SUMMARY OF THESIS

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the University of Glasgow, 1957.

INTELLIGENCE AND SOCIAL CLASS: AN INVESTIGATION OF THE INTELLIGENCE TEST PERFORMANCE OF CHILDREN OF DIFFERENT SOCIAL GROUPS

by John Macdonald, M.A., Ed.B.

The problem of social differences in intelligence test performance has been much discussed over the past forty years; there is still, however, little agreement among psychologists as to their interpretation. The first part of this thesis, which incorporates a summary of previous researches on the problem, also makes some attempt to explain the strong differences in opinion which have often arisen. It is perhaps unfortunate that no reasonably satisfactory explanation has yet been suggested, in view of the widespread use of intelligence tests for the purposes of selection.

An examination of previous studies makes it clear that the relationship of IQ or test score to socio-economic level, social class, or occupational level is one of the best documented facts in the whole field of mental testing. Since the early experiments of Binet and his colleagues/

/colleagues there has been a consistent finding on the part of investigators that subjects in good social circumstances tend to score more highly than subjects in poor social circumstances. Some psychologists have explained this superiority of the upper social classes in terms of assumed hereditary and innate differences, while others have pointed to the known differences in environment. Recently a few American investigators have claimed that differences between social groups in test performance are due to the inclusion in tests of a disproportionate number of problems with which working-class subjects have had no opportunity to become familiar.

Sufficiently detailed analysis either of tests or of the social environment of subjects is still relatively lacking; it may therefore be said that where psychologists have come to definite conclusions as to the meaning of social differences in test performance they have done so on the basis of unsatisfactory evidence. It is hoped that the present investigation, the first comprehensive item-analysis study of the problem to be carried out by a British psychologist, provides some new facts against which to assess possibilities.

The chief findings of the study, an analysis of the performance on a group intelligence test of 1000 children belonging to five social classes characterized as upper-middle-class, middle-middle-class, lower-middle-class, upper-working-class, and lower-working-class, may be summarized briefly as/

/as follows.

1. When the mean test scores of the five social class groups were computed, it was found that these declined consistently with social level.

2. When the incorrect responses of children to test items were analysed, it was noticed that as social level declined a steadily increasing proportion of these incorrect responses was due to subjects omitting or failing to reach items, and a steadily decreasing proportion due to their answering items wrongly.

3. When the performance of children on individual items in the test was examined, it was evident that on the great majority of items the performance of subjects in the upper social classes tended to be superior to that of subjects in the lower social classes. Nevertheless items varied markedly in the extent of superiority of performance shown, and there were a few cases where lower social groups were superior, though the difference in their favour was nowhere statistically significant.

4. When the relative tendency of items to exhibit large or small differences in performance in one comparison of social groups was compared with this tendency in another comparison, it was found that in general no close association existed, the individual discrimination of items between groups being clearly dependent to some extent/

/extent on the nature of the groups being compared.

5. When items were classified into three categories by the nature of the knowledge or previous experience which they demanded from subjects, no large differences in discrimination between the children of the different social classes were established, although there seemed to be a slight tendency for items dependent on special knowledge or experience to discriminate more than items dependent on knowledge gained in school.

6. When items were classified into two categories according to the degree of abstract thought which they were considered to require, no clear tendency was found for "abstract" items to discriminate more than "concrete" items between the children of the different social classes.

7. When items were classified according to the type of symbolism which they employed, it was found that in general "verbal" items did not discriminate more than "non-verbal" items.

8. When items were classified according to the type of question which they asked, it was found that some types of question showed significantly greater discrimination than others.

9. When a small number of items were isolated which showed significant differences in wrong answer pattern between children in the upper-middle-class and children in the/

/the lower-working-class, evidence was found that the lower-working-class children appeared to answer items in a more random manner; but it also seemed that children found answers differentially attractive, according to the social class to which they belonged.

10. When the performance on individual items of two groups matched for test score and drawn respectively from the upper-middle-class and the lower-working-class was compared, no sound evidence of significant differences could be established.

INTELLIGENCE AND SOCIAL CLASS

AN INVESTIGATION OF THE INTELLIGENCE TEST PERFORMANCE OF CHILDREN OF DIFFERENT SOCIAL GROUPS

by

John Macdonald, M.A., Ed.B.

Submitted in fulfilment of the
requirements for the degree of
Doctor of Philosophy
University of Glasgow
1957

TABLE OF CONTENTS.

Chapter.		Page.
I.	INTRODUCTION...................	1
	PART I. A CRITICAL EVALUATION OF PREVIOUS RESEARCH ON THE RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT.	6
II.	SOME REASONS FOR THE EXISTENCE OF OPPOSITION TO THE PRACTICE OF INTELLIGENCE TESTING.......	7
III.	THE SOCIAL BACKGROUND OF DIFFERENCES OF OPINION AMONG SOCIAL SCIENTISTS.............	15
IV.	CONCEPTIONS AND DEFINITIONS OF SOCIAL CLASS...............	22
V.	CONCEPTIONS AND DEFINITIONS OF INTELLIGENCE................	32
VI.	EARLY RESEARCHES ON THE DEGREE OF RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT........	39

Chapter.		Page.
VII.	LATER RESEARCHES ON THE DEGREE OF RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT	59
(i)	Researches on pre-school children	62
(ii)	Researches on children of school age	72
(iii)	Researches on students at establish--ments of higher education	94
(iv)	Researches on adults	99
(v)	Researches on intellectual deviates	
	(a) superior deviates	105
	(b) inferior devaites	110
VIII.	STUDIES OF THE RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND INDIVIDUAL FACTORS OF THE SOCIAL ENVIRONMENT	115
IX.	ITEM-ANALYSIS STUDIES OF THE RELATION--SHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT	133
X.	THE SOCIAL AND EDUCATIONAL RELEVANCE OF RESEARCH ON THE RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT	174

Chapter. Page.

PART II.

Chapter		Page
	AN INVESTIGATION OF THE INTELLIGENCE TEST PERFORMANCE OF 1000 CHILDREN OF FIVE DIFFERENT SOCIAL GROUPS	181
XI.	PROBLEMS AND DIFFICULTIES IN SELECTION OF SUBJECTS	182
XII.	SELECTION OF SUBJECTS AND SECURING OF TEST MATERIAL	193
XIII.	DESCRIPTION OF TEST	199
XIV.	SOCIAL CLASS DIFFERENCES IN GENERAL TEST PERFORMANCE	206
XV.	VARIATION IN TEST SCORE WITH SOCIAL CLASS, LOCATION OF HOME, AND SIZE OF SCHOOL	219
XVI.	ERRORS IN TEST MARKING	232
XVII.	METHODS USED IN ITEM - ANALYSIS : SOCIAL CLASS DIFFERENCES IN PERFORMANCE ON INDIVIDUAL ITEMS	241
XVIII.	FAILURE TO REACH ITEMS : SOCIAL CLASS DIFFERENCES	258
XIX.	OMISSION OF ITEMS : SOCIAL CLASS DIFFERENCES	273
XX.	SIZE AND STATISTICAL SIGNIFICANCE OF SOCIAL CLASS DIFFERENCES FOR INDIVIDUAL ITEMS	294
XXI.	SPECIFIC AND GENERAL DISCRIMINATION OF ITEMS	317
XXII.	ITEMS SHOWING SMALL SOCIAL CLASS DIFFERENCES	339

Chapter.		Page.
XXIII.	ITEMS SHOWING LARGE SOCIAL CLASS DIFFERENCES........................	370
XXIV.	SOCIAL CLASS DIFFERENCE IN RELATION TO THE FORMAL STRUCTURE OF ITEMS....	381
XXV.	SOCIAL CLASS DIFFERENCES IN RELATION TO THE KIND OF KNOWLEDGE OR EXPERIENCE IMPLIED BY ITEMS.........	405
XXVI.	SOCIAL CLASS DIFFERENCES IN RELATION TO THE DEGREE OF ABSTRACTION INVOLVED IN ITEMS..................	412
XXVII.	SOCIAL CLASS DIFFERENCES IN RELATION TO THE TYPE OF SYMBOLISM EMPLOYED BY ITEMS..........................	420
XXVIII.	SOCIAL CLASS DIFFERENCES IN RELATION TO THE TYPE OF QUESTION ASKED BY ITEMS..............................	426
XXIX.	SOCIAL CLASS DIFFERENCES IN WRONG ANSWER PATTERN.....................	438
XXX.	SOCIAL CLASS DIFFERENCES IN PERFORMANCE ON INDIVIDUAL ITEMS BY TWO GROUPS MATCHED FOR TEST SCORE...	455
XXXI	SUMMARY AND CONCLUSIONS.............	459
	BIBLIOGRAPHY........................	470
	APPENDIXES.	
	APPENDIX I. FACSIMILE OF TEST EMPLOYED IN INVESTIGATION...........	487
	APPENDIX II. DETAILS OF VERY DIFFICULT, VERY EASY AND "UNREACHED" ITEMS.............................	496

LIST OF TABLES.

