of merit for the prediction of success in grammar schools by the component parts of the Eleven-plus selection examinations was (1) intelligence tests (2) English (3) Arithmetic". However in two surveys he conducted, in Lancashire and N. Ireland, Arithmetic appeared to be the best individual predictor. Yates and Pidgeon (1957: 109) accepted Wrigley's work on the importance of the

Intelligence test as a predictor when linked with English and Arithmetic as a result of their own survey in which they contrasted traditional type examinations with standardised tests correlating their scores with teacher assessments after two years of secondary education.

The Fife investigation reported by McIntosh (1959) showed Arithmetic in third place in all sets of correlations between tests and achievement, including Leaving (School) Certificate. Though English often ranked higher than IQ, the significance of the difference especially in correlation with the School Certificate is questionable. Consistently best as a predictor was the entire battery comprising two Intelligence tests and Teachers' estimates as well as tests in English and Arithmetic. (Coefficients were again boosted to counteract the effect of selection and so cannot easily be compared with those of other studies.) In predicting success in individual subjects in the Leaving Certificate, the battery was more efficient for English, French and Latin, than for subjects with a practical bias such as Mathematics and Science.

Hockey (1968:140-7) used the number of subjects passed at GCE O-level as his criterion when investigating the effectiveness of the Common Entrance examination for admission to Marlborough College at the age of 13. He confirmed that the best single predictor of success was the Moray House Intelligence test then used, while the best predictors of O-level performance in each subject were the traditional-type papers in those subjects in the Entrance examinations.

In the 1960's interest in the U.K. turned away from testing for selection as comprehensive education gained ground and there is little incentive to investigate the validity of methods of selection in areas where they are still applied. Most selection batteries include Teachers' estimates in English and Arithmetic which usually correlate significantly with test results in those subjects, and to avoid resentment at the selection procedures in N. Ireland, the Government have implemented the proposal that selection should depend basically on a Transfer Report compiled by the Principal Teacher of the pupil's primary school containing general information about the pupil, about his

attainment in English and Mathematics and whether he is above average, average or below average in each subject and in verbal reasoning (N.I. DOE 1977:10). The wheel has gone full circle from the early selection procedure of the 1920s where subjectivity in evaluation was the order of the day. Even in 1952, Emmett and Wilmut (1952:53) had to point out that "the widespread tendency to allow good character to compensate for lack of ability when selecting grammar school entrants is not considered to be in the interests of the child".

Other factors in prediction. In their study of admission procedures to grammar schools already referred to Yates and Pidgeon (1957:160-65) recommended that age, sex differences and differences in primary schools should be taken into account. They demonstrated too that there is a significant relationship between a child's chances of being allocated to a grammar school course and the amount of encouragement that he is judged to be likely to receive from his parents. Concerning the assessment of personality, they concluded that there were at that time "no means available to education authorities of effectively identifying and assessing the personal

qualities that influence a child's chances of success in a grammar school course" (p.163).

More recent research has been stimulated by American studies in the area of personality factors and school achievement. Butcher et al. (1963:276-85) showed that intelligence, conscientiousness and self-sufficiency all correlate positively with achievement, but they attempted no intercorrelations with other predictors such as a full Intelligence test.

In his survey of studies in this area Rushton (1966:183) pointed out that 11 out of 13 studies with children connect stability, as opposed to anxiety, with ability while some 17 out of 19 studies among older children showed a significant connection between extraversion and school success. He concluded his own study of 11-year old school children by asserting that "the stable extravert children . . . have superior attainment". It would seem from Douglas and Ross (1964:185-196) that those who reach puberty early also have superior attainment. Perhaps they are the same children!

Achievement is obviously the cumulative result of ability, personality and motivation operating over many

years. Personality and motivation especially are both subject to change, much more so than ability measures, and the present simple additive model, Catell et al. (1966:293) felt, "falls far short of what is theoretically required. Thus the fact that these measures at a given year predict between a half and three-quarters of the variance is remarkable". They had in fact shown from the results of a study among adolescents from the Seventh, Eighth and Ninth grades in American Schools that personality and motivation source trait measurements contributed to the prediction of school achievement. However, correlations, though significant, should not be expected to run in general above 0.1 and 0.2 when what are essentially independent trait factors are correlated with achievement. They found that seventy-five per cent prediction of the criterion variance is the limit for such measures since they take no account of variations in schools, teaching quality, home facilities and the intervention of illness and accident in the examination room. They concluded:

> The encouraging magnitude of this prediction does, however, justify our position that educational prediction should not be based on ad hoc tests and scales but on comprehensive tests of the main source traits in personality structure and the

> main dynamic traits in the total interest area. (p.294)

Their research gives further indication of the results to be expected from the present study. Correlations in the order of .60 to .70 are all that can be expected (when not boosted for restriction of range due to selection factors) from relating tests taken early in a child's secondary school course with his results at the end, while the inclusion of other factors (outside the scope of this research) could be expected to produce correlations of up to .80.

In the British context, Sumner and Warburton (1972) differentiated between parts played by personality, ability, home background, school and tone of school in school achievement, while Naylor (1972) produced an extensive survey of research already undertaken in this area.

Prediction of achievement at post-secondary level. As far as prediction of achievement at higher levels of education in the U.K. is concerned, there does not appear to be much interest other than the reports by the National Foundation for Educational Research (NFER), one of which described the Test of Academic Aptitude

(Choppin et al., 1973) designed to offer an alternative to GCE A-level for prediction of success at university. The test was not found to have greater predictive validity than the traditional examinations.

Certificate of Secondary Education (CSE) grades were used in research on their predictive value for further education by Williams and Boreham (1972), who took as their criteria pass/fail categories in certain City & Guilds and Royal Society of Arts examinations. When difficulties were encountered in constructing a sampling frame, the entire population was studied. Each CSE subject was considered separately and pairs of subjects that had the highest correlations with the criterion were used in multiple correlation. The weights were then applied to scores in the two subjects of the best pair and a mean score was calculated for each candidate in a randomly selected group consisting of about half the candidates (the development group). Mean scores were computed for those who passed in the criterion and for those who failed and a cut-off point was established about mid-way. This was applied to the other group (the 'trial' group) and the pro-

portion of those whose results were correctly predicted was tested for significance to see if there was an important or random classification.

2.2 Prediction studies in the U.S.A.

It is at the transition to tertiary education that prediction studies in the U.S.A. proliferate. Educational and Psychological Measurement, for example, has published numerous reports on studies designed to predict success at College based either on achievement at high school or College Entrance test batteries. (Vide Vol. 19, No. 4). It is regrettable that many studies seem to ignore research already undertaken in the field, and even reported in the same journal, and as each researcher forms his sample at his own convenience it is difficult to compare results. Only three articles will be singled out for consideration.

Klugh and Brierley (1959:625-6) found that the grade point average at high school of the freshmen at their college improved the predictive value of the School and College Ability Test (SCAT). Men and women were considered separately but no comparisons

were made. Hills and Klock (1965:145-50) used the Gulliksen-Wilks technique to determine whether there was a significant difference between the different regression equations using the College Entrance Examination Board's Scholastic Aptitude Test (SAT) scores and high school grade-point-averages to predict success in different schools in the university. They concluded that courses in the first year were sufficently similar not to justify distinguishing them for prediction purposes and a single multiple regression equation could be used for both male and female, as the only significant difference between the equations for the two sexes was in the constant. In a report on the use of the General Aptitude Test Battery (GATB) as a predictor of college success, Sharp and Pickett (1959:617-23) distinguished between students majoring in engineering and others when making predictions.

Predicting success in college seems to be big business and the Free Press in New York produced a volume (Astin, 1971) giving selectivity data for 2,300 American colleges permitting the applicant to consider his chances of success in various institutions. As

Bloom and Peters (1961:25) point out, "school grades, achievement test scores, or aptitude test scores will ordinarily correlate with college grades from 0.40 to 0.60. However, multiple correlations using two or more ot these in combination will usually range from 0.55 to 0.65". Astin not only controls for sex, high school grades, academic ability and college selectivity in predicting freshmen College grade-point-averages but also lists student characteristics that he considered relevant. As they might well apply to some extent to our own situation one of the tables he gives (Table 1.13, page 13) is reproduced on next page.

2.3 Prediction studies in Africa

Testing in occupational selection. On the African continent, early prediction studies were more concerned with occupational selection than success at school. Ord (1972:130) and Napier (1972:87-8) both pay tribute to the work of Bisheuvel and the National Institute for Personnel Research (NIPR) in Johannesburg in developing a General Adaptability Battery (group administered performance tests) for the occupational selection and

Table 2.2

Factors affecting achievement at College*

Students who do better than expected	Students who do worse than expected
Member of scholastic honor society	Turned in papers late
Attended private high school	Came late to class
Rates self high on academic ability	Made wise-cracks in class
Rates self high on desire to achieve	Went to the movies frequently

*Table taken from Astin (1971:13)

classification of industrial workers. Validities in terms of job performance of 0.60 to 0.70 are reported and in Ghana, Taylor (1962) reported validities as high as 0.77 in prediction of tradesman skills of West African Shell-BP employees. NIPR demonstrated as early as 1952 that if reasonable facilities and expertise are provided, performance and paper and pencil tests of European origin can be adapted so as to have as high a predictive validity and selection and classification usefulness in African communities as in the communities for which they were originally designed (Ord 1972:131).

In other countries, including French-speaking countries, Raven's Progressive Matrices received inordinate attention. Clarke (1948, 1949) reported their use amongst adults in what is now Bauchi State, while Napier (1972:94-5) describes work in Ghana by Bulley who considered that the Matrices were not of sufficient validity to be used alone:

> Training even in Ghana involves the understanding of English - the official language and medium of instruction in all schools and institutions of training. Used with vocabulary and comprehension tests based in each case on the type of vocabulary required in training, the Matrices should furnish a useful tool for selection. (Bulley, 1964:8)

A large company in Zaria still claims to use the Matrices as part of their personnel selection procedures.

Testing in educational selection. Care must be taken when comparing test scores across cultures.

> If one were using the same test to predict success on a common criterion . . . then as far as evaluation among either European or African peoples were concerned, the test score could be held to have finite and common meaning. This predictive meaning should be held separate from descriptive or environmental meaning. (Irvine 1966:27-8)

Adams (1964:219) insists that:

> The most important single principle in aptitude

> test interpretation is that the test results must be considered in the context of all other information on the student's abilities (available in school records or supplied by students and parents).

Irvine, in an earlier article (1963:45) drew attention to the heterogeneity of factors operating in unknown ways contributing to error varience, obscuring individual differences and reducing the ability to appraise comparatively the potential of individuals, and in his reports on his surveys in Malawi and Zimbabwe (1968: 209) he came to the conclusion that:

> The major influence on the scores was the quality of school attended. Also males did significantly better than females on verbal tests, especially those demanding information about scientific or social studies. Age (as reported, but not verified, since births were not registered by law) was negatively correlated with attainment even after holding number of years constant. Analysis of variance showed little or no significant association of scores with level of father's occupation, number of sibblings, family position or number of languages spoken.

Aware of the factors mentioned above, the present researcher has felt obliged to restrict this study to the purely predictive value of the tests used. No claims as to what the tests measure will be made. It is regrettable that the final results of Irvine's Central African survey do not appear to have been

published as they were to include regression on School certificate results with the groups of tests available in consecutive years of this longitudinal study. The most striking finding so far reported is that the entrance examination by which children were originally selected for secondary school was less related to Form 2 examination performance than any of the ability and attainment tests used in that year. A combination of Standard 6 examination results and psychological tests showed the closest association with the Form 2 results. This resembles the findings in the U.K. that assessments of attainment at primary school either through tests or teachers' estimates make a significant contribution to the predictive value of selection procedures for secondary schools.

Another factor influencing achievement in Africa is of course personality. Kline (1966:91-94) confirmed the Eysenckian theory predicting relationship between introversion and academic success among students by his study in Ghana (in what was then called the University College of Science Education, Cape Coast).

Development of the I-D tests. However, we are

most concerned in this study with the work of Schwarz who headed the team in Lagos which devised many of the I-D Tests in the early 1960s. "Africa has been the scene of much testing research but few testing programs", he wrote in 1963 (Schwarz, 1963:673) and less than ten years later Snider (1972:176) reported that during the period 1965-1971 approximately 600,000 I-D tests were administered in West Africa, and they have also been adapted for use in Brazil, Korea and Thailand. As a starting point, Schwarz selected tests of aptitude from a great variety of American batteries, aware that ". . . a typical unadapted American test can fail irrespective of content because it can contribute too massive a learning task for the African examinee" (Schwarz 1963:675).

He also stressed the desirability of an easily understood consistent format for all tests. He advocated spoken rather than written instructions, and particularly 'dynamic instructional techniques' using visual aids such as charts, cut-outs, and articulated models (which can even be handled by the testee if need be), extended practice and live demonstrations

that ensure feedback on practice items (Schwarz & Krug, 1972;79-109).

An examiner's manual was prepared in both English and French (Schwarz, 1964a) in such a way that it could be used as a text book to train new examiners. Complementary to this was a technical manual (Schwarz, 1964b) giving data on reliability, validity and standardisation of the 21 tests of the final battery. An important feature of the I-D tests was that they all proved equally applicable in three French-speaking countries (Tunisia, Mali and Cameroon). The tests could be divided into four general categories:

(a) "intelligence",
(b) perceptual and spatial abilities with emphasis on speed and accuracy,
(c) attainment, knowledge, interests and skills,
(d) psychomotor.

For each test Kuder-Richardson 20 reliability coefficients were calculated using groups of candidates in West Africa, Brazil and Thailand, while validity studies concentrated on correlating individual tests and series of tests with instructor ratings, examination averages

and course grades. There are basically three kinds of test "package":

(a) semi-skilled, skilled, technical and scientific,

(b) junior clerical, senior clerical and commercial,

(c) three academic levels according to years of schooling.

The Test Development and Research Division of the West African Examinations Council is continuing its research on these tests and at the same time devising a new series given the title "Potential Abilities Tests".

Validity Studies on the I-D tests. The I-D tests are sufficiently varied in conception to have a great variety of applications in the field of educational and occupational prediction, and certain tests are intended for certain levels of education. Of foremost concern in this study is the Scholastic Aptitude Test Battery which is actually the lowest level of the Academic series. The original validity study carried out by Schwarz and his team (1964b) in the early 1960s using schools throughout Nigeria which included Government College, Kaduna and Barewa College, Zaria gave a com-

bined unweighted validity coefficient of 0.56 when concurrent achievement ratings were correlated with the test scores.

When the testing programme got under way in Kaduna State, Ndahi (1975) in a project for the Department of Education at Ahmadu Bello University, Zaria, collated information on test scores and examination results at the end of the third and fourth years in English, Mathematics and Science from Zaria schools, and concluded that there was significant evidence indicating a relationship between the two sets of scores, but added:

> The degree of the correlation, however, is subject to further investigation and cannot readily be estimated. (p.41)

At the same time, the Test Development and Research Division administering the tests in the State collected data on the subsequent progress of the entire population tested. The results should be available some time in 1978.

Akeju (1966) outlined the use of three of the higher level academic tests (in Verbal Analysis, Reading Comprehension and Arithmetic) for selecting students for the Nigerian School of Science, Lagos. Applicants who

scored above average were invited for interview and 109 admitted. Correlations of these students' scores with their end of year marks in each subject were on average very low and in fact not one was significant! This could in part be attributed to the use of the tests in selection. Schwarz and Krug (1972:179) insist that for a good validity study "it is important to ensure that the test results are not used as a basis for selection decisions". It is of interest to note that Akeju found that the Arithmetic test correlated "very insignificantly" with Physics, Applied Mathematics and even Pure Mathematics. This may not be as surprising as he suggests, for U.K. studies already discussed show that success in Mathematics does not necessarily correlate highly with Arithmetic tests used in selection procedures at 11+.

In Majasan's study (1968) of the predictive validity of the I-D tests as instruments for selection into technical and commercial courses of study, nineteen of the tests were used (i.e. all except the lower level of Verbal Analogies and Reading Comprehension - VAL, RDL). 398 applicants to the Government Technical

College, Ibadan, were tested, and course grades of the 139 admitted were collected at the end of each term for three terms, and an average grade was calculated. The results show that "most of the tests . . . predict the course grades fairly well" (p.3). In actual fact, not all the tests even correlated significantly at the 5% confidence level and no single test predicted performance in all the courses in any group. (Unfortunately the course groups tended to have few students, ranging from 6 to 31). The Arithmetic test (RTH) correlated significantly with secretarial studies in each term and also in the overall results for Civil Engineering. The Memory Test (MEM) only showed a significant correlation once! A possible explanation for these correlations has already been mentioned - the tests were used as a means of selection.

Other validity studies on various batteries of I-D tests were reported at the 18th Annual Meeting of WAEC in 1970, and are not of immediate concern to this study.

Prediction of success at secondary school. The only research located using WASC grades as the criterion is Yoloye's (1973:48-59) in which 150 Form I pupils and 70 of the sample in Form 5 were given the Lorge-Thorndike

Intelligence Tests. In Form 5 it was the verbal IQ that correlated more highly with the aggregate scores and scores in individual subjects in WASC, while in Form I it was the non-verbal IQ which gave a better prediction (0.48) of all subjects except English Language, Mathematics, and Physics. Yoloye deduces that the language handicap for Nigerian students is very great and variable at that lower level of education. It is conceivable that a test of English at this stage might have more relevance as a predictor, as an understanding of English is a basic requirement for success at the secondary level. Napier (1972:97) quotes M. A. Brimer who had specialised in test design at the University of Ibadan as saying that any differentiation among students must be based to a large degree on their ability to cope with verbal, oral and written aspects of the English language.

Schwarz was aware of this situation and reported (1972:31) the existence of data showing that students entering a particular secondary school with high scores having come from a privileged primary school showed little effect of this advantage on completion of their

secondary course.

Comparing admission test scores of a sample of students from advantaged and disadvantaged locations it was found that the two distributions were almost non-overlapping. Further samples were taken and in addition to the normal secondary school admission test, a set of I-D scholastic aptutide tests were administered.

> Comparative data were available for approximately 500 students. On the admission tests the difference between the advantaged and disadvantaged samples were as large as before. But on the aptitude tests, the average scores of the two groups were almost exactly the same. (p.33)

It would appear that the aptitude tests were relatively unaffected by variations in school quality. It must therefore be assumed that the I-D Reading Comprehension test (RDL) is not merely an attainment test in English measuring the extent of the English teaching the pupil has received during the previous 6-8 years.

Attention should be drawn to Pilliner's complaint (1968:5-7) that test specialists often fail to conduct a thorough-going study of the validity of their tests. This is partly so because "the criterion in a predictive study is often many years away". "Since not all

educators are equally impressed by statistical results", he recommends that the study of the success and failure of students be undertaken, as such types of studies "are regrettably rare".

By way of conclusion, a comment by Aleyideino is appropriate (1973:5):

> The rationale for using tests for selection is partly based on need to devise a fair method of sharing out limited educational resources among a large number of competing candidates. This however is not the only rationale: tests are also used to determine whether or not candidates possess the capacity to profit fully from courses of instruction by institutions. In other words, tests are meant to predict how well a candidate could perform after undergoing some kind of training.

To what extent the I-D tests of Verbal Analogies, Reading Comprehension, Memory and Arithmetic can fulfil these functions is the basic purpose of this study.

2.4 Summary

In this chapter the relative merits of using standardised tests of intelligence and attainment, traditional tests of English and Arithmetic, estimates by teachers and school grades in predicting future success have been reviewed. The intelligence test would appear

to be the best single predictor in many studies, but prediction can be improved by the use of a battery of tests and results, and even by taking into account personality and environmental factors. Too many studies use subjective criteria for concurrent validity rather than the more acceptable criterion of final public examinations. Attention has been drawn to the part played by Arithmetic testing in prediction and to the importance of the English language both in the U.K. and in Africa.

Chapter 3

PROCEDURE

This study was designed to ascertain the relationship that could be expected to exist between the I-D Scholastic Aptitude Test Battery and results in the West African School Certificate. When the sample was decided upon, the data were collected, independent and dependent variables selected, and the hypotheses tested, after the scores had been adjusted to compensate for lack of homogeneity in the distributions.

3.1 The sample

The ideal prediction study requires a simple random sample from the population, which in this study should consist of all the pupils who have taken the I-D tests in Form 2 in Kaduna State secondary schools. As Catell and Butcher (1968:159) point out:

> The sampling problem is a difficult one in educational research because theoretical requirements and administrative and practical possibilities are generally opposed. Random or probability sampling is theoretically desirable, but often virtually impossible, and has only been employed in a tiny

> minority of published studies. Where it is not practicable, careful intelligent choice of samples and detailed reporting are essential.