Table.		Page.
I	Mean I.Q.'s According to Father's Occupation.. (Terman and Merrill, 1937)........................	68
II.	Ordering of 23 Sims Items by Degree of Association with Intelligence Test Performance (Cuff, 1934).................................	124
III.	Distribution of Thirty-Six-Day Sample by Occupational Class.............................	196
IV.	Distribution of Correct Responses for Multiple-Choice Items by Positional Order.....	203
V.	Distribution of Test Score by Social Class....	207
VI.	Mean Test Score by Occupational Class and Social Class...................................	208
VII.	Mean Test Score and Standard Deviation of Test Score by Social Class....................	210
VIII.	Level of Significance of Differences between Standard Deviations of Test Score.............	215
IX.	Distribution of All Subjects by Location of Home..	221
X.	Distribution of All Subjects by Size of School Attended.................................	221
XI.	Mean Test Score by Location of Home...........	222
XII.	Mean Test Score by Size of School Attended....	222
XIII.	Mean Test Score by Social Class by Location of Home..	225

Table		Page
XIV.	Mean Test Score by Social Class by Size of School Attended..................	228
XV.	Mean Test Score by Location of Home by Size of School Attended..................	230
XVI.	Distribution of Marking Errors by Social Class..........................	236
XVII.	Corrected and Uncorrected Mean Test Scores and Standard Deviations by Social Class..........................	239
XVIII.	Davis' "Difficulty Index" for conversion of Percentages to Normalized Scores......	245
XIX.	Distribution of Test Items by Difficulty of Item for Five Social Class Groups.....	251
XX.	Proportion of Items Passed by 50 per cent or More of Subjects in Five Social Class Groups..............................	253
XXI.	Mean Item Difficulty and Standard Deviation of Item Difficulty by Social Class..................................	256
XXII.	Distribution of "Unreached Responses" by Social Class.........................	259
XXIII.	Contribution of "Unreached Responses" to Total Incorrect Responses for Five Social Class Groups......................	267
XXIV.	Contribution of "Unreached Responses" to Differences in Mean Test Score betweem Five Social Class Groups (1).............	269
XXV.	Contribution of "Unreached Responses" to Differences in Mean Test Score between Five Social Class Groups (2).....	270

Table		Page
XXVI.	Distribution of Omissions by Social Class..............................	274
XXVII.	Contribution of Omissions to Total Incorrect Responses for Five Social Class Groups.......................	284
XXVIII.	Contribution of Omissions to Differences in Mean Test Score between Five Social Class Groups(I).	285
XXIX.	Contribution of Omissions to Differences in Mean Test Score between Five Social Class Groups(2).	286
XXX.	Contribution of Wrong Answers, Omissions, and "Unreached Responses" to Differences in Mean Test Score between Five Social Class Groups (I)...................	287
XXXI.	Contribution of Wrong Answers, Omissions, and "Unreached Responses" to Differences in Mean Test Score between Five Social Class Groups (2)...................	288
XXXII.	Mean number of Wrong Answers, Omissions, and "Unreached Responses" for Five Social Class Groups..........................	289
XXXIII(I)	Distribution of Test Items by the Level of Significance of Differences between Class A and Class B Responses.....................	296
XXXIII(2)	Distribution of Test Items by Size of Index Differences between Class A and Class B Responses......	296
XXXIV(I)	Distribution of Test Items by the Level of Significance of Differences between Class B and Class C Responses..................	297

Table.		Page.
XXXIV(2)	Distribution of Test Items by Size of Index Differences between Class B and Class C Responses.......	297
XXXV(I)	Distribution of Test Items by the Level of Significance of Differences between Class C and Class D Responses...........................	298
XXXV (2)	Distribution of Test Items by Size of Index Differences between Class C and Class D Responses.......	298
XXXVI.(1)	Distribution of Test Items by the Level of Significance of Differences between Class D and Class E Responses...........................	299
XXXVI (2)	Distribution of Test Items by Size of Index Differences between Class D and Class E Responses.............	299
XXXVII(1)	Distribution of Test Items by the Level of Significance of Differences between Class A and Class C Responses...........................	300
XXXVII(2)	Distribution of Test Items by Size of Index Differences between Class A and Class C Responses.............	300
XXXVIII(1)	Distribution of Test Items by the Level of Significance of Differences between Class B and Class D Responses...........................	301
XXXVIII(2)	Distribution of Test Items by Size of Index Differences between Class B and Class D Responses.............	301

Table		Page
XXXIX(1)	Distribution of Test Items by the Level of Significance of Differences between Class C and Class E Responses.	302
XXXIX(2)	Distribution of Test Items by Size of Index Differences between Class C and Class E Responses.................	303
XL (1)	Distribution of Test Items by the Level of Significance of Differences between Class A and Class D Responses.	304
XL (2)	Distribution of Test Items by Size of Index Differences between Class A and Class D Responses......................	305
XLI (1)	Distribution of Test Items by the Level of Significance of Differences between Class B and Class E Responses.	306
XLI (2)	Distribution of Test Items by Size of Index Differences between Class B and Class E Responses.................	307
XLII(1)	Distribution of Test Items by the Level of Significance of Differences betwenn Class A and Class E Responses.	308
XLII(2)	Distribution of Test Items by Size of Index Differences between Class A and Class E Responses.................	309
XLIII	Items Showing Significant Differences in Response in Comparisons of Contiguous Social Class Groups........	318

Table.		Page.
XLIV.	Distribution of Test Items by the Level of Significance of Differences in Response for Comparisons of Non-contiguous Social Class Groups..	323
XLV.	Distribution of Significant Item Discriminations Among Comparisons of Social Class Groups...............	325
XLVI.	Correlations between Item Discrimin--ations for Four Comparisons of Social Class Groups.................	328
XLVII.	Mean Index Differences for other Comparisons of Specifically Discriminating and Generally Discriminating Items in Comparison A-B................................	332
XLVIII.	Mean Index Differences for Other Comparisons of Specifically Discrim--inating and Generally Discriminating Items in Comparison B-C...............	332
XLIX.	Mean Index Differences for other Comparisons of Specifically Discriminating and Generally Discriminating Items in Comparison C-D.................................	333
L.	Mean Index Differences for other Comparisons of Specifically Discriminating and Generally Discriminating Items in Comparison D-E.................................	333
LI.	Mean Index Differences for 16 Items Showing Small Tendency to Discriminate between Social Class Groups...........	341
LII.	Mean Index Differences for 6 Items showing Large Tendency to Discriminate between Social Class Groups.............................	371

Table.		Page.
LIII.	Mean Index Differences for 40 Multiple-Choice Items............	382
LIV.	Mean Index Differences for 39 Multiple-Choice Items by Number of Choices Provided..............	384
LV.	Mean Index Differences for 30 Five-Choice Items by Type of Question Asked....................	386
LVI.	Mean Index Differences for 20 Five-Choice Items by Type of Question Asked....................	388
LVII.	Mean Index Differences for 30 Five-Choice Items by Positioning of Correct Response..............	391
LVIII.	Mean Index Differences for 11 other Items offering Choice of Response.........................	396
LIX.	Mean Index Differences for 11 Items offering Choice of Response by Type of Question Asked........	397
LX.	Mean Index Differences for All Items by Page of Test.............	399
LXI.	Mean Index Differences for 70 Items by Nature of Experience or Knowledge Implied.............	409

Table.		Page
LXII	Mean Index Differences for All Items by Degree of Abstraction Involved.....	416
LXIII	Mean Index Differences for All Items by Type of Symbolism Employed.........	422
LXIV	Mean Index Differences for All Items: Verbal/Non-Verbal Symbolism...........	423
LXV	Categorization of All Items by Type of Question Asked....................	428
LXVI	Mean Index Differences for All Items by Type of Question Asked.............	430
LXVII	Level of Significance of Differences between Mean Index Differences for Items Categorized by Type of Question asked..........................	435
LXVIII	Social Class Differences in Wrong Answer Pattern on Item 8..............	442
LXIX	Social Class Differences in Wrong Answer Pattern on Item 20.............	445
LXX	Social Class Differences in Wrong Answer Pattern on Item 34.............	446
LXXI	Social Class Differences in Wrong Answer Pattern on Item 53.............	448
LXXII	Social Class Differences in Wrong Answer Pattern on Item 56(a).........	450
LXXIII	Social Class Differences in Wrong Answer Pattern on Item 64.............	451
LXXIV	Obtained and Expected Distribution of Values of Chi-Square in Comparison of Performance of Two Matched Groups..	457

LIST OF APPENDIX TABLES

Appendix Table		Page.
I.	Items Correctly Answered by More than 92 Per Cent of Subjects......	498
II.	Items Correctly Answered by Less than 8 Per Cent of Subjects.......	499
III.	Items Not Reached by More than 5 Per Cent of Subjects............	500

LIST OF FIGURES.

Figure.		Facing Page.
1.	Frequency Polygons Showing Distribution of Test Score for Five Social Class Groups..................................	217
2.	Distribution of Incorrect Responses for Social Class Groups......................	290
3.	Distribution of Test Items by Level of Statistical Significance of Differences between Performance of Social Class Groups................................	310
4.	Distribution of Test Items by Size of Index Differences between Social Class Groups................................	315
5.	Variations in Performance by Social Classes on Items, I, 22,27,64,65. (Item 33 presented for purposes of comparison)...........................	342
6.	Variations in Performance by Social Classes on Items 22,23,24,25...........	346
7.	Variations in Performance by Social Classes on Items 1,2,4,6,13,14,16,18, 22,27,31,44(c),45,60,64,65.............	348
8.	Variations in Performance by Social Classes on Items 6,7,8,12,20,39........	351
9.	Variations in Performance by Social Classes on Items 60,61,62,63,64,65.....	362
10.	Variations in Performance by Social Classes on Items 30,56(c)...............	364
11.	Variations in Performance by Social Classes on Items 32,33,41,42,59,66(b)..	372

PREFACE.

The interpretation of social differences in intelligence test performance is still uncertain, in spite of considerable investigation. The author of this study first became interested in the subject in 1950, partly as a result of his own experience as a university student of working-class origin. The present thesis details an investigation begun in 1954; it should not be thought that the interest of the author ends with its completion.

It is hoped that the following pages, presenting as they do a critical evaluation of previous experimental study and an original inquiry into differences in intelligence test performance between children/

/children of different social classes, will fill a gap in British psychology, the literature of which lacks a detailed investigation of social differences in intelligence. The relevance of the investigation to educational practice, particularly with reference to selection for secondary education, has been kept in mind throughout.

Acknowledgement has been made of all printed sources consulted. The author wishes to state his indebtedness to the Carnegie Trust for the Universities of Scotland, a research grant from whom made the study possible. Thanks are also due to the Mental Survey Committtee of the Scottish Council for Research in Education, who willingly made available the experimental material on which the thesis is based; and to the officers of the Council for their invaluable help and assistance. In addition, the author is greatly obliged to Professor Stanley Nisbet for his sympathetic encouragement at all stages of the work; to Dr. R.A. Robb for his careful checking of statistical/

/statistical method; to Mrs. Mary Mackintosh for her many helpful suggestions; and to Mr. Robert Vallance for his assistance in the compilation of item-analysis charts. Finally, he must record his gratitude to his wife, whose loyal support carried him over many difficulties.

CHAPTER I.

INTRODUCTION.

Intelligence tests have in the past been attacked and defended by a variety of people for a variety of reasons. Differences of opinion still exist as to their value. In practice, however, testing is accepted as an irreplaceable method of assessing individual ability.

It is no part of the task of the present investigator to consider past controversies in any detail. It will be found that discussion in succeeding chapters refers closely to one particular problem, that of the differences in test performance between subjects of different social backgrounds. Since the early experiments of Binet there has been a consistent finding on the part of investigators that subjects in good social circumstances tend to score more highly on intelligence tests than subjects in poor social circumstances. The relationship between test performance and social environment has been demonstrated with many different types of test, and for many different methods of classification of social background.