The restrictions imposed by circumstances on the choice of sample for this study will therefore be outlined.

Due to the fact that the tests were first administered in 1973 only one class of pupils tested have sat for the West African School Certificate (in 1976); the population is of necessity reduced to the pupils in Form 2 in 1973. Of that population, some pupils did not complete their secondary school course, some were transferred to other schools, some concentrated on technical and commercial courses in WASC while others changed their names making the identification of their WASC extremely difficult. In addition owing to the unprecedented increase in secondary school enrolment, and the considerable variety of problems encountered in certain schools, pupils did not all have an equal chance of achievement, and there are schools where WASC results were so unsatisfactory that they can hardly be attributed to the lack of aptitude of the pupils alone. As this situation should not repeat itself, it did not seem relevant to include such schools in our sample. Choice

of sample was restricted to secondary schools in Zaria which are all well-established and with which personal contact was possible in order to acquire the necessary data.

Pupils in Form 2 of these three schools - Barewa College, Government Secondary School, and Kufena College - showed different levels of aptitude, thereby permitting a stratified sample which could improve accuracy in estimating parameters. Within each school a random sample of 50 (Minium 1970:279) was arrived at by first listing all the pupils who took the I-D tests in 1973 and the WASC examinations in academic subjects (as distinct from technical and commercial) in 1976 and whose results could be identified. This gave in all 101 pupils in Barewa College (BCZ) out of a total of 139 who took the tests, 63 in Government Secondary School (GSS) out of 124 and 69 in Kufena College (KCZ) out of 114. Then the required number of cases for each school was obtained by the use of random number tables. The procedure outlined gave rise to a stratified sample which was disproportionate, but the eventual findings could not be biassed as no generalisations were made from the means of scores in each stratum of the sample. On the other hand, a better representation of the subjects of the different strata was obtained. It should be noted that

some pupils whose results could have been very poor did not take the WASC examinations and were not given a chance of being included in this sample. The use of statistical techniques related to simple random samples without taking into consideration the particular nature of our sample should give results which err on the 'conservative' side. (Butcher 1966:10). As prediction concerns individual pupils it is the guiding principle of interpretations in this study to err on the 'safe' side and in testing hypotheses stringent levels of significance will be applied.

3.2 Collection of Data

A copy of the West African Examination Council's computer print-out of the scores in the Scholastic Test Battery of all the pupils tested in 1973 in the three schools of the sample was made available by the Ministry of Education, Kaduna, while the results of WASC were furnished by the Principals of the individual schools. As far as was possible with the help of staff of the schools, the WASC results were collated with the test results of the pupils who had taken both in the same school.

3.3 The Independent Variables

Five scores for the I-D Scholastic Test Battery were provided for each candidate: an aggregate score for the tests (termed here for convenience: ALL) and scores for the four individual tests. These tests are (quoting from Kure and Bennett, 1974:L-16):

> VAL (Verbal Analogies Low form): a verbal reasoning test used to predict success in school or a job requiring formal studies.
>
> RDL (Reading Comprehension Low form): a test of ability to read and understand written material, used to predict academic potential.
>
> MEM (Memory): a test of ability to learn and remember material organised in a meaningful way.
>
> RTH (Arithemtic): a test of speed and accuracy in doing simple computations.

In addition, the aggregate of the first two tests, VAL + RDL, was computed and treated as a sixth variable.

The scores in the above tests are published in Stanines, i.e. standardised scores with a range from 9 to 1, based on the Normal curve. The mean is 5 and each interval is half-sigma giving a standard deviation of approximately 1.96. Unfortunately the test results of 1973 in Kaduna State were not awarded according to the Normal distribution but according to national norms

established 10 years earlier. (In 1974, 1975 and 1976 the results were normally distributed, while in 1977, norms for 1976 were applied.)

3.4 The Criterion Variables

The criterion in this study is the WASC examination as

> performance in this examination serves as a basis for selection to jobs and higher institutions of learning. The entire secondary grammar school curriculum is therefore geared to this examination. (Yoloye 1972:49)

Candidates usually take eight out of a wide range of subjects and the certificate is awarded on the basis of the aggregate of the six best subjects. (To earn a Division I or Division II certificate the candidate must satisfy certain standards in English and in Mathematics or an allied subject.) Marks are converted into scores on a 1 to 9 grading system, but the distribution is far from Normal, as it appears that in some subjects it is possible for 50% of the candidates in Nigeria to receive the lowest grade of 9. The grades from 2 to 8 are assumed to be on an interval scale.

According to Ohuche and Akeju (1975:106) the WASC

examinations are standardised

to the extent that

(a) the marks are assigned after a standardised fashion,
(b) the subject grades are awarded on the basis of the raw marks obtained and
(c) the final certificate divisions are given according to fixed standards.

They fall short of the I-D standardised tests in that

(a) the standards fixed at a standardisation meeting are subjective
(b) no norms were obtained on previous samples and
(c) the assignment of grades to subjects by Chief Examiners and Moderators is also subjective and not fixed.

Though the aggregate in 6 subjects is the most considered score there is no pass mark as such because of various conditions imposed. It is possible in theory to be awarded a certificate with an aggregate of 48, but in actual practice candidates with an aggregate of more than 40/42 are seldom successful. To err on the 'safe' side we consider an aggregate of more than 44 as highly undesirable and deem it a failure, for the purpose of the statistical demands of this study. In fact 77 out of 470 candidates in Zaria in 1976 received aggregates between 45 and 48, but only two were awarded passes - with an aggregate of 45.

The criterion variables treated are:

ENG (English Language)

MATHS (Mathematics)

AGG (Aggregate of grades in English Language, Mathematics, Geography and Biology)

GEN. AGG (General aggregate of six best grades)

No information on the reliability of the WASC examinations is available, and though no claims of validity are made for them they have been taken as the criterion in this study, because of their unquestioned acceptance by employers and educational administrators in the country.

3.5 Processing of Data

The ten variables referred to in 3.3 and 3.4 were punched on Computer cards for each of the 150 pupils of the sample. The computer used, an ICL1903A, was already programmed for basic statistical operations, and so a print-out was obtained of the distribution of each variable, the mean and standard deviation, and the skewness and kurtosis for the curve.

In addition a Pearson's Product-Moment Correlation matrix was produced, and multiple regression equations

formulated for certain variables.

It was very evident that the distributions of most of the variables in this sample lacked normality. The WASC grades tended to give very skewed distributions owing to the high failure rate in many subjects, while the application of inappropriate norms caused the I-D test scores to be skewed, thereby preventing the Stanine scores allegedly used from conforming to the distribution such a scale is supposed to reflect.

For satisfactory regression analysis normality is usually assumed and in interpreting correlation coefficients, according to McNemar (1969:186-7)

> any effect of skewness will disturb the comparability of r's when the relationships are relatively high.

He goes on to recommend the application of the normalizing T score transformation to the variates.

Scores were therefore normalised and expressed as T scores with a mean of 50 and Standard Deviation of 10, by first calculating the rank (making a slight correction for ties) and then dividing by the total number of cases (i.e. 150) to give a cumulative proportional frequency for each score for which the T-score could be

read from Guilford's table (1965:420).

Using the transformed scores, the same procedure for obtaining details of the distributions, Pearson's r and regression equations was carried out.

3.6 Hypothesis Testing

The null hypotheses based on the first three hypotheses given in chapter 1.5 were tested using a significance level of 0.001 with two tailed tests. In addition, confidence intervals were computed to estimate the population parameters, following the recommended procedure of Popham & Sirotnik (1973:60-61):

> Interval estimation operations are considered to answer different kinds of questions than those allowed by hypothesis testing operations. Hypothesis testing tells whether a relationship exists, interval estimation procedures indicate how strong the relationship is.

Scattergrams were made to check for linearity and homoscedasticity before accepting results of the regression analysis. The most useful equations were formulated giving error of prediction and F-value with level of significance.

The procedure for testing the two remaining hypotheses has been outlined in a presentation of the

results. Chapter 4 is designed not merely to present the results but also to make clear the processes involved and the interpretations possible when certain statistical techniques are applied. In this way, it is hoped to facilitate the replication of this study.

Chapter 4

RESULTS AND DISCUSSION

As stated in the previous chapter, the independent variables - the I-D Aptitude Test Battery - have Stanine scores from 9 to 1, 9 being the highest score and 1 the lowest, while the dependent variables - the WASC examination grades are based on a non-standard scale from 1 (highest) to 9 (failure). Correlation co-efficients which indicate a strong relationship will therefore be negative. Rather than invert one of the systems it was felt simpler to make the transference in interpreting the negative co-efficients computed - they will be recorded as positive - while taking care to watch out for the possibility of an unexpected positive correlation which would indicate a negative relationship.

4.1 Hypothesis I

That there is a significant relationship between scores in the I-D Scholastic Aptitude Test Battery

and the aggregate scores in the West African School Certificate.

Pearson's Product-Moment Correlation Co-efficients for each test and the aggregate was computed using both the original and the normalised scores. A Multiple R was computed taking the two most significant tests and all the tests. This information has been set out in Table 4.1 to which the original validity coefficients using concurrent criteria as reported in the I-D Technical Manual (Schwarz, 1964b) have been added for the sake of comparison.

In view of the principle of requiring a very stringent confidence level, the Memory Test cannot be claimed to correlate significantly with the criterion of General Aggregate in WASC.

The null hypothesis, for all tests excluding the Memory Test, that r = 0 can be rejected and the alternative hypothesis that there is a significant relationship can be accepted. Normalising the scores does not actually make a significant difference to the correlation coefficients.

However, our concern is with the extent of the

Table 4.1

Correlation of test scores with General Aggregate in WASC

r	GEN. AGG (WASC)		Validity Coefficients (Schwarz, 1964)
	Original Scores	T-Scores	
VAL	.61	.62	.45
RDL	.63	.64	.45
MEM	.21*	.21*	.20
RTH	.33	.30	.32
ALL	.63	.61	-
VAL + RDL	.68	.69	-
R			
VAL/RDL	.68	.69	-
ALL TESTS	.68	.69	.56

*p < .01. For other results p < .001 (2-tailed test)

relationship and not merely its existence.

The highest co-efficients are for the VAL and RDL taken as an aggregate and the multiple R for tests taken together. From the computation of multiple correlation coefficients, it is evident that the Memory and Arithmetic Tests add little to the predictive value

of the test battery. This is clear from Table 4.2 where it will be seen that 41% of the variance in WASC scores can be attributed to RDL and 7% to VAL, and that the contribution of the two remaining tests is not significant.

Table 4.2

Increase in correlation coefficient (R) as each test is added

	Multiple R	R square	R sq. change	F-ratio	Level of significance
RDL	0.6426	0.4129	0.4129	28.34	.001
VAL	0.6949	0.4829	0.0700	16.21	.001
RTH	0.6952	0.4833	0.0004	0.12	n.s.
MEM	0.6953	0.4834	0.0001	0.03	n.s.

It can therefore be concluded that RDL and VAL are the best predictors of success in WASC. When the aggregate of the two equally-weighted tests (the actual weights are 6 and 5) is used in simple correlation, the coefficient is similar, suggesting that there is considerable

overlap in what the two tests actually test. In fact, they correlate significantly with each other. (r = 0.64).

It is interesting to note the improved correlation between VAL and RDL scores and WASC grades over that of the criteria used in the original validity study mentioned in 2.3, especially when the greater homogeneity of the present sample is likely to have reduced the size of the correlation coefficient. It is tempting to surmise that there is here some confirmation of McClelland's theory (1942:79) that 'intelligence' scores "rise in predictive value relative to the other measures as the years go on".

In order to predict aggregate scores in WASC, it is evident that VAL and RDL will make the most satisfactory regression equation. The regression line giving the least sum of squares for the residual, as shown in Table 4.3, is based on taking RDL and VAL separately in multiple regression, though little is sacrificed by using the aggregate of the two tests.

The most accurate regression equation which can be arrived at using the test scores is:

Table 4.3

Residual Sums of Squares for different predictions of GEN. AGG (T-scores)

Tests used in Regression	Residual SS
RDL	8420.34
RDL + VAL	7430.16
RDL, VAL	7415.93

$$\hat{Y}' = - 0.357\ X_1' - 0.435\ X_2' + 89.58 \qquad (4.1)$$

where $\hat{Y}'$ is predicted T-score in GEN. AGG

X_1' = T-score for VAL

X_2' = T-score for RDL

For this equation F = 68.64 with degrees of freedom of 147 and 2, and the standard error of estimate is ± 7.05. The confidence limits for 95% of the cases would be ± 13.82. The alternative regression equation using the equally-weighted aggregate of VAL and RDL in T-scores would be

$$\hat{Y}' = - 0.396\ X_{1+2}' + 89.57 \qquad (4.2)$$

(F = 135.67 with degrees of freedom 148 and 1).

X_{1+2}' = aggregate of T scores for VAL and RDL

The standard error of estimate is the same as for Equation 4.1. The scattergrams of Figure 4.1 demonstrate the effect of the transformation from original scores to T-scores. The T-score arrays can be seen to satisfy the requirement of homoscedasticity. For the use of the Guidance Teacher in subsequent years when the actual test scores awarded follow the normal distribution, Equation 4.2 can be expressed thus:

$$\hat{Y}' = - 1.845\ X''_{1+2} + 68.32 \qquad (4.3)$$

(F = 140.13 with df 148 and 1)

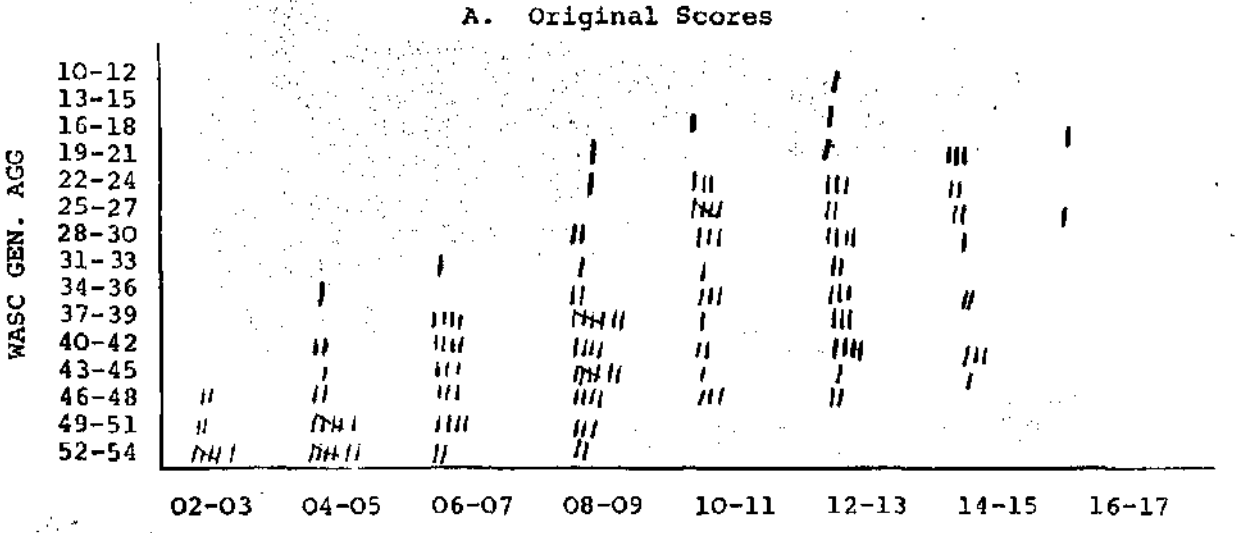
where X''_{1+2} = sum of VAL and RDL in true stanine scores and $\hat{Y}'$ = predicted T-score for GEN. AGG.

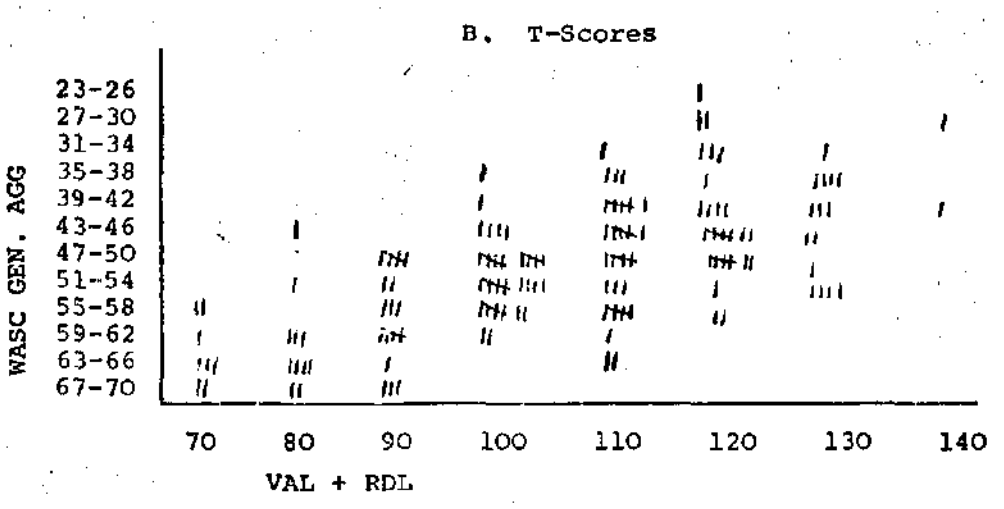
The standard error of estimate is ± 7.03.

The T-scores predicted for GEN. AGG may be converted into percentiles which can be interpreted more readily to the student. Table 4.4 is designed to give a guide to the relationship between T-scores, Centile Ranks and Stanine scores. A more accurate and complete table can be derived from the tables giving the area under the normal curve in terms of Z found in the appendix of any basic textbook in statistics, or from the table specially prepared by Guilford to aid the calculation

Figure 4.1

Scattergrams showing relationship between VAL + RDL and WASC GEN. AGG





of T-scores (1965:520).

Table 4.4

Corresponding scores in three scales - T-scale, Centile Rank, Stanine

T-scale	Centile Rank	Stanine
73.5	99.0	9
67.5	96.0	
65.0	93.0	8
62.5	89.5	
60.0	84.0	7
57.5	77.5	
55.0	70.0	6
52.5	60.0	
50.0	50.0	5
47.5	40.0	
45.0	30.0	4
42.5	22.5	
40.0	16.0	3
37.5	10.5	
35.0	7.0	2
32.5	4.0	
29.5	2.0	1
26.5	1.0	
Mean: 50 S.D.: 10		Mean: 5 S.D.: 1.96

In addition, it is possible to make approximations with original scores by use of graphs (Figures 4.2 and 4.3) showing the transformation to T-scores of the original scores in the two tests, VAL and RDL and in the General Aggregate in WASC. A few cases have been selected to demonstrate the application of the prediction equation 4.1 to our original data, and this information is presented in Table 4.5.

It is not possible to obtain satisfactory predictions from a regression equation based on the original scores for GEN. AGG as these scores were not intended to be normally distributed and predicted scores inevitably follow the normal distribution.