No/

/No firm conclusions as to the meaning of this relationship have yet been reached. This may seem a strange statement to make, in view of the large number of studies carried out over the last forty years. Yet the fact is that few investigators have gone beyond the measurement of the degree of the association between test performance and social environment. Studies attempting a more comprehensive and detailed analysis are infrequent. Thus the size of average differences in performance between social groups is now accurately known ; but their interpretation still presents many difficulties.

It seems abundantly clear that no further need exists for research which simply measures the extent to which social groups differ in test performance. Repetition of experiments is as desirable in psychology as in the exact sciences; but there must come a point in the examination of a problem when it is recognised that research has to <u>proceed from</u> certain findings.

There are several possible interpretations of social differences in test performance, and each has had its supporters among psychologists and educationists. It may be argued that these differences reflect real differences in natural ability; in that case it follows that the lower social groups or classes in the community are inferior to the upper/

/upper social groups in genetic endowment. Alternatively, it may be argued that these differences reflect the obvious differences in physical environment; in that case it would be anticipated that an improvement in the social conditions of the lower social groups would lead to an improvement in their test scores. A third line of argument is that these differences are indicative of bias on the part of the test constructors, who include in their intelligence tests a disproportionate number of items the style and content of which is familiar to subjects from the upper social groups; in that case hypothetical "culture-free" tests would show no differences between social groups.

These interpretations are, of course, not mutually exclusive. Conceivably real differences between social groups in innate ability may exist, but be exaggerated by the differences in physical environment and by bias on the part of test constructors.

There is also a possibility, little considered so far, of qualitative differences in mental functioning between social groups. Biesheuvel (1943) has suggested that the general inferiority of Africans on tests designed by European psychologists may be due to the fact that their thought processes and perceptual habits are qualitatively different from those shared by Europeans. It may be that

/cultural

/cultural differences within European society are large enough to produce similar qualitative mental differences among Europeans themselves.

It will be clear that much research remains to be carried out before psychologists can decide which interpretations are to be accepted and which to be rejected. At the moment there is no firm reason why one should be preferred over another.

The problem is an extremely complicated one, which invites a lengthy programme of research. Only limited aspects can be investigated by a single researcher. The scope of the present study has been closely defined. It does not attempt to consider directly whether or not there are "real" differences in ability between social groups, since the investigator is of the opinion that in the present state of knowledge such an undertaking would be premature. Its main method is the analysis of the performance of children in five different social groups on individual items and categories of item in a group intelligence test. Although it naturally leans on previous work to some extent, the investigator believes that it may be considered as original in its treatment of many aspects of the problem. No other comprehensive study of the intelligence test responses of children of different social classes has so far been undertaken by a British psychologist. None of the American investigations reviewed in later chapters have/

/have compared the test responses of children from as many as five social groups. No former experimenter has examined in such detail the incorrect responses of subjects or analysed so closely the variation in social discrimination shown by items at different points on the social scale.

Although little is said here of the practical relevance of research on social differences in intelligence, it must be obvious that there are few problems in psychology the solution of which would have more influence on social and educational policies. If it is found as the result of detailed investigat--ion that these differences probably do not reflect real differences in natural ability, then clearly injustice is being done by modern selective procedures to adults and more especially to children from the lower social classes. If, on the other hand, investigation verifies that these differences are due to real differences in natural ability, then the selective procedures are empirically justifiable. Such verification, however, might have considerable implications for democratic political theorists.

PART 1.

A CRITICAL EVALUATION OF PREVIOUS RESEARCH ON THE RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT.

CHAPTER 11.

SOME REASONS FOR THE EXISTENCE OF OPPOSITION TO THE PRACTICE OF INTELLIGENCE TESTING.

There can be little doubt of the poor integration of the work of the various groups of social scientists working in the field of community studies. This is partly due to the fact that each discipline interprets social phenomena in terms of its own special concepts. Partly relevant also is the real hierarchy of prestige which exists among the social sciences, history coming somewhere near the top and the psychology of social behaviour floundering at the lower levels. Students of the older and better established studies are often woefully ignorant and even disdainful of the findings of the newer disciplines. Text books by distinguished historians and economists match a sophisticated interpretation of events with a naive and mechanical conception of the springs of human action. At the same time the works of, for instance, Geoffrey Gorer (1948,1949) on human behaviour in the United States of America and the Soviet Union/

/Union are almost aggressively impressionistic and unmethodical -exalted journalism rather than social science.

Parity of esteem among social studies will undoubtedly come, although its arrival is delayed by the very nature of society itself. In general history concerns itself with the past, and the degree of remoteness of its concern is of little matter. In a very real sense the nineteenth century is as far away from mid-twentieth century man as is the twelfth century. Social anthropology has in the main occupied itself with the study of primitive and distant tribes. Classical economics tends to work within its own abstract model of society. Workers in these fields, so long as they do not step outside their closely defined limits of study, may expect a cordial acceptance of their professional skills. But sociologists, social psychologists, and their comrades from other disciplines who have contracted out of the unwritten agreement not to turn too close and critical an eye to the examination of real life as it is lived meet with considerable resistance and opposition, both at conscious and unconscious levels. Writing of social psychology Krech and Crutchfield (1948, p.6) say: "In so far as social psychology attempts to gain an understanding of social thinking and behaviour, it must perforce delve into the nature of the basic goals of the various individuals of society. In so doing, it is likely to uncover facts that are/

/are difficult for some people to take, and it may violate strongly ingrained prejudices, beliefs and vested interests." And again:" Society's resistance to a science may reflect the differences in the number and class of people and ideas jeopardized by the body of knowledge built up by the science." It may be added, although Krech and Crutchfield do not make the point, that the social resistance to investigation to which they refer will be particularly marked in an era of pronounced social change such as the one through which Western man is passing.

It may be thought that the present investigation is not the kind of investigation to which such considerations apply. Intelligence tests have become an accepted part of the national life, and are continuously employed for selection or diagnosis in schools, prisons, the armed services and to some extent in industry. Yet an observer who merely recorded this fact would not be doing justice to his material. The exist--ence of opposition to the use of intelligence tests and to the arguments of psychometrists regarding their value can be substantiated by reference to the most diverse sources :- debates in the House of Commons, discussions in local councils, letters in the correspondence columns of newspapers, the statements of speakers, sometimes distinguished academics or men of affairs, in radio and television programmes, the recorded/

/recorded opinions of parents and teachers. Even psychologists themselves are not immune from the general unease, as may be illustrated by the often - quoted remark of Zangwill (1950) that intelligence testing is ' a technology whose theoretical foundations are distinctly insecure'.

It may be agreed that part of the reason for the opposition to intelligence tests is the natural egoistic dislike of the individual for classification and assessment. But another, and more important part, is the fear aroused by the intrusion of scientific method into fields of human activity hitherto controlled by pre-scientific 'prejudices, beliefs and vested interests'. If it is accepted that the scientific measurement of human abilities is technically possible, then it may be argued that for the first time in history man possesses the means of organising his society in a really efficient manner, each individual performing the task to which he is best fitted. There is small doubt that many psychologists and educationists do argue along these lines, although perhaps not in public, and that they exert a continuous, if gentle, pressure for the widening application of the new scientific tools which they claim to have invented. As an example reference may be made to the many assertions in periodicals and text books of the superiority of testing to traditional methods in the selection of university students. The/

/The attitude of Eysenck (1955, p.153) may be quoted as typical. He says "........the case for the introduction of selection procedures by properly qualified psychologists (into student selection) is unanswerable". In a recent series of articles Himmelweit and Summerfield (1951) make precisely the same point. Certainly if and when intelligence tests do come to play a large part in student selection in this country, the psychometrists will have won a major victory, since they will in effect be determining who are to be the leaders of the next generation. They will thus have moved a long step nearer to the logical organisation of society in terms of the abilities of its members.

There have been hints from some psychologists and educationists that the longer society is run along non - scientific lines the more likely it is to do itself irremediable damage. Burt (1947) and Cattell (1937,1940) have both suggested that national intelligence is declining, since the lower social classes in the community are less intelligent than the upper classes, but have larger families. The Scottish 1947 Mental Survey was planned and carried out to investigate this hypothesis, which was also put forward by Thomson (1947). In fact the investigation did not confirm the hypothesis, although it was not definitely disproved. Naturally this idea of working - class inferiority has not been/

/been greeted in a friendly manner by the articulate representatives of the lower social classes.

It is difficult for one trained in objective thinking to sympathise with the attitudes of those who oppose the spread of testing simply because they dislike the scientific approach, or are opposed to change, or find the present organisation of society very much to their interest. But there exists an opposition whose arguments are more import--ant, and which mental testers would be unwise to ignore. It divides itself into two groups. In the first are the people who tend to admit the validity of testing, but are convinced that a society in which position was based purely on ability as defined by the tester would be objectively worse to live in than present- day society. In the second are those who are willing enough to change, but maintain that the testers are deluding themselves about the validity of their findings, the general significance of which is strongly limited by their dependence on data collected within a special set of social circumstances. (There are also, of course, a few people, mainly professional psychologists or social scientists themselves, who claim that the actual techniques of testing are fundamentally inapplicable to the material on which they are exercised, or at least very faulty.)

The attitude of the first group can scarcely be discussed/

/discussed here. The views of individual members of the second group are discussed and illustrated in later chapters, particularly in reference to the varying interpretations placed on the apparent distribution of intelligence through--out the community.

Although the wider social implications of whatever viewpoint an individual may take on the question of intell--igence testing and the interpretation of test results are extremely interesting and profitable to consider, it will be in the sphere of education that the effects of this viewpoint will be most clearly evident. There are few extreme hereditarians left nowadays except among the geneticists. But in the mind of Karl Pearson (1919), one of the fathers of modern psychometrics, no doubt existed about the inter--pretation of test results. These results showed hereditary differences in intelligence between individuals which were positively associated with social class ; the Darwinian process of selection had gradually insured that the more intelligent individuals rose to the top of society, and the less intelligent sank to the bottom. Education therefore should in general be differentially designed for the different social groups in the community. There seems still to be a hint of this approach in the suggestion of Sir Cyril Burt(1950) that examiners should take home conditions and social background into account when selecting for secondary education./

/education. At the opposite pole is Bagley (1926), who, in his paper written only six years after Pearson's Function of Science in the Modern State, rejected forcibly the notion of an hereditary and unchanging intelligence, and insisted on the powerful influence of good schooling. Obviously the writings of Bagley and his successors have had considerable influence in the United States, as may be judged by the fact reported by Eysenck (1953) that there exist American colleges where the average intelligence, as measured by tests, of the student body is below the average intelligence for the whole country.