4.2 Hypothesis 2

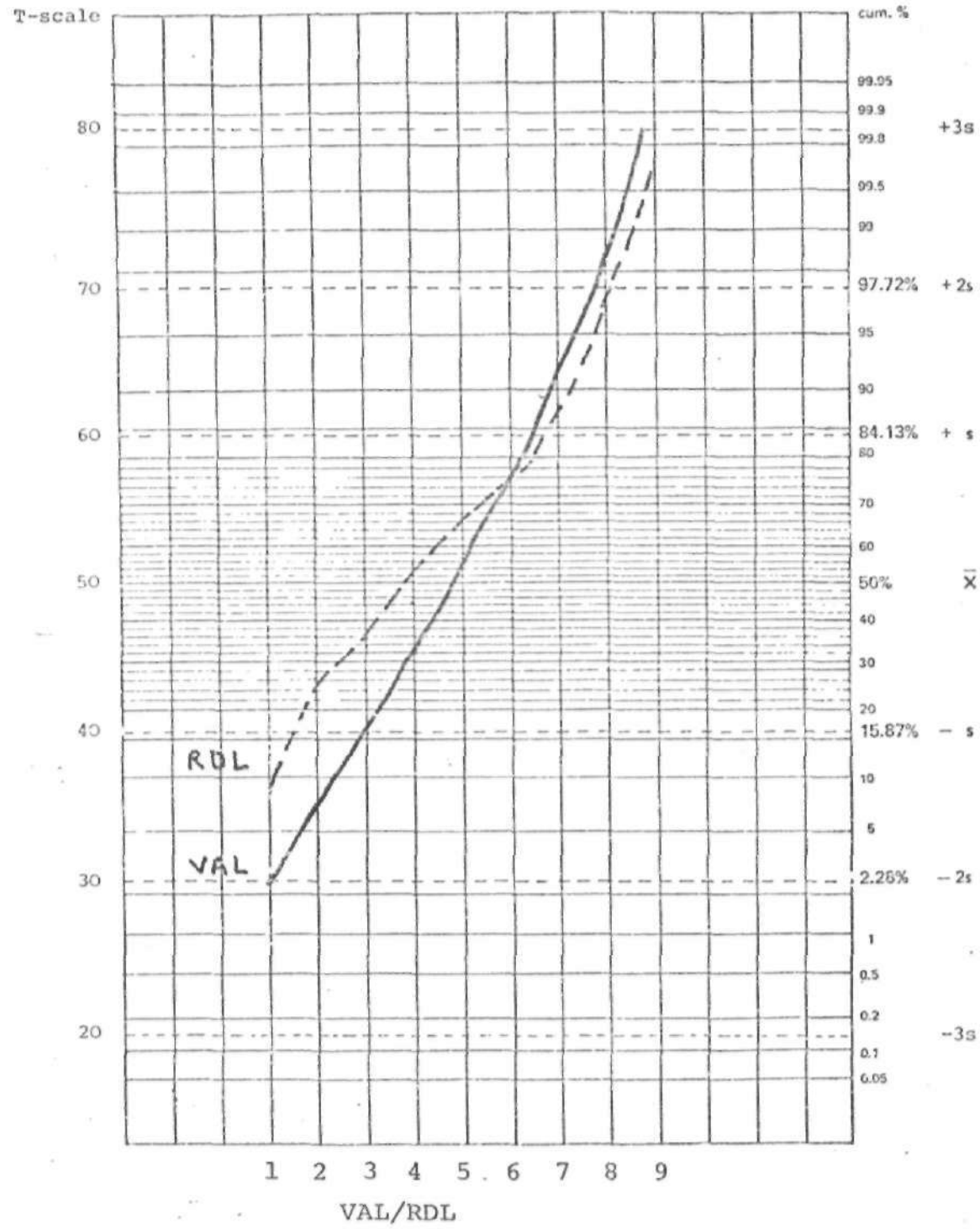
That there is a significant relationship between the scores in the I-D Scholastic Aptitude Test Battery and WASC grades in English Language and Mathematics.

Pearson's product-moment correlation coefficients were computed for both original and normalised scores and are presented in Table 4.6.

All the coefficients (except that of MEM) are

Figure 4.2

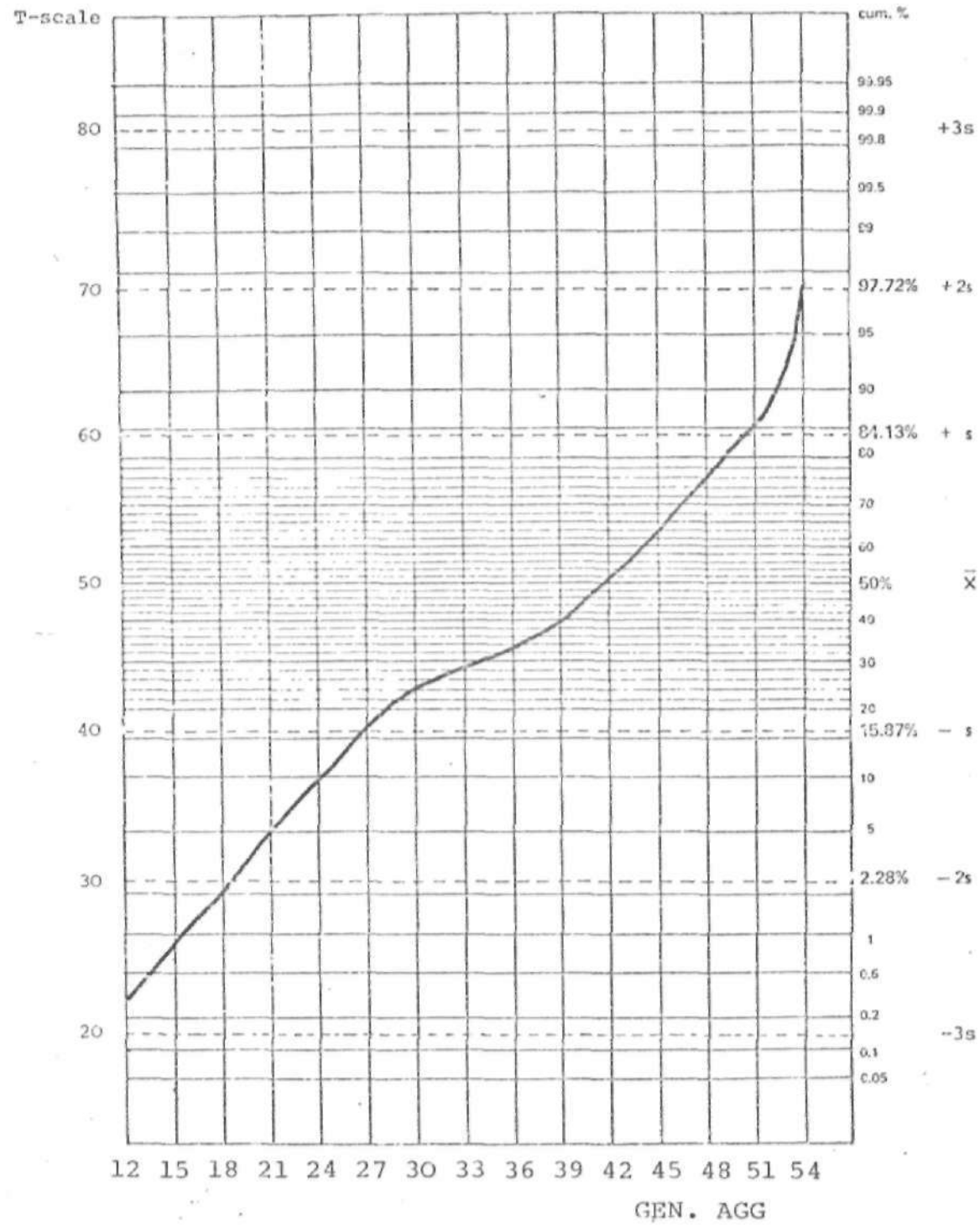
Graphs showing Transformation from Original Scores To T-Scores for VAL and RDL



Cumulative percentage frequencies on arithmetic probability scale (y) : v. class boundaries (x)

Figure 4.3

Graph Showing Transformation from Original Scores to T-Scores for GEN. AGG (WASC)



Cumulative percentage frequencies on arithmetic probability scale (y) : v. class boundaries (x)

Table 4.5

Examples of the Application of The Multiple Regression Equation (4.1) Predicting GEN. AGG from VAL and RDL

Case No.	X_1	X_1'	X_2	X_2'	$\hat{Y}'$	Y'	Residual	$\hat{Y}$ (approx.)	Y
067	1	30.8	1	33.3	64.1	57.1	+ 7.0	52	48
042	2	35.8	2	40.8	59.1	62.8	- 3.7	50	52
071	3	40.4	3	45.2	55.5	59.1	- 3.6	46	50
017	4	45.8	4	49.1	51.9	41.8	+10.1	43	28
020	5	51.5	5	52.9	48.2	41.8	+ 6.4	40	28
015	6	57.8	6	55.9	44.6	46.8	+ 3.8	34	27
004	7	64.2	7	59.5	40.8	36.4	+ 4.4	27	22
003	8	71.1	8	65.0	35.9	29.6	+ 6.3	23	16
133	8	71.1	9	73.1	32.4	40.8	- 8.4	20	27
No. of cases	150	150	150	150		150			150
Mean	4.64	49.98	4.41	50.08		49.95			38.87
S.D.	1.66	9.50	2.24	9.47		9.81			10.66

Equation: $\hat{Y}' = - 0.357\ X_1' - 0.435\ X_2' + 89.58$

where $\hat{Y}_1'$ = predicted T-score for GEN. AGG

X_1' = T-score for VAL

X_2' = T-score for RDL

$\hat{Y}$ = approximate predicted score for GEN. AGG using original scale

Y' = T-score in GEN. AGG

Y, X_1, X_2 = original scores in GEN. AGG, VAL and RDL

Standard error of estimate: ± 7.05; 95% confidence limits: ± 13.82 (T-scores)

Table 4.6

Correlation of Test Scores with English and Mathematics in WASC

r	ENG		MATHS	
	Original	T-Score	Original	T-Score
VAL	.56	.62	.48	.52
RDL	.59	.64	.48	.47
MEM	.17+	.17*	.27*	.27*
RTH	.25	.28	.47	.43
ALL	.56	.59	.59	.55
VAL + RDL	.63	.69	.53	.55
R				
VAL/RDL	.63	.69		
VAL/RTH			.56	.57
ALL TESTS	.63	.70	.60	.59

+ $p < .05$ * $p < .005$ Other coefficients: $p < .001$

statistically significant. Therefore the null hypothesis that $r = 0$ can be rejected in all cases except that of the Memory Test. Each test except the Memory Test as well as the aggregates of the test scores relates significantly to the criterion, both English

Language and Mathematics. Not unnaturally the Arithmetic test (RTH) plays a more important part in predicting Mathematics grades, though generally the composite variable (VAL + RDL) seems the most useful general predictor and the administration of the entire battery gives little improvement in prediction except in the case of Mathematics.

To formulate the regression equation to predict scores in English, it is relevant to consider the significance of the contribution of each test to the multiple regression. From Table 4.7 it is clear that only RDL and VAL can be used. The contribution of the Memory Test is not significant and that of RTH was insufficient to warrant computation. The F-ratio used to test for the significance of the addition to the coefficient when a new variable is added consists of dividing the change in R^2 by $(1 - R^2)/N - k - 1$ where k is the number of variables used (with degrees of freedom of $N - k - 1$

and 1).

Table 4.7

Increase in correlation coefficient (R) as each variable is added - ENGLISH LANGUAGE

	Multiple R	R Square	R^2 change	F-ratio	Significance
RDL	0.6370	0.4058	0.4058	101.05	.001
VAL	0.6950	0.4830	0.0772	21.97	.001
MEM	0.6966	0.4852	0.0022	0.62	not sign.

The regression equation for predicting English grades in WASC using T-scores would therefore consist only of VAL and RDL:

Either

$$\hat{Y}' = - 0.332\ X_1' - 0.369\ X_2' + 84.55 \qquad (4.4)$$

(F = 68.67 with df 147 and 2)

or

$$\hat{Y}' = - 0.350\ X_{1+2}' + 84.54 \qquad (4.5)$$

(F = 138.12 with df 148 and 1)

$\hat{Y}'$ = predicted T-score for ENG

X_1' = T-score for VAL

X_2' = T-score for RDL

The standard error of estimate is ± 6.27 for two-thirds of the cases, giving ± 12.29 for the 95% confidence limits.

When the T-scores for VAL and RDL are converted into Stanine scores with a mean of 5 and standard deviation of 1.96, the equation 4.5 can be written

$$\hat{Y}' = - 1.605\ X''_{1+2} + 65.60 \qquad (4.6)$$

(F = 132.19 with df 148 and 1)

where X''_{1+2} is the aggregate of Stanine scores in VAL and RDL. The standard error of estimate is ± 6.32.

In the sample used in this study, the normalising effect of the transformation into T-scores reduced the skewness of the distribution of the English grades from - 1.319 to - 0.609. The skewness could not be completely eliminated as 63% of the candidates had the lowest grade of 9. Equations 4.4, 4.5 and 4.6 are therefore valid for establishing a percentile rank for each candidate, which could be interpreted according to the pass rate expected in the institution concerned.

In considering the prediction of Mathematics grades it will be noticed that the criterion correlated similarly with three of the tests, VAL, RDL and RTH. In stepwise regression, RTH was entered in Step 2

rather than RDL because presumably VAL and RDL overlap considerably in their predictive value. The effect of adding each variable in the multiple regression and the F-ratio for the significance of the addition are presented in Table 4.8.

Table 4.8

Increase in correlation coefficient (R) as each variable is added - MATHEMATICS

	Multiple R	R square	R^2 change	F-ratio	Significance
VAL	.5200	.2704	.2704	54.85	.001
RTH	.5671	.3216	.0512	11.09	.01
RDL	.5906	.3488	.0272	6.09	.05
MEM	.5929	.3515	.0027	0.60	not sign.

As the level of significance chosen for this study is 0.001, the contribution of RTH and RDL in the multiple regression cannot be considered as statistically significant. The regression equation derived from this situation is therefore:

$$\hat{Y}' = - 0.428\ X_1' + 70.68 \qquad (4.7)$$

(F = 54.85 with df 148 and 1)

where $\hat{Y}'$ = predicted T-score for Mathematics

X_1' = T-score for VAL

The standard error of estimate is ± 6.68.

Using the aggregate of VAL and RDL again, a slightly more accurate prediction is possible:

$$\hat{Y}' = - 0.249\ X'_{1+2} + 74.209 \qquad (4.8)$$

(F = 63.48 with df 148 and 1)

where $\hat{Y}'$ = predicted T-score for Mathematics

X'_{1+2} = aggregate of T-scores in VAL and RDL

The standard error of estimate is ± 6.54.

When the scores in VAL and RDL are converted into stanines, the equation becomes:

$$\hat{Y}' = - 1.150\ X''_{1+2} + 60.75 \qquad (4.9)$$

(F = 62.75 with df 148 and 1)

where X''_{1+2} = aggregate of stanine scores in VAL and RDL.

The standard error of prediction in this equation is ± 6.55.

Although the aggregate of VAL and RDL correlates more highly with English Language than Mathematics, it does appear that it is the most useful predictor of grades in both subjects.

4.3 Hypothesis 3

That grades in the four subjects taken by all students in WASC - English, Mathematics, Geography and Biology - may be predicted with more accuracy than the aggregate of the six best subjects.

This hypothesis originated from the assumption that the four subjects chosen freely by the candidates would be tackled with more success, being more related to the pupils' interest and including what are alleged to be 'easy options'. It therefore seemed possible that the aggregate in the four 'compulsory' subjects would reflect more closely the ability of the pupils shown in the Scholastic Aptitude Tests. (The aggregate of the six best subjects includes at least two of the four 'compulsory' subjects.) The aggregateswere converted into T-scores after being computed from the original scores.

Once again, the aggregate of VAL and RDL appears to be the most useful predictor. However it correlates equally with the two criteria, and in the case of other tests the difference is so obviously not significant that no test has been applied. It could therefore be

Table 4.9

Correlation between tests and aggregate in "compulsory" subjects (AGG) and aggregate in six best subjects (GEN. AGG)

r	AGG (4 subjects)	GEN. AGG
VAL	.60	.62
RDL	.65	.64
MEM	.24*	.21*
RTH	.30	.30
ALL	.62	.61
VAL + RDL	.69	.69
R		
VAL/RDL	.69	.69
ALL TESTS	.69	.69

* $p < .01$ Other coefficients: $p < .001$

deduced that requiring students to take certain basic subjects does not significantly affect their aggregate in 6 subjects.

The correlation between the two aggregates is in fact 0.88.

4.4 Hypothesis 4

That a cut-off point in the Scholastic Aptitude Test scores can be fixed below which a pupil will only have a 5% chance of success in WASC.

McClelland's (1942:106) method of fixing the pass-mark for the Qualifying Examination for entrance into secondary schools in Scotland was to find the mark giving the least number of "misfits" i.e. those who were above the pass-mark but failed the criterion tests, and those who were below but passed.

For our example, taking VAL + RDL as the best predictor, and 44 as the pass mark in General Aggregate in WASC, the minimum number of misfits would be 28 (using original scores).

Table 4.10

Number of misfits from each pass-mark in VAL + RDL

VAL + RDL	Misfits	Reject - Pass	Admit - Fail
04 or less	38	2	36
05 or less	31	4	27
06 or less	28	6	22
07 or less	30	14	16
08 or less	40	27	13

McClelland's pass-mark would have been fixed between 06 and 07 - ensuring that more pupils would be taken and would fail than would be rejected and subsequently pass (see Figure 4.4). However a pupil with 06 would have an almost 50% chance of passing.

An alternative method would be to use a graph of the percentage pass for each score in our sample (Figure 4.5), and assert that those with a score of 03 or less have less than a 5% chance of success.

Figure 4.5

Percentage of pupils in sample who passed at each level of VAL + RDL

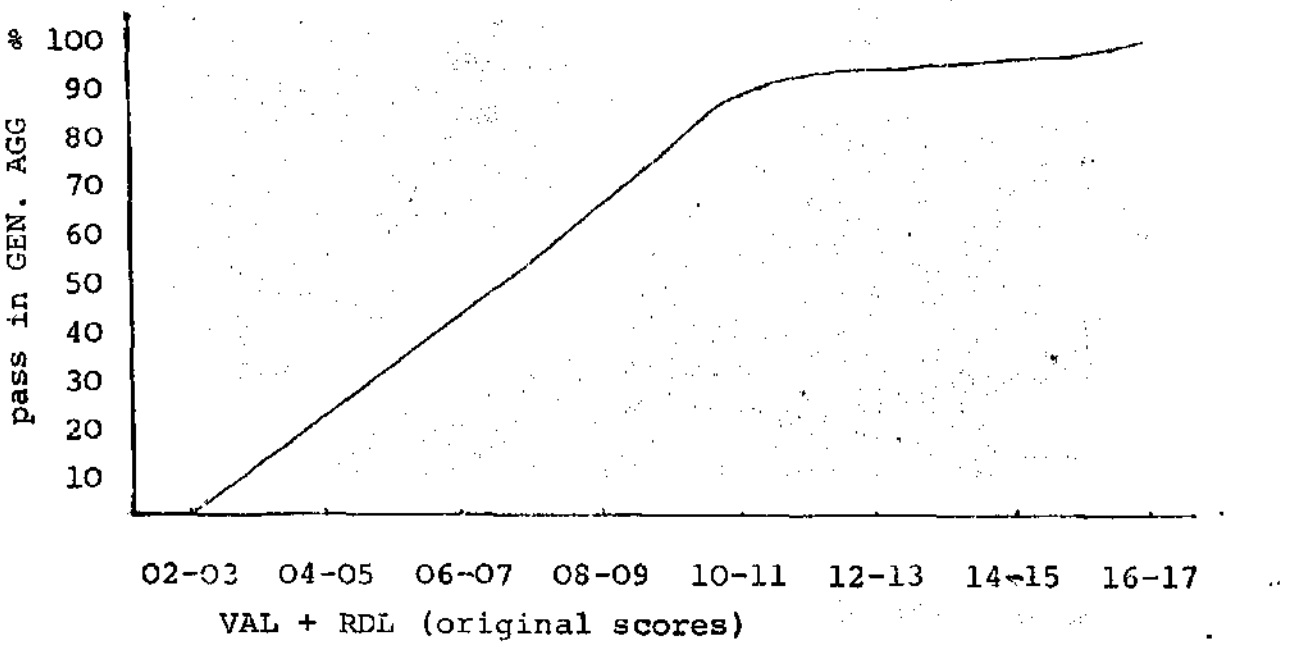


Figure 4.4

Application of McClelland's pass-mark to date of study

	18	Pass	Fail
	17	X	
	16	Admit-pass X	Admit-fail
	15	XXXXX	
VAL	14	XXXXXXXX	O
+	13	XXXXXXXXXXXXXXX	OO
RDL	12	XXXXXXXXXX	
	11	XXXXXXXXX	O
	10	XXXXXXXXXXX	OO
	09	XXXXXXXXXXX	OOOOOOO
	08	XXXXXXXXXXXXX	OOO
	07	XXXXXXXX	OOOOOO
	06	XX	OOOOO
	05	XX	OOOOOOOOO
	04	XX	OOOOOO
	03	Reject-pass	OOOOO Reject-fail
	02		OOOOO

Strictly speaking this only applies to this sample, but the relevance to the whole population can be shown by reference to particular features of the sample.

1. Most of the pupils who received scores as low as 02 and 03 in the two tests did not take the WASC examinations. (Only two out of twelve in Barewa College did.) The sample therefore reflects a higher level of ability than the whole population.
2. The sample is taken from well-established schools where the chances of passing are expected to be greater than in many other secondary schools in the state.
3. A higher percentage of pupils were awarded scores of 01 and 02 in the VAL and RDL tests than occurred in subsequent years, owing to the fact that the scores did not follow the normal distribution for Stanine scores. For instance, 26% had 01 or 02 in VAL instead of 11% of the candidates.