CHAPTER III.

THE SOCIAL BACKGROUND OF DIFFERENCES OF OPINION AMONG SOCIAL SCIENTISTS.

The integration of social studies so obviously required will follow the full acceptance of the newer sciences, and there will be a pooling of social models. At present psychologists, educationists, sociologists, and social/cultural anthropologists may often be found proffering competing interpretations of the same data, each researcher fighting for his own hand and his own science. The followers of different disciplines are predisposed to construe a common store of factual evidence in different ways. With regard to the evidence drawn from intelligence testing, Vernon (1955), for instance, while admitting the modification of earlier simplistic inter--pretations, on the whole considers that it can be accepted at face value. The sociologist Glass (1954), on the other hand, is very sceptical of the value of tests, and even more so of the mode of organisation of education in England/

/England for which they have provided so much theoretical justification. The social biologist Darnott (1958) simply refuses to accept that differences in mean test score between social or racial groups have any real meaning. Typical also is the reaction of the human geneticist Penrose (1949) to psychological prognostications of a continuing national decline in intelligence; he constructs a genetical model of the community which, if accurate, would maintain the level of intelligence from generation to generation.

These differences of view cannot be explained only in terms of objectively varying intellectual positions. In the first place the very social circumstances under which social scientists live themselves influence the nature of scientific opinion, in such a way that certain theories tend to be propagated and other theories tend to be denigrated or ignored. The kind of theory which a social scientist espouses tends to define him as a conservative or a radical. Polanyi (1949) makes something the same point when discussing why particular scientific discoveries, of apparently equal merit, are " recog--nised and developed further, or discouraged and perhaps even smothered at birth ". It is possible to see this process at work within the field to which the present study is directed. Studies which have been carried on over a number of years at the University of Iowa, and which have been described in a series of research reports by Skodak (1939), Skeels (1940), Skodak/

/Skodak and Skeels (1949), Skeels and Harms (1948), and Schmidt (1946) have seemed to show that environment exercises a much greater influence on intelligence level than it has been fashionable to suppose. These studies have come under heavy attack from other psychologists, notably McNemar (1940), Simpson (1939) and Kirk (1948). An examination of their technique of criticism brings the uneasy feeling that they are committed to disbelief before they begin to sift the evidence. Vernon (1950, p.41) himself says of the Iowa experiments, "It would be advisable .... to suspend judgement until similar experiments have been repeated under more strictly controlled conditions". It should be added that, although the further verification which he suggests is clearly a matter of extreme importance, there would seem to be reluctance on the part of the critics of the Iowa findings to undertake their own experiments.

Similarly the research into the intelligence test performance of children of different social classes carried out by the Committee on Human Development of the University of Chicago and reported by Davis (1948) and Eells et al. (1951) has received marked lack of support among psychologists and educationists. In what is probably the best recent survey of modern developments in intelligence testing Vernon (1955) does not at any point refer to this work, although he quotes from other/

/other American researches of the same period.

While the "radicals" are accorded cavalier treatment, their colleagues on the other side of the heredity/environment fence receive a much more gentle handling. Vernon says, for instance, in the paper just noticed, "Competent experts like Burks in America...... calculate..... that the hereditary contribution to I.Q. variance is three or more times as great as the environmental ".(Vernon,1955, p.8) Burt (1957) also is highly commendatory of the famous experiment of Burks (1928a), in which the relationship between the intelligence of foster - children and foster - parents was compared with that between the intelligence of ordinary children and parents. Yet her procedures are open to considerable question, according to Loevinger (1940), and this has been recognised for a considerable time. It has not prevented her conclusion being repeatedly quoted and supported for the past thirty years.

In all fairness it must be stated that faults are not confined to one side. Spiker and McCandless (1954), looking at the whole vexed question of intelligence testing from an "operational" viewpoint, criticise impartially both environmentalists and non - environmentalists, pointing to an equal laxness in terminology and equal errors of procedure in crucial experimental designs. They believe that differences can be resolved/

/resolved by rigorous terminological analysis; to the present experimenter this appears an unduly optimistic conclusion.

There are, of course, qualities of behaviour charact--eristic of experimenters and research workers in the social sciences for which society cannot be held to account. Fear (1955) has remarked on the tendency for experimental groups to consist of children, students, members of lower social groups in the community, etc. He maintains that this cannot be explained entirely by their greater availability for investigation, but is connected with the need of investigators to bolster their own prestige. Discussing Warner's conceptualization of social class, Bernard (1951) asks why it met with so much response. "Did it satisfy something in the sociological public ? -- in the researcher ? Did he enjoy his own class security as he described those "lower" than himself ? Did he enjoy a vicarious if snobbish identification with those "higher" than himself ------ or perhaps a kind of malicious pleasure in revenging himself on them by describing their mores dispassionately and non-sneeredly ? Or was it because, as some have alleged, it was an academically respectable way to demolish the Marxian concept of class based on power relationships rather than on status ? "

There/

/There is no doubt that it is difficult for workers in the social sciences to preserve the necessary scientific objectivity, and unfortunately this becomes all the more evident the more closely their work is related to actual social problems. An interesting example of the effect of political bias on educational and psychological theory is an article by Simon (1949). Writing as an orthodox Marxist, she condemns the whole science of psychometrics as "bourgeois"; she hints that intelligence testers are motivated by a desire to retain the present order of society; she claims that fellow - Marxists who retain some faith in testing are tainted by "capitalist" ideologies. MacRae(1953) goes so far as to assert that most writers who show an interest in social stratification are conscious or unconscious egalitarians(although he does not think that this biases their findings unduly.) In view of the recent appearance of books by, for instance, Lewis and Maude (1949) and Eliot (1948), who are far from egalitarian in philosophy, perhaps it would be more accurate to say that writers on stratification are usually committed, whether they realise it or not, to a definite set of political and social attitudes. On the basis of an examination of a number of academic writings, Maude(1953) decides that English intellectuals are invariably political radicals. It is important to determine whether this tells us more/

/more about Maude or about English intellectuals.

Of all subjects of social investigation, the social distribution of intelligence is the more bedevilled by imponderable considerations which hamper the discovery of true facts. Sprott (1952, p.137) comments wryly that whenever we turn to consider the role and nature of intelligence in society we enter upon a battlefield : "the trained war horses of educational psychology paw the ground". It has been the purpose of the present writer to show some of the forces present on this battlefield, together with the reasons for their presence.

CHAPTER IV.

CONCEPTIONS AND DEFINITIONS OF SOCIAL CLASS.

In no investigation of social class characteristics is it wise of the investigator to concentrate on experimental findings at the expense of preliminary consideration of the social variables with which he is working. Unfortunately this has often been done in pyschological and educational investigations of the social distribution of intelligence, perhaps understandably, since in an age of specialism psychologists and educationists can seldom be expected to be as well - read in sociology as in their own fields. The recent development of educational sociology, especially in Britain, is probably a sign of better things to come.

Even from the point of view of the sociologist, however, the study of social class leaves much to be desired. There is little community of agreement as to definition. Considered purely as a sociological variable, little is known of its mode of operation in the community. Although it has been/

/been written about for many years, innumerable crucial experiments remain to be made. According to Centers (1949), many who employ the term "social class", even prominent researchers, have only a very vague concept of its meaning. Marshall (1950, p.86) points out, "It is a bad thing for a subject when few people write on it at length, but many in brief". In view of the considerations referred to in the last two chapters, this state of confusion, conflict, and semi-neglect is hardly surprising. It has been remarked that, although few are unaware of the pervading influence of social class, not many people care to discuss this taboo subject (cf Pear,1955).

Nevertheless the examination of writings on social class in recent years shows that two main approaches towards the problem have crystallised out. One insists that social class is best defined and interpreted by reference to objective social facts such as occupation and income. The other main--tains that if social class is accepted as a determining influence on human behaviour, what matters is not the economic position of people in society, but how they individ--ually see that position. In other words, there exist two main competing interpretations of social class, one stressing an objective, the other a subjective, approach.

That/

/That is not to say that other views cannot be held. It may be argued that although social stratification in general is a reality, social class is largely an illusory concept created by political dogma, a sociological myth. This is roughly the position adopted by Sorokin (1937). He cites forty - eight different kinds of definition of social class, decides that they are all unhelpful, and concludes that it is much more useful to talk separately of economic, occupational and political strata in society. However, in spite of his massive reputation, Sorokin's approach to many problems in sociology is highly individ--ualistic, and not shared by the vast majority of his colleagues. But it should be noted that his attitude would find some support in many quarters. Conant (1948), a prominent American educationist, and a scientist by training, claims that social classes are a European phenomenon, and do not exist in the United States "in any real sense" ; he does not of course deny the existence of differences in level of income and standard of living.

Other educationists have been so convinced of the real existence of social classes that they have not bothered about niceties of definition. Finney (1928), for example, is so taken with the virtues of "middle- class" life that he sees the major aim of education as ensuring that every child/

/child becomes middle - class./

Many investigators have tended to avoid the problem of definition and interpretation by writing of an amorphous variable which they call "socio - economic background", and which in fact includes without differentiation all systems of social stratification. In a large proportion of the researches to be reported in later chapters, this concept will be found to have been used. It would be unfair to criticise the researchers too harshly. But the point must be made that in evading the problem of definition they are automatically creating difficulties in the assessment of results both for themselves and others.

In recent years, especially in the United States, investigators have tended to use subjective valuations of social class, and to decry the efficiency of objective indices. Typical examples of their methods in practice may be studied in Warner and Lunt (1942). Warner may be regarded as the leader of this school, and he has influenced many young investigators. Looked at from a more generalised point of view, the movement owes much to the method and style of social anthropology. A survey of modern studies of primitive communities show how much anthropologists depend on the personal reports and statements of members of these communities to build up a picture of their inner structure/

/structure. For instance, in Firth (1956) the working of Tikopia society gradually reveals itself through the con--versation and the answers to questions of the Tikopia themselves. It is not surprising, in view of the spectacular success of the methods of the anthropologists, that these should have transferred to studies of modern industrial societies ; whether their use is legitimate for the purposes to which they have been put is another matter.