It is evident therefore that the limit of 03 is a generous one and that the probability of pupils with

such a score passing in the state as a whole is very slight indeed, and certainly less than the 5% set by the hypothesis.

It is of interest to apply the regression equation 4.1 predicting GEN. AGG from the two tests, VAL and RDL, though this is not usual practice in fixing cut-off points. Using Table 4.5 on page 64 there are two possible combinations of X_1 and X_2 to give an aggregate of 03 in original scores. The predicted T-scores for $\hat{Y}'$ would therefore be 63.39 or 63.38. The 95% Confidence Interval of ± 13.84 gives a range of 49.5 to 77.5, which, using the graph (Figure 4.3) gives approximate original scores of 41 to 54. This range includes several scores considered in this study as pass (though in actual practice it is not uncommon to fail with such an aggregate score). This arises from the nature of the standard error of prediction which is the same for all levels of Y, not taking into account the characteristics of the distribution. The scattergram (Figure 4.1B) shows a certain lack of homogeneity in variance at each end of the distribution of arrays, so the average error of estimate is in fact too small at the centre and

too great at the limits. The range of predicted values of Y for the lowest scores should therefore be much more restricted and it is certainly likely that a true 95% confidence interval would reinforce our selection of the cut-off point if we had access to the statistical techniques to take account of this.

In conclusion, it must be stated that, although it is reasonable to accept 03 as a cut-off point in the aggregate score of VAL and RDL, it is not possible to demonstrate beyond all measure of doubt that there would be less than a 5% chance of success at this level in the whole population.

4.5 Hypothesis 5

That students who take Science subjects score higher grades in WASC than students who take Arts subjects, when the effect of differences in Aptitude test scores has been controlled.

Guidance teachers are familiar with pupils seeking to change from Science courses to Arts courses because it is thought that there is a greater chance of success in Arts subjects.

Details of the distribution of scores in VAL + RDL and GEN. AGG for both Science and Arts students were tabulated (Table 4.11) and used in the computation of the necessary statistics to effect an analysis of covariance (McNemar 1969:421).

Table 4.11

Comparative statistics for Arts and Science students

	N	VAL + RDL		GEN.AGG		r_{xy}
		M_x	$S.D._x$	M_y	$S.D._y$	
Arts Students	55	90.6	16.88	54.6	9.22	- .613
Science Students	95	105.5	14.93	47.3	9.14	- .661
All	150	100.1	17.20	50.0	9.81	- .694

The F-value of 2.00 is not significant at the level required. It therefore cannot be concluded that there are significant differences in WASC results between the two groups over and above those which would be expected because of the differences in test scores. It cannot be substantiated that Arts subjects make the easier option.

Table 4.12

Analysis of Variance for Y Variable of Table 4.11 by covariance adjustments for uncontrolled X

	Total	Within Groups	Between Groups
Sum of products: Σxy	- 17455	- 13625	
Sum of squares: Σy^2	14342	12440	
Sum of squares: Σx^2	44081	36330	
Adjusted Σy^2	7430 -	7330	= 100
df	148	147	1
Variance estimate:		49.86	100

F-ratio = $\frac{100}{49.86}$ = 2.00 (df 147 and 1)

The mean scores do indicate however that the above average students choose Science and the below average Arts, and this dichotomy is reflected in the WASC results.

In conclusion, the null hypothesis that there is no significant difference between the WASC results for students who take Science and those who take Arts cannot

be rejected.

The findings of this study and their implications will be summarised in the following chapter.

Chapter 5

SUMMARY AND CONCLUSIONS

It is intended in this chapter to give a synopsis of the research project just reported, suggest interpretations of some of the results and propose topics for further investigation.

5.1 Summary

The study was undertaken with the aim of showing how the continued administration of the I-D Scholastic Aptitude Test Battery to Form 2 pupils in Kaduna State could be useful in predicting results in the West African School Certificate, in helping in the choice of subjects and even in eliminating a small number of pupils of insufficient ability. In addition, it became a validity study of the tests, as the previous validity studies had not used the WASC examination as their criteria.

As a background to the study, the development of testing in Africa for occupational and educational selection was considered as well as the relative merits of different types of testing and evaluation used for prediction of success in schools in the U.K. and U.S.A. Apart from the considerable contribution of assessments of attainment (through school examinations and teachers' estimates) and the less tangible effects of personality and environment, the most useful predictors of success appeared to be the standardised tests of intelligence and tests of English.

Hypotheses about the predictive value of the four I-D scholastic tests (Verbal Analogies, Reading Comprehension, Memory and Arithemtic) in estimating grades in WASC were formulated and tested, using a sample of 150 pupils taken from three secondary schools in Zaria where the I-D tests had been administered in 1973 and the WASC examinations taken by the same students in 1976.

A significant relationship was found to exist between the Scholastic Test Battery and all the criterion scores used. With English Language, the aggregate of the six best subjects give multiple correlation coefficients

reached 0.69 to 0.70, while for Mathematics the coefficient was 0.59. The validity coefficient quoted in the I-D Technical Manual is 0.56 when school grades and assessment concurrent with the administration of the tests were used. Taking the tests individually, the Memory Test (MEM) did not correlate once with the criteria at the level of significance expected in this study (viz. $p < .001$), nor with any other test in the Battery, while Arithmetic, though correlating significantly with all criteria, made no significant contribution to the multiple correlations, not even when predicting grades in Mathematics. Presumably the Arithmetic test overlaps with the other two tests in what it measures - it correlated significantly with Reading Comprehension (RDL) (0.37) and with Verbal Analogies (VAL) (0.43). These two tests, which correlated significantly with each other (0.64) did not show any significant difference when correlated with the criterion scores. There was a slight increase in the coefficient when the aggregate of the two equally-weighted tests was taken, and multiple correlation using the two tests gave the same coefficient, for the Beta weights were approximately 8 and 7. The most

useful predictor for WASC grades and the most convenient is therefore the aggregate of scores in the low forms of the I-D Verbal Analogies and Reading Comprehension Tests.

Regression equations, based on normalised T-scores for both tests and criteria in order to satisfy the demands for homogeneity of variance and normality of distribution, were computed, enabling the prediction of scores in WASC. There was no significant difference in the performance of the students who chose Arts rather than Science when the WASC scores were adjusted to take into account the difference in the Test scores for the two groups.

Consideration was given to the possibility of establishing a cut-off point in the test scores below which pupils could be expected to have almost no chance of success. An aggregate score of 03 in VAL and RDL was shown to be an effective cut-off point, but its relevance for the whole population was not proved conclusively.

5.2 Generalisations

It is difficult to make inferences about the

population from a correlation study such as this. As Guilford points out (1965:105):

> A correlation is always relative to the situation under which it is obtained, and its size does not represent any absolute natural fact.

In this study detailed attention has been given to the sample, which, if not truly representative of the population, should give results which are less significant than might be expected in the population.

Because of this principle of erring on the 'safe' side, it is not unreasonable to draw some general conclusions from the results of this study.

(a) VAL and RDL are both good predictors of results in WASC and are more efficient than the I-D Test Manual suggests (over 0.60 rather than 0.45).

(b) VAL, RDL and to a lesser extent RTH (the Arithmetic Test) overlap in what they measure. VAL and RTH may reflect the "intelligence" of the candidate, while RDL shows the command of English without which the other tests could not be attempted. The effort, however, to avoid discriminating in the English test against those with disadvantaged educational

backgrounds, may have made RDL depend more on "intelligence" than on comprehension of English. RTH made no significant contribution in multiple correlation, although, by itself, it correlated significantly with the criteria.

(c) VAL and RDL could be used as screening tests to ascertain pupils who have little chance of success in academic subjects in WASC.

(d) The Memory Test (MEM) does not correlate significantly with the criterion, and its validity as a test in the Scholastic Battery must be put in doubt.

(e) The mean scores in all four tests were lower than those of the norming group which was tested in 1963. The difference was not pronounced in VAL (the "intelligence" test) but it was evident in RDL where 56% of the pupils in the sample had a score of 4 or under as opposed to 40% in the norming group. In RTH, 80% of the candidates were below average when compared with the norming group. Though this

study has not attempted to analyse what is measured by the tests, it is tempting to conclude that the pupils in secondary schools in this decade are not reaching the levels of competence in English and Arithmetic of ten years earlier.

(f) As far as the prediction of results in WASC is concerned, it must be emphasised that the regression coefficient will vary from school to school and from year to year. It is therefore highly desirable for each institution to develop its own expectancy tables.

To conclude, it is necessary to record that the findings and the generalisations mentioned above are only valid within the parameters of this research. It is not claimed that the Aptitude Test scores alone are adequate as predictors of WASC grades, as personality, environmental and other factors may be equally effective in prediction. Further, studies quoted in this research have been mentioned on account of their use of intelligence in predicting achievement, without consideration of the hierarchical importance of intelligence as one of several factors, as this was not within the specific objectives and parameters of the present study.

5.3 Recommendations

Arising from the conclusions of this study, certain recommendations could be tendered to the Ministry of Education in Kaduna State and to the Test Development and Research Division of the West African Examinations Council concerning the future use of the Aptitude Tests.

(a) Two of the tests used in the Scholastic Aptitude test battery are very efficient predictors of grades in the West African School Certificate. They permit an objective assessment of the pupils' ability as they are administered under the direct supervision of the TDRD and can therefore be used to draw attention to individual pupils who do not have the ability to succeed in an academic course and whose presence in the classroom acts as a discouragement to both teachers and classmates, with a resultant lowering of standards attained. The tests also allow objective comparisons to be made at school, state or national levels, and from year to year. In addition they make available scores for each pupil in the state which provide an excellent starting point for educational research, a useful alternative to age which is

the most common point de départ overseas. The administration of the I-D testing programme offers distinct advantage for the educational progress of the State and should be continued.

(b) More use of the test scores should be made at the school level, with assistance being provided in undertaking the necessary statistical analyses. The analysis of covariance, for example, is a useful statistical tool for comparing results between classes which have been subjected to different conditions.

(c) The relevance of the Memory Test has been questioned in this study, and unless its inclusion in the Battery can be justified, its continued administration seems unnecessary.