Not only the methods, but also the concepts and general--isations of social anthropology have been borrowed by Warner and his colleagues. Centers (1949) says, "Classes are psycho - social groupings................ class lines of cleavage may or may not conform to what seem to social scientists to be logical lines of cleavage in the objective or stratification sense ". The legitimacy of this borrowing must be very seriously questioned ; it would be hard to deny that social class has its psychological aspects, but any assumption that hence social class is psychologically determined is very debatable.

Lipset and Bendix (1951) put down the popularity of this subjective, anthropological approach to social class to the proclivity of social scientists to react against Karl Marx. Bernard (1951) talks of a "cult of culture" and remarks that if a science were a rational structure, the subjectivist approach of/

/of the anthropologists would long since have been modified.

It would seem, however, that recognition of the fact that subjective definition of the concept of social class has many weaknesses is slowly taking place. Lipset and Bendix (1951), criticising the way in which Warner has treated the information provided by his informants in his analysis of the structure of small-town society in the United States, show that he places undue weight on the evidence of people in the upper social groups, sometimes neglecting lower-class views entirely. They believe there is a danger that his conclusions may merely reflect the outlook of the upper social groups on social stratification ; whatever the defects of objective indices of social class they are not liable to this kind of subtle distortion. They are also inclined to feel that in so far as his complicated work on subjective status placement leads to objective conclusions it is unnecessary, since objective variables get at the same stratification factor. Warner himself has acknowledged, without seeming to see that his earlier work is thus compromised, that subjective status evaluation and economic position are highly correlated. (Warner, Meeker and Eells,1949).

Objective indices of social class are not, of course, without obvious deficiencies. In particular the implication of psychological similarities from economic similarities has to/

/to be done with great caution. For an example of the really skilful handling of this problem, reference may be made to Marshall (1950, pp 112-115). Accepting Marxian terminology with a number of reservations, he indicates how psychological traits characteristic of the "bourgeoisie" may be deduced from their position in society. He says, "Capitalists, in the sense of owners of property, do show common features when examined in terms of social class, but these are not derived from the fact that capital gives power over labour, nor primarily from the fact that capital yields an income, but rather from the fact that property, however small, gives security and insurance against misfortune and liberty for new adventure, thus cultivating a sense of proprietorship in a civilisation, of independence of status, which makes governments appear as servants, not as masters, and institutions as the means to freedom, not to servitude". Unfortunately this statement would tend to be viewed with suspicion by the very people who are most interested in the psychological characteristics of which he speaks, because it is logically, and not experimentally, deduced.

The most interesting theoretical work on social class to be published in Britain in recent years by a psychologist is that of Burt(1955). Not very enthusiastic about objective methods of assessment, on the grounds that they are of doubtful/

/doubtful value in a swiftly changing and highly mobile society, and sceptical of the typical American methods, because of doubts as to their applicability in other countries, he suggests a number of "social" criteria, for instance speech, dress, manners, equality of social intercourse, which can be applied with a fair measure of objectivity by psycho--logists trained in observational techniques. Irrespective of our agreement to this approach, his tabulation of the problems which psychologists studying the phenomenon of social class should seek to solve is of value. In the first place, he says, sound methods of differentiating the different social classes in the community must be discovered ; secondly, the way in which mental characteristics differ from one class to another must be shown ; and thirdly, the true relationship between psychological differences and class differences must be determined. In a small way the present investigation may be regarded as an attempt to carry out a part of Burt's scheme of work, since intelligence may be treated as a psychological characteristic varying from class to class. But the field of intelligence testing does not lend itself easily to the employment of observational techniques, although Piaget has used such techniques in his studies of the nature of intelli--gence, and they are a necessary part of the equipment of psychologists working with pre - school children.

The/

/The attitude of the present investigator may be defined as follows. As yet there is still a marked lack of detailed information about the way in which intelligence test score varies from social class to social class, although the fact that it does vary has been illustrated often enough. What is important therefore is the accumulation of more information. Under these circumstances it is most sensible to use some mode of definition of social class in the investigation by which the class of subjects is readily determinable, such as occupational level in the community. (In fact, as will be seen in Part II, this experiment is carried out with school children ; it is assumed that, once the social class of parents is known, the children's social class is automatically determined.) It is recognised that some of the more subtle manifestations of class behaviour are therefore likely to escape notice ; but, so long as this deficiency is remembered in the interpretation of results, findings should still be of value. There is too great a tendency for social scientists working in this field to claim an unwarranted universality of application for their conclusions ; no such tendency will be found to operate in this investigation.

One effect of the conflicting schemes of definition of social class is difficulty in determining the number of social/

/social classes which can be differentiated within the community. Marxists accept the existence of only two real classes in capitalist society, the bourgeoisie and the proletariat. Lay terminology usually refers to the upper classes (the aristocracy, land owners, etc.), the vaguely conceived middle classes, and the working or lower classes (usually seen as manual workers of one type or another .). Anthropologically - biased American researchers have claimed to find six social classes within industrial communities of normal complexity, which they have designated as upper upper, lower upper, upper middle, lower middle, upper lower and lower lower (Eells et al, 1951.) Pear (1955) employs a five--fold differentiation ; the upper class, the middle class (1), the middle class (2), artisans and skilled workers, and unskilled workers.

In default of exact information, it is up to each researcher to make his own decision as to the scheme of social classification he is to adopt. It will be found in Part II. that the present investigation assumes the existence of five social classes among school children in Scotland who are receiving a normal education ; these have been called upper-middle, middle-middle, lower-middle, upper-working and lower-working.

CHAPTER V.

CONCEPTIONS AND DEFINITIONS OF INTELLIGENCE.

The literature on many aspects of mental testing is so bulky that any researcher placing his work within an historical context has to limit severely his conception of what is relevant. Care has been taken here not to enter at length into the discussion of problems which are not directly associated with the subject under investigation.

Nevertheless, as there will be continual references in the text to "intelligence", it was thought necessary that some space should be devoted to discussing how this concept is defined by the investigator.

There are many definitions of intelligence. At one time every leading psychologist seemed to be interested in this problem of definition, and whole symposiums were concerned with its discussion. (Brit. J. Psychol, 1910, ~~[illegible]~~). A useful summary of the important earlier definitions is contained in Knight (1933). Many of these precede, or take little account of, factor analysis ; they tend to be "dictionary" definitions chosen/

/chosen for their comprehensiveness as, for instance, that of Tolman ——— " the sum total and degree of development of the organism's capacities to perceive, differentiate, integrate and manipulate its environment " . In more recent years definition has often been in terms of "factors" which various experimenters claim to have isolated by means of mathematical procedures. These "factors" have multiplied in number especially as the result of work done by American psychologists. The best summary of the later definitions is contained in Vernon(1950). Thurstone has talked of various special (statistically determined) vectors of intellectual capacity such as memory, verbal comprehension, verbal fluency, space visualization, number facility, induction, deduction, speed of reaction time, perception, judgement, closure, and rate of reversal of ambiguous perceptions. Neill has mentioned fluency, plasticity (the breaking down and reorganization of structures), complexity (the ability to realise complicated intellectual structures), and globalization(uniting separate data into a single whole). In Britain, under the influence of Burt, Spearman, and Thomson there has been much less tendency to differentiate the original concept, and the idea of a fundamental unifying force has been retained. Even so, it has become more common to refer, for example, to the distribution of "abilities"/

"abilities" rather to the distribution of "intelligence". And it has been suggested that "intelligence" tests should now be called "academic aptitude" tests.

Recently, however, there have been new attempts to construct a comprehensive theoretical framework. These spring from the observational work of Hebb (1949) and Piaget (1950). Both show the importance of contacts with the environment, and hence of learning and experience, in developing mental behav--iour. Hebb is therefore led to distinguish between two aspects of intelligence, one fundamental, the other derived. Intelligence A is an innate potentiality for forming and combining assemblies of neurones in the association areas of the brain, and is probably genetically determined. Intelligence B is the intelligence which is met in daily life, which, although dependent on Intelligence A, is in detail largely acquired, and does not develop unless adequate environmental stimulation is present. In practice Intelligence A cannot be measured, but has to be called into being to explain the existence of Intelligence B. Intelligence B can be measured with varying degrees of adequacy by intelligence tests. It cannot therefore be assumed that a poor score on an intelligence test indicates an innate inferiority ; there is greater likelihood of an unsatisfactory environment, especially in very early life.

Although/

/Although Vernon (1955) has stamped this approach with his approval, it has not yet won general acceptance. Perhaps the reason is not far to seek. At one stroke it removes from the field of practical discussion the whole nature/nurture, heredity/environment controversy. The environmentalists find that the principle of hereditary determination of intelligence is retained, but in a way that makes it very difficult for them to attack. The hereditarians find that while in a very general sense their ideas are accepted, these can no longer be applied to the interpretation of intelligence test performance.

It would seem, however, that this formulation would be accepted by Davis (1948). While willing to admit the existence of fundamental genetic differences between individuals, he lays great importance on the role of learning in the development of systems of mental behaviour. Much of this learning is "cultural learning", i.e. induction into the characteristic patterns of particular "cultures". Since there are different social groups within a modern society, each constituting a "culture", individuals tend to learn and develop differently, irrespective of their genetical structure, according to the group to which they happen to belong. As intelligence tests are compiled by the middle social groups in society, the items which they choose for inclusion tend to reflect middle-class learning/

/learning experience, and therefore discriminate against subjects who belong to the lower social groups. Intell--igence testing can thus arrive at no satisfactory conclusions until tests of "general problem-solving ability" are devised.

Davis nowhere commits himself on the question of whether a middle-class "culture" can be considered more stimulating to personal development than a lower-class "culture", or middle-class behaviour patterns "better" than lower-class behaviour patterns. He insists that the aggression and hostility so prevalent in slum areas are realistic adaptations to circumstance, and in no sense pathological. He claims that among the lower classes there is much less guilt and anxiety with regard to sex.

In Britain there would tend to be some distrust of Davis' anthropological terminology. While the existence of qualitative differences between social groups might be accepted in theory, in practice there would be a strong feeling that a middle-class environment was superior to a working-class environment, and that middle-class behaviour was better than working-class behaviour. British environ--mentalists would therefore tend to say that differences in test score between social groups betokened a real, if conditional, inferiority which would not be eliminated by using new tests but only by the "bourgeoisification" of the lower classes.

A/

/A growing school of psychologists, namely the "operationists", would maintain that the progress of intelligence testing has been hampered by unscientific controversies over questions of definition and interpretation. Some radical operationists assert that the meaning of a quantitative concept lies exclusively in the set of measuring operations which determine the application of the concept ; hence intelligence is defined as what intelligence tests test. This approach can be illustrated by the statement of Pratt (1945). "Scientific explanation in the last analysis is a statement of observed correlations, the relations which have been discovered to obtain among observed events. My impression is that science is a vast and impressive tautology.........This is not a discouraging thesis except to those who would demand of science some penetration into a world not compounded of palpable stuff".

Clearly operationists---the more rigorous at any rate--- would dismiss much of the work of Davis as theorising ; they would regard Hebb's hypothesis of mental structure as belonging to the realm of metaphysics rather than science.

In general it will be found, however, that the present investigator tends to accept Hebb's formulation. Historically speaking, science seems to have advanced through a more liberal conception of the role of hypothesis than that suggested by the operationists. But care has been taken, as will/

/will be seen in Part II, not to introduce unwarranted speculation. So far as possible, discussion has been confined to the facts which emerge from the data collected, and would be very much the same, no matter what the viewpoint of the experimenter. It may of course justifiably be pointed out that an experimenter who defined intelligence differently would have carried out a different experiment.

CHAPTER VI.

EARLY RESEARCHES ON THE DEGREE OF RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT.

It is easily forgotten that the belief in the existence of differences in ability or intelligence between the different social groups in a community is much older than intelligence testing. This belief has in fact been held by the majority of political thinkers from Plato onwards. Many examples are quoted by Sorokin (1928). Attempts to justify it experimentally do not appear, however, until the end of the nineteenth century. These may perhaps be attributed not only to the desire for a more accurate and scientific approach towards social problems, but also to the uneasiness caused by the advance of democracy and the growing political and industrial power of the working classes. Darwin re-interpreted Malthusian economic theory in terms of biology. His biological hypotheses were then re-applied to society by many thinkers who did not possess his careful and cautious temperament/

/temperament ; they tended to crystallise as slogans which were often accepted uncritically ; men talked of the "struggle for survival" and the "survival of the fittest". The new science of genetics, founded by Bateson and Morgan in the first years of the twentieth century, appeared to provide detailed evidence for theories of "social Darwinism". There is no doubt that Galton and Pearson, the founding fathers of psychometrics in Britain, were strongly influenced by these developments. They in turn influenced a generation of psychologists, with the result that ideas which have been much modified or even discarded by biologists and geneticists still lurk in psychological writing in one disguised form or another, and cloud the interpretation of experimental investigations.

Since the tools necessary for the direct investigation of intelligence were not available until the publication of the Binet-Simon tests in 1908, early studies of the social distribution of abilities invariably proceeded by analysis of the early social backgrounds of distinguished men. There are a great number of these studies, all very similar in procedures and findings. Among the best known are those of Galton, Ellis, Cattell, Woods, and Clarke.

The earliest study is that of Galton (1869) . He found that 977 eminent men each defined as the most eminent of 4000 random persons had 535 relatives as close as grandfathers/

/grandfathers or grandsons of at least the same degree of eminence, whereas 977 men chosen at random would only have had 4 such relatives. In a later study (1904) he obtained family schedules from 200 Fellows of the Royal Society which showed a large number of relatives as distinguished as the Fellows themselves. As relationship became more remote there was a consistent dropping off in numbers of such relatives. In other words, eminent men tended to belong to eminent families and thus to middle-class families with a background of professional work.

Ellis(1904) studied the family origins of 108 British men and women whose professional achievement was sufficiently high for them to be regarded as geniuses. He showed that in the main they were descended from professional, and usually upper-class, families .

Cattell (1903) investigated the family background of 1000 important American scientists. He concluded that these had far more professional men among their fathers than an equivalent number of men from the general population. (In a re-analysis of Cattell's data Brimhall(1922,1923) showed that the brothers of these distinguished scientists were 72.8 times as likely to be in the American Who's Who as the brothers of an unselected group, and the sisters 70.8 times as likely as the sisters of an unselected group.)

Woods/

/Woods (1925) obtained the names of 3500 eminent Americans from biographical dictionaries. Claiming that the ordinary man had only one chance in five hundred to have an eminent man as a close relative, he pointed out that one in five of the 3500 had a close relative among the 3500. He concluded that one per cent of the population was as likely to produce a man of genius as all the rest of the population.

Clarke (1916) collected information about 666 prominent American literary men. He found that 49 per cent had fathers belonging to the professional classes, and that only seven per cent had fathers who were clerical workers or skilled or unskilled manual labourers.

These studies have been included to preserve the historical continuity of the narrative and as pioneer invest--igations, rather than because they are considered to-day to be of major importance. They tended to be conceived as illustrative examples of the social weight of heredity rather than as experimental investigations in the modern sense. Sometimes the distinctions which they made between groups in the community were based on a criterion of intellectual achievement rather than on occupational criteria, and there--fore they bear only indirectly and at long range on the present problem. There are many more which it has been felt unnecessary to discuss. For instance, there exists a whole category/

/category of studies which group eminent historical personages in terms of their social origins, and show how they tended to belong to the upper social groups. Cox(1926) analysed the occupations of fathers of 282 of the most eminent men living between 1450 and 1850 ; 52.5 per cent had fathers belonging to the professional classes or the nobility, whereas only 1.1 per cent had fathers who were unskilled workers. There are also some well-known studies which deal only with the biographical details of individual families. Galton investigated the famous Darwin - Wedgwood family group, with which he was himself associated. In America the similarly distinguished Jonathan Edwards family was studied by Winship, while family connections were also discovered with singularly bad social and intellectual records------the Jukes, the Kallikaks, the Nams, the Hill Folk. As the science of intelligence testing grew and statistical methodology improved, the faultiness of the earlier type of investigation was gradually realized ; while it often presented facts in a striking and interesting manner, too much scope was allowed in their interpretation for the predilections of the experimenter.

Although Binet was probably aware of some of the early work prior to 1908 , and there is evidence that he considered some kinds of test question to be subject to environmental effect (cf Chapter IX, p. 135), it seems clear that he felt originally/

/originally that the scales which he and Simon developed were unlikely to show much differences between children from different social backgrounds ----- even though he does say in discussing the 1908 Scale " All our successive experiments and certain new ones have been made either at l'Asile Sainte- Anne, or at la Salpetriere, or in the primary and maternal schools of Paris. They bear therefore always upon children of the so-called working class. This is a fact to be emphasized". (Kite, 1916, p.184). His attention was quickly drawn to the fact that fellow-workers applying the 1908 Scale under new conditions in Belgium and England were obtaining different results from those he and his collaborators had obtained in Paris, and that his tests were hence being criticised as being too easy. In his last article in l'Annee Psychologique(Kite,1916,pp.274-329), in which he proposed certain emendations of the 1908 Scale, he took up these criticisms, particularly in regard to an experiment carried out by Decroly and Degand in Brussels, but also with some references to the results of testing carried out by a Miss Johnston in English schools. Decroly and Degand (1910) had applied the 1908 Scale to 45 children ranging in age from 3½ - 12½ years in a private school in Brussels ; obtaining norms for every age group above those of Binet, they had come to the conclusion that the Scale had been incorrectly standardized./

/standardized. Re - analysing their figures, Binet found that on the average the Belgian children were a year and a half in advance of his Parisian subjects. He explained the difference in social terms, pointing out again that his subjects belonged to a district of Paris "poor without being indigent", while the Belgian children belonged to a "social class in easy circumstances". He showed that the superiority of the Belgian children was particularly manifest in individual tests involving the correct use of language, and attributed this to their generally superior home environment.

In the same way he attributed the discrepancies in Miss Johnston's experiments in Sheffield, in which for the earlier years she obtained norms considerably above, and for the later years considerably below, those he had determined for the 1908 Scale, to the fact that she worked with children of very unequal social standing. He urged her to calculate new averages, taking account of the state of poverty or wealth represented by the parents of the children.

He also reported a number of experiments carried out in Paris with a variety of collaborators. In the first case he looked up some of the data which had been gathered in connect--ion with the standardisation of the 1908 Scale ; differences had then been noted, as he puts it, between children " in indigent conditions" and children "in easy circumstances". He/

/He was however unable to find any appreciable difference in norms between the two categories ; he concluded that this was due either to the fact that differences had been care--lessly noted, or that they were in fact too small to have any influence. (The latter seems a likely explanation, in view of his statement in 1908 already quoted that all his experiments had been carried out in working-class areas).

Dissatisfied with the inconclusive nature of this invest--igation, he asked a headmaster to test 50 children in his own school, at the same time classifying them with care into four social categories; indigence, poverty, mediocrity, ease. The data which he obtained from the headmaster, a M. Morle, were entirely negative. There appeared to be as many children superior to the average in the indigent group as in the group in easy circumstances; nor did there appear to be any difference with regard to the number of children inferior to the average. Binet again interpreted this as being due to the existence of social differences in M. Morle's school too small to have any appreciable effect.

Still not satisfied, Binet asked a headmistress, Madame Thevenot, to carry out an experiment in her school. She tested 18 children, 15 between eight and nine years of age, and three between seven and eight. Here a very different trend was evident in the results. No child scored below the norms of his age. Some were three years ahead,

and/

/and 6 were two years and over ; on an average there was an advance of 1.7 years on Binet's norms. Thus findings were similar to those of Decroly and Degand .

Madame Thevenot interpreted these results as meaning that the children were of a higher intelligence than children in other schools in which she had taught. Binet considered that the social condition of the parents had to be considered an important factor ; the school was situated in a wealthy commercial district of Paris.

As a final check, Binet asked M. Morle to measure the level of intelligence in two primary schools representing extreme social differences. 50 children from one of the poorest schools in Paris having been equated for chronological age with 50 children from a school in a good area, Morle found that 16 children in the good school had a mental age greater than normal by one or more years, compared with only five children in the poor school. At the same time 12 children in the poor school were backward by one or more years, compared with only four children in the good school. On average the children in the poor school were retarded by a quarter of a year and the children in the good school advanced by half a year, a difference in mental age of three-quarters of a year. While this was not as large a difference as in the Belgian experiment, it nevertheless pointed/

/pointed in the same direction.