(d) The inclusion of the Arithmetic test should also be justified, as it does not appear to measure any ability not measured by the Verbal Analogies Test.

(e) Consideration should be given to the low scores obtained in the Arithmetic Test. According to Schwarz and Krug (1972:135) it should be regarded as a "fairly broad spectrum test that uses the vehicle of computation problems to approximate

the examinees' overall aptitude in the quantitative domain". If this is so, then there is considerable deterioration in the standard of numeracy of our pupils, which will require correction at the primary school level.

5.4 Suggestions for further study

The availability of the I-D test scores for all the Form 2 pupils in the State offers many opportunities for further study.

(a) There are seven other tests at present being administered in the Kaduna State Testing Programme. Their usefulness should also be investigated, as it is possible that VAL and RDL may be the only significant predictors in the batteries designed to show aptitude for Commercial and Technical courses.

(b) The aptitude tests studies a count for less than 45% of the variance in the WASC results. The inclusion in multiple correlation of other independent variables - personality and environmental factors, for example, might help to draw attention to relevant factors influencing academic

success. (Lavin, 1965).

(c) The possibility that a standardised attainment test of English might be a better predictor of WASC grades could also be considered.

(d) The effect of repeating Form 2 and taking the tests a second time could make a contribution to the present debate on "automatic promotion".

(e) Further appraisal of what is actually measured by the I-D tests and the WASC examinations would be relevant to the development of an improved testing programme.

(f) This study has concentrated primarily on aggregate scores in WASC as criteria. More consideration needs to be given to predicting grades in individual subjects in order to guide the individual student into the course most suited to his ability.

In fact, much more remains to be investigated than has already been investigated, and in concluding the report of this piece of research, the author is aware that the work has not been concluded, but has only just begun.

References

Adams, G. S. 1964. Measurement and Evaluation in Education Psychology and Guidance. Holt Rinehart and Winston. New York. 654p.

Akeju, A. 1966. Report on Validity of I-D Aptitude Tests for predicting academic success in Nigerian School of Science. WAEC. Lagos. 8p.

Aleyedeino, S. C. 1973. The nature, place and influence of Tests and Measurement Procedures in Education. West African Journal of Educational and Vocational Measurement. 1:1-13.

Astin, A. W. 1971. Predicting Academic Performance in College. Selectivity Data for 2300 American Colleges. Free Press. New York. 299p.

Bloom, B. S. and F. R. Peters. 1961. The Use of Academic Prediction Scales for Counseling and selecting College entrants. Free Press of Glencoe. New York. 146p.

Bulley, H. C. A. 1964. The Progressive Matrices as an Instrument for Determining Educability in Ghana. Quoted from Napier (1972).

Butcher, H. J. 1966. Sampling in Educational Research. Manchester U.P. 30p.

Butcher, H. J., M. Ainsworth, and J. E. Nesbitt. 1963. Personality factors and school achievement. A Comparison of British and American Children. British Journal of Educational Psychology. 33:276-85.

Cattell, R. B. and H. J. Butcher. 1968. The Prediction of Achievement and Creativity. Bobbs Morrell Co. New York. 386p.

Cattell, R. B., A. P. Fealy, A. B. Sweney. 1966. What

can personality and motivation source trait measurements add to the prediction of school achievement? British Journal of Educational Psychology. 36:280-95.

Choppin, B. H. L. et al. 1973. The Prediction of Academic Success. NFER. London. 70p.

Clarke, J. W. 1948, 1949. Performance Tests of Intelligence for Africa. Overseas Education. 19:777-789; 20:822-826.

Cronbach, L. J. 1961. Essentials of Psychological Testing. Second Edition. Harper & Row. New York. 650p.

Douglas, J. W. B. and J. M. Ross. 1964. Age of Puberty. Journal of Child Psychology and Psychiatry. 5:185-196.

Durojaiye, M. O. A. (Ed.) 1972. Psychological Guidance of the Schoolchild. Evans Bros. Ltd. Ibadan. 128p.

Emmett, W. G. 1945. An Enquiry into the Prediction of Secondary School Success. University of London Press. 58p.

Emmett, W. G. and F. S. Wilmut. 1952. The prediction of school certificate performance in specific subjects. British Journal of Educational Psychology. 22:52-56.

Gulliksen, H. and S. S. Wilks. March 1950. Regression Tests for several samples. Psychometrika. 15:91-114.

Guildford, J. P. 1965. Fundamental statistics in psychology and education. Fourth edition. McGraw Hill Book Company. New York. 605p.

Hills, John R. and J. A. Klock. November 1968. Predicting Grades for all schools in a large university. American Educational Research Journal. 2:3:145-150.

Hockey, S. W. 1968. Intelligence, Common Entrance and the GCE. British Journal of Educational Psychology.

38:140-147.

Irvine, S. H. 1963. Ability Testing in English-speaking Africa. An overview of comparative and predictive studies. Rhodes-Livingstone Journal. 34:44-45. Quoted in Ord (1972).

Irvine, S. H. 1964. Selection of Africans for Post-primary education in Southern Rhodesia. Bulletin of the Inter-African Labour Institute. 11:69-93. Quoted in Ord (1972).

Irvine, S. H. 1968. A five-year follow-up of secondary school selection procedures in Central Africa 1962-67. British Journal of Educational Psychology. 38:208-211.

Irvine, S. H. 1966. Towards a rationale for testing attainments and abilities in Africa. British Journal of Educational Psychology. 36:24-32.

Lavin, D. E. 1965. The Prediction of Academic Performance. A Theoretical Analysis and Review of Research. Russel Sage Foundation. New York. 182p.

Kline, P. 1966. Extraversion, Neuroticism and Academic Performance among Ghanaian University students. British Journal of Educational Psychology. 36:91-94.

Klugh, H. E. and R. Brierley. 1959. The School and College Ability Test and High School grades as predictors of College achievement. Educational and Psychological Measurement. 19:625-626.

Kure, D. H. and J. Bennett. 1974. Handbook for the use of Guidance Teachers. Ministry of Education. Kaduna. 60p. (mimeo)

McClelland, W. 1942. Selection for secondary education. University of London Press. 264p.

McIntosh, D. M. 1959. Educational Guidance and the Pool of Ability. University of London Press. 199p.

McNemar, Q. 1969. Psychological Statistics. Fourth Edition. John Wiley & Sons Inc. New York. 529p.

Majasan, J. K. 1968. The predictive ability of the I-D tests for selecting candidates into technical and commercial courses. WAEC. Lagos. 36p.

Minium, E. W. 1970. Statistical Reasoning in Psychology and Education. John Wiley & Sons. New York. 465p.

Ndahi, S. W. December 1975. A follow up study of the 1972-73 Placement of Students who took TEDRO Aptitude Tests in five secondary schools in Zaria. A.B.U. Zaria. Unpublished M.Ed. project.

Napier, R. W. 1972. School Guidance Services. Focus on the Developing Nations. Evans Bros. Ltd. Ibadan. 200p.

Naylor, F. D. 1972. Personality and Educational Achievement. John Wiley & Sons. Sydney. 162p.

N.I. Department of Education. 1977. Report of Working Party on Alternative Transfer Arrangement. H.M.S.O. Belfast. 12p.

Ohuche, R. O. and S. A. Akeju. 1975. Testing and Evaluation in Education. African Educational Resources. Monrovia. 145p.

Ord, I. G. 1972. Testing for education and occupation selection in developing countries - a review. Occupational Psychology. 46:3:123-182.

Peel, E. A. and D. Rutter. 1951. The predictive value of the entrance examination as judged by the school certificate examination. British Journal of Educational Psychology. 21:30-35.

Pilliner, A. E. G. 1968. In H. J. Butcher (Ed.) Educational Research in Britain. University of London Press. 408p.

Popham, W. J. and K. A. Sirotnik. 1967. Educational

Statistics. Use and Interpretaion. 2nd Edition. Harper and Row. New York. 413p.

Rushton, J. 1966. The relationship between personality characteristics and scholastic success in eleven year old children. British Journal of Educational Pyschology. 36:178-84.

Rutter, D. 1950. An enquiry into the predictive value of the Grammar School Entrance Examinations. Research Review. University of Durham. 1:3-11.

Schonell, F. J. 1949. The Development of Educational Research in Great Britain. Part IV. British Journal of Educational Psychology. 19:160-175.

Schwarz, P. A. 1963. Adapting Tests to the Cultural setting. Educational Psychological Measurement. 23: 673-686.

Schwarz, P. A. 1964a. I-D Examiners' Manual/Manuel de l'examinateur. American Institutes for Research. Pittsburgh.

Schwarz, P. A. 1964b. Development of Manpower screening tests for developing nations: a technical manual and report submitted to the U.S.A. I.D. American Institutes for Research. Pittsburgh.

Schwarz, P. A. and Krug, R. E. 1972. Ability Testing in Developing Countries. A Handbook of Principles and Techniques. Praeger Publishers. New York and London. 264p.

Sharp, H. C. and L. M. Pickett. 1959. The GATB as a Predictor of College Success. Education and Psychology Measurement 19:617-623.

Snider, J. G. 1972. Aptitude Tests for West Africa. West African Journal of Education. 16:2:171-177.

Stubbins, T. E. 1940. The prognostic Values of the subjects of a secondary school entrance examination.

British Journal of Educational Psychology. 10:16-24.

Summer, R. and F. W. Warburton. 1972. Achievement in Secondary School. Attitude, Personality and School Success. N.F.E.R. London. 208p.

Taylor, A. (Ed.) 1962. Educational and Occupational selection in West Africa. O.U.P. London. 219p.

Thorndike, R. L. 1949. Personnel Selection. Test and Measurement Techniques. John Wiley & Sons. New York, Chapman & Hall. London. 358p.

Williams, I. C. and N. C. Boreham. 1972. The predictive value of CSE grades for further education. Evans/ Methuen. London. 88p.

Wiseman, S. 1966. Correlation Methods. Manchester U.P. 44p.

Wrigley, J. 1955. The relative efficiency of intelligence and attainment test predictors of success in grammar schools. British Journal of Educational Psychology. 25:107-116.

Yates, A. and D. A. Pidgeon. 1957. Admission to Grammar schools. Newnes. London. 260p.

Yoloye, E. A. 1973. The predictive value of Lorge-Thorndike Intelligence Tests for Achievement in Nigerian Grammar Schools. West African Journal of Educational and Vocational Measurement. 1:48-51.