On the basis of these investigations Binet concluded that there was a definite tendency for the children of the upper social classes to perform better in his tests than the children of the lower social classes, the degree of super--iority being dependent on the groups compared. Nevertheless he maintained the general usefulness of his tests as aids to teachers. He did not commit himself on the question of whether this superiority was innate or acquired, although it is clear that he laid great stress on the stimulating influence of a good home environment, and may therefore be placed among the environmentalists rather than the heredit--arians. Possibly it is not too fanciful to suggest that if he had lived longer he might have attempted to compensate in some way for class differences by computing different norms for different social groups ; there is a suggestion of this in the advice to Miss Johnston to take account in future investigations of the social level of the children she tested.

Binet's views have been illustrated here at some length, although the experiments he and his collaborators carried out are by modern standards relatively unsophisticated. It was felt necessary to show in detail the approach and attitude of the authentic founder of intelligence testing towards/

/towards this problem of social class differences in test performance. It was also thought proper to show that these differences have been evident from the very beginning of practical testing.

Binet does not at any point discuss the practical educational implications of social class differences in performance. But almost immediately after his death they were brought sharply into consideration by work carried out in Germany. In 1911 Stern and a group of associates tested pupils in a Vorschule and a Volksschule in Breslau, using Bobertag's revision of the Binet - Simon Scale. At this period in Germany Vorschulen catered for the children of wealthy parents, and were considered as preparatory schools for the Gymnasion ; Volksschulen were attended by working - class children. The German Social Democrats were strongly opposed to this undemocratic segregation in primary education, and wished all primary children to be taught in the same schools. Stern's experiment was designed to compare levels of ability in the two types of school ; it is described by Hoffman(1914) and Whipple (1914).

Stern found a definite superiority on the part of Vorschule pupils. His nine-year-old Vorschule pupils passed on average 70 per cent of the tests in the Binet - Simon series ; the nine-year-old Volksschule pupils passed 60 per cent of the tests. On the other hand ten-year-old pupils in the Volksschule also passed 70 per cent of the tests. Nine-year-old pupils in the/

/the Vorschule were thus equivalent to ten-year-old pupils in the Volksschule. Although he believed that this superior performance by the children of wealthy parents indicated their greater mental maturity, he was careful, in commenting on results, not to commit himself as to its causes. He sugges--ted that the only way to arrive at definite conclusions was to find out in which tests the superiority of the "cultured classes" was most evident, and in which tests the performance of the two social groups was similar. In view of the obvious--ness of this method it is astonishing that his suggestion has been so little followed up ; item - analysis has been employed extensively in the standardization of tests, but only to a very small extent in the investigation of test responses by subjects of different social classes.

The whole problem was taken up in the United States the following year by Weintrob and Weintrob (1912). Their study was the most extensive up to that date. They selected three groups of 70 children from good, medium, and poor environments respectively. Three institutions were picked to represent the three sets of social circumstances. These were (1)The Horace Mann School, which was attended by the children of wealthy parents,(2) The Speyer School, attended by the children of moderately prosperous parents (apparently including some from the working classes), and (3) The Hebrew Sheltering Orphan Asylum, a Victorian-style orphanage. In the Orphan Asylum/

/Asylum only children who had spent at least four years there were tested.

The test applied to the three groups in the three institutions was a modification of the Goddard Revision of the Binet Scale. The Weintrobs worked out an idiosyncratic method of scoring which was supposed to give the same amount of retardation from normal at all points on an age scale.

Their study is specially interesting because it does not show a definite relationship between social level and test performance. Differences between the three groups were always small. Only for the 8 - 9 age group was the order of performance what might have been expected, the Horace Mann School at the top, the Speyer School next and the Orphan Asylum last. For the other age groups (9 - 10, 10 - 11, 11-12, 12-13) and for the subjects as a whole the order was the Horace Mann School first, the Orphan Asylum second, and the Speyer School last.

In the analysis of these inconclusive results the Weintrobs placed most emphasis on the racial composition of their groups ; many children at the Horace Mann School were Jewish and the Orphan Asylum was an entirely Jewish institution. The present writer has been unable to find any investigation in the literature of intelligence testing specially designed to study the test performance of different social groups among the Jewish population ; the Weintrobs' experiment may perhaps be taken as evidence that differences are less than among other racial groups/

/groups.

It is possible that this pioneer investigation may have done something to stimulate interest in the comparative performance of different racial groups. Naturally in the United States the first comparisons were made between white and negro children. One investigation of the early period quotes some incidental findings which are of value. Using the Binet-Simon Scale, Strong(1913) tested 350 white and coloured children in primary schools in North Carolina. As might have been expected, she found that the white child--ren were superior. Out of curiosity, she divided the white children into two social groups, according to whether they belonged to the City of Columbia or to the mill-districts on the outskirts (workers in the mills being considered a specially depressed class). 95 white children lived in the city; 60 lived in the mill villages. All were between six and twelve years of age. She found that less than 6 per cent of the city children were retarded, while 10 per cent were over a year above the mental level of their age. In contrast no mill children were mentally accelerated, while 10 per cent were retarded.

A study reported in 1915 by Yerkes and Anderson (1915) was directly concerned with measuring the true extent of differences in performance between children from different social environments. It was carried out as part of the standardization/

/standardization of the Yerkes revision of the Binet Scale --- the most important feature of which was a new method of scoring on a point basis rather than with tests assigned to specific age levels. By early standards its design was care--ful and methodical. 54 children of from four to eight years of age were selected from the kindergarten and first grade of a grammar school in a good neighbourhood in Cambridge, Mass. and matched individually for age and sex with children from a school in a poor neighbourhood. The mean score for boys in the grammar school was 7.7 points higher than that for boys in the school in the poor neighbourhood ; girls in the grammar school had a similar advantage of 8.4 points. Although he had no means of assessing the statistical sign--ificance of these differences, Yerkes considered that they showed a 20-21 per cent inferiority on the part of children in the school in the poor district. Only at age 4 were they superior to their counterparts in the grammar school, and then only slightly. Above 4 the direction of difference reversed, and the difference in favour of the grammar school group showed a steady increase as the children became older. Yerkes attributed the minor discrepancy in the results for the very young children to the greater self-confidence of the poorer children at this stage of development. This seems to be an early formulation of the kind of approach towards the problem of/

/of class differences represented by Davis (1948) ; he points to the conformist, adult-respecting behaviour required of their children by middle-class parents in American society, and contrasts it with the freely aggressive behaviour encouraged by working-class parents.

On the basis of this investigation, Yerkes and his associates adopted a standardization procedure which has not been favoured by later test designers. They provided separate sets of age norms for two different levels of social environ--ment, for boys and girls, and for English and non-English speaking children. It would therefore appear , although there is no real discussion of the point, that they considered differences in performance to reflect environment rather than natural endowment.

From the experimental work in intelligence testing so far quoted it will be seen that none of the early experimenters cared to commit themselves as to the reasons for the inferior performance of children of the lower social groups. In fact, a careful reading of their work leads this investigator to conclude that, if anything, they were prepared to blame poor environment rather than poor heredity. This situation was changed by Terman's investigations, many of which seem designed to show a lack of environmental influence on test score(e.g. 1918, 1919,1925). The standardization data for his/

/his Stanford-Binet Revision of the Binet Scale, (1917) published in 1917, includes the most complete investigation of differences in test score between children of different social groups up to that date.

Terman's original standardization group for the Stanford - Binet Revision consisted of 1000 children. He had 492 of these children classified on a five - point social scale, social levels being defined as very superior, superior, average, inferior, and very inferior. A median IQ was computed for each of the five groups which had been differentiated. The results obtained were as follows. The very superior group had a median IQ of 106, the superior group an IQ of 107, the average group an IQ of 99.5, the inferior group an IQ of 93, and the very inferior group an IQ of 85. Ignoring the extreme groups because of the small number of cases, Terman emphasized the difference of 14 points of IQ between the superior and the inferior groups, and declared that this was equivalent to a difference of one year of mental age at age 7, and two years at age 14.

Clearly the discovery of such a large difference between social groups in test score must have been a matter of considerable concern to Terman, engaged as he was in the design of a new and important test. The development of different norms of performance for different social groups would have greatly hampered the employment of the Stanford-Binet in schools/

/schools or educational establishments ; children would have had to be classified socially before being tested, and close instruction would have had to be issued on how this classification was to be performed. On the other hand, if evidence could be called to suggest that differences between social groups in median score mainly reflected differences in native endowment in much the same was as differences in individual score between the members of a single group, it could be argued that no need existed for the development of different norms.

In order to determine how far differences in score between groups were caused by environmental differences Norman adopted a novel statistical procedure. Grouping IQs into a five-point classification, he built up a 5 x 5 correlation table showing the relationship between IQ and social level. As an estimate of the amount of relationship present he obtained a Pearsonian correlation coefficient of .40. He then divided up his 493 subjects into three age ranges, namely 5-8, 9-11, and 12-14 and worked out the Pearsonian correlation between IQ and social level for each age range, obtaining correlations of .43, .41, and .39. He interpreted this decrease of correlation with increasing age as showing the superiority of endowment over environ--ment as an influencing factor on test score. He claimed that/

/that if social and home environment were an important factor it would become more closely associated with IQ as the age of the subjects increased.

Terman's statistical method would not now be accepted as adequate, nor would it be agreed that the deductions which he draws from the data he presents are legitimate. From a modern point of view it might be said that he offers no sound reasons for his conclusion that difference in test score between social groups are mainly due to differences in natural endowment. Certainly it was more convenient for him and indeed for the teachers and psychologists who were to use his test to assume that they were ; and the weight of his considerable prestige gave the hereditarians a head start in the later controversies with the environmentalists.

Nevertheless Terman's study marked a considerable advance in methods of tackling the problem. Subjects were assigned individually to social groups, thus giving less scope for error of assessment ; formerly schools or neighbourhoods had been compared with each other. Correlation procedures were used for the first time to express relationships ; this was to become typical of later studies. And it was much easier to express differences in terms of IQ than in terms of the slightly clumsy concept of mental age.

It will be seen from the testing studies reported in this chapter that the problem of social differences in performance was/

/was one of the chief matters of concern in the period of early development. After a few years of indecision, the belief seems to have hardened that these differences were in the main innately determined. Practical considerations aided the popularisation of this belief, which in any case fitted well into the framework of popular opinion. Racial differences in performance had also by this time been identified, and there was a strong tendency to believe that these reflected the natural superiority of white and the natural inferiority of coloured races. If environment could have been held to account for apparent differences in intelle--ctual ability between social groups, then it would have been much more difficult to argue for hereditary racial differences. On the other hand, as evidence was accumulated to weaken the case for hereditary racial differences, the case for heredit--ary social differences was also weakened. Thus, after an initial acceptance of their points of view, hereditarians such as Terman found themselves engaged in a lengthy and often bitter controversy with colleagues who maintained that both racial and social differences in intelligence test perform--ance were chiefly determined by environment.

CHAPTER VII.

LATER RESEARCHES ON THE DEGREE OF RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT.

Since the end of the first World War a very large number of investigations have been carried out into this problem of the relationship between intelligence test performance and social background. Comprehensive experiments were made easier through the appearance of group intell--igence tests, which enabled large numbers of subjects to be tested at one time ; various methods of measuring social status were also developed, and these will be explained where this is relevant to the general discussion.

At intervals investigations have been summarized and analysed, notably by Stoke and Lehman (1930), Schwesinger(1933), Lorimer and Osborn (1934), Cattell (1937), Neff (1938), Loevinger (1940), Fleming (1943) and Eells (1948). Accumulation of evidence, however, has not brought about agreement as to its interpretation. On the one hand Cattell (1940, p.225)/

/(1940, p.225) states his belief that Burks' assessment of 80 per cent of IQ variance as due to heredity assigns too much importance to environment ; on the other Neff declares, (1938, pp.754-755) "All.....of the mean difference in IQ found to exist between children of the lowest and highest status may be accounted for entirely in environ--mental terms."

It is only fair to say that, while regarding Cattell's statement as extreme, the majority of psychologists would definitely not accept Neff's point of view. They would assert the superior influence of heredity. Nevertheless the existence of a large minority of environmentalists is a fact to be borne in mind. In part this is due, not to technical differences of opinion in the interpretation of data, but to social and personal factors of the type described above. However, many of the researches do offer considerable scope for such differences of opinion ; often the very results which an experimenter adduces to prove hereditarian hypothesis can be interpreted by a hostile reviewer in favour of an environmental hypothesis, and vice versa.

Researches reported in this chapter are those which have measured the difference in test performance and social status in terms of a coefficient of correlation. These researches in which social environment has been analysed into various factors/

/factors or in which an item-analysis has been carried out have been reserved for a fuller treatment at a later stage (Chapters VIII, IX), although they may be mentioned in passing.

It will be found, moreover, that discussion is fairly strictly confined to studies having a direct bearing on the stratification aspect of social environment. The following are thereby excluded from consideration, at least in any detail ;- (1) researches dealing with differences between urban and rural subjects ; (2) researches dealing with intra-familial resemblances in intelligence ; (3) researches dealing with the relationship between intelligence and size of family ; (4) researches dealing with the effects of foster-home training or the placement of foster-children ; (5) researches dealing with the general effect on intelligence of variations in environment ; (6) researches dealing with the effect on intelligence of malnutrition or deprivation of essential dietary constituents ; (7) researches dealing with differences in school achievement between children of different social backgrounds ; (8) researches dealing specifically with the test performance of specially depressed or educationally backward groups (e.g. Gordon's canal-boat and gypsy children or Hirsch's Kentucky mountain children.)

Even this process of exclusion leaves a very large number/

/number of studies for consideration. For that reason a purely historical mode of treatment would be complicated and difficult to follow. The discussion will therefore deal with researches in the following order ;- (i) researches on pre-school children ; (ii.) researches on children of school age ; (iii) researches on students at establishments of higher education ; (iv.) researches on adults ; (v.) researches on intellectual deviates (a) superior deviates (b) inferior deviates.

(i) Researches on pre - school children

The most careful early examination of the test performance of pre-school children was carried out by Goodenough(1920). She tested 380 children between the ages of 18 and 54 months in Minneapolis. Her measure of intelligence was the Kuhlmann Revision of the Binet tests, administered twice, with an interval of six weeks between the two testings. Children were classified by occupation of father into six socio- econ--omic groups in accordance with the standards of the Minnesota Scale of Occupational Intelligence. (This scale divided up the population into six categories, namely Professional, Semi-professional and Managerial, Clerical, Skilled Trades and Retail Business, Semi-skilled, and Unskilled Day Labourers. It was an early variant of the better-known Minnesota Scale for/

/for Occupational Classification, which also took into account rural occupations). For both administrations of the test she found that a definite occupational hierarchy emerged, although this was more evident on the second testing. The mean Kuhlmann- Binet IQ for children in the Professional group was 116 on the first testing, 125 on the second ; for children in the Unskilled group it was 96 on both testings. The same hierarchical pattern held constant at all ages, although for younger children the differences between the occupational groups tended to be smaller.

Furfey (1928) used a younger age group for his investi--gation of this problem. His experimental subjects consisted of 45 to 49 infants at each of ages 1, 2, 4, 6, 9, and 12 months. As a test of intellectual development he employed the Linfert-Hierholzer Infant Scale. As a measure of socio-economic level he employed the Chapman-Sims Scale. (This scale was developed in 1925. It is the earliest composite socio-economic scale, utilising both occupational and non-occupat--ional information. From a questionnaire concerning the cultural, economic, educational and occupational status of a family a score rating is computed which then defines the family's socio-economic level in the community.) He found that no significant relation existed between the scores of infants at any age on the Linfert-Hierholzer Scale and the ratings/

/ratings of parents on the Chapman-Sims Scale. Retesting 60 of his group at an average age of 56 months with the Stanford-Binet, he obtained a small and unreliable negative correlation between their Stanford-Binet scores and their Linfert-Hierholzer scores. At this higher age level, however, there was some indication that Standord-Binet score was related to Chapman-Sims rating.

In a less carefully planned study including pre-school children between the ages of 3 and 5 years Kawin (1934) found a difference of 18 points of Stanford-Binet IQ between a group of high socio-economic level and a group of low socio-economic level. The difference between the same groups in terms of Merrill-Palmer IQ was 5.7 points. (The Merrill-Palmer test is a performance test for pre-school children.)

Coffey and Wellman (1936) have published a detailed account of testing carried out on 417 children entering the pre-school laboratories of the Iowa Child Welfare Research Station between 1921 and 1934. At entry these children were between three and four years of age. Their intelligence was measured either by the Kuhlmann-Binet or the Stanford-Binet ; results for the two tests were not treated separately. They were to some extent a selected group; the majority were the children of professional parents, and no children of unskilled manual workers were included. When they were divided up into socio-economic/

/socio-economic categories by means of the occupations of their fathers, no very consistent picture of differences emerged. A difference of 8 points in mean IQ between the children of professional men and the children of clerical and skilled manual workers was not significant, although this may have been due to the small number of cases in the latter category. There was, however, a significant, though relatively small, difference between the mean IQ of children in the professional category and that of all the other children.

Undoubtedly the most thorough and important invest--igation of the relationship between test score and socio-economic level among pre-school children is that of Bayley and Jones(1937). They carried out a series of tests on 61 children born in Berkeley, California in 1928 and 1929 from infancy to six years. Various measures of intelli-gence were employed. The California First-Year Mental Scale was administered several times during the first year. During the period from 18 to 60 months the California Pre-school schedules I and II were administered on twelve occasions. At six years a Stanford-Binet IQ was taken for each child. In addition at 66 months a vocabulary test from the Thorndike CAVD test and a series of form boards were administered. The families of the children were rated on/

/on The California Socio-Economic Index. (This is a complex socio-economic scale which consists of a weighted average of the following factors : family income translated into logarithmic scale scores ; years of formal education of the father ; years of formal education of the mother ; a rating of the father's occupation ; and a combined rating of home, living room, and neighbourhood.) Their social circumstances were slightly superior to the average of the community, but not sufficiently to render findings invalid.

Bayley and Jones found that in the first six months of life correlations between test score and socio-economic ratings were usually negative, that from six months to eighteen months they hovered around zero, and that from six months to six years they became increasingly positive. When the particular correlations between test score and individual environmental variables making up the Index were computed, a consistent order of closeness of association emerged which did not vary from 21 months to six years : viz., formal education of mother, formal education of father, occupation of father, home and neighbourhood rating, and family income. At 30 months the following correlations were obtained:- between test score and the total socio-economic scale .20 ; between test score and father's occupation .25 ; between test score and family income .08. The respective correlations obtained/

/obtained at six years for the same associations of variables were .41, .58, and .59.

Since this clear demonstration of the relationship between social background and test performance in the early years of life little need appears to have been felt for further investigation. However, for the sake of completeness, a few other studies will be reviewed.

The study of Dubnoff(1938) is remarkable as one of the few which have been carried out inside the Soviet Union. Using family income as a measure of socio-economic level, he tested three samples of children at ages 3, 6 and 10 months with the California First-Year Mental Scale. At the successive ages correlations between test score and family income were .37, .44, and .44, the probable error in each case being around .07 . This result is in some conflict with the findings of Furfey and Bayley and Jones; assuming the accuracy of Dubnoff's figures, it may be explained by differing principles of distribution of income in the Soviet Union and "capitalist" countries.

The most representative sampling of pre-school children, at least over the age of two years, has been that undertaken by Terman and his associates in connection with the standard--isation of the revised Stanford-Binet (Terman and Merrill 1937). 662 children/

/662 children between the ages of 24 and 66 months were classified by occupation of father into the seven categories of the Minnosota Scale for Occupational Classification. The following table reproduces the distribution of mean IQs given in Table 12 of measuring Intelligence (p.48)

TABLE 1.

MEAN I.Q.'s ACCORDING TO FATHER'S OCCUPATION.
(Terman and Merrill,1937)

Father's Occupational Classification.	Chronological Age. 2 - 5½
1. Professional.	116.2
2. Semi-professional and Managerial.	112.4
3. Clerical, skilled trades and retail business.	108
4. Rural owners.	99.1
5. Semi-skilled, minor clerical, minor business.	104.3
6. Slightly-skilled.	95.1
7. Day labourers, urban and rural.	93.6

This pattern of results compares closely with that of several smaller studies, and notably with that of Goodenough(1927).

An investigation very similar to that of Bayley and Jones/

/Jones, but carried out on a larger sample, is that of Honzik (1940), at the Institute of Child Welfare in the University of California. 252 children born in Berkeley, California within the period January 1928 to June 1929 were tested at intervals between the ages of 21 months and 8 years. The mental tests used were the California Pre-school Schedules I and II, between the ages of 21 months and 5 years ; the Stanford-Binet at 6 and 7 years and the Terman-Merrill, Form L, at 8 years. The homes of the children were rated individually by a socio-economic index based on 4 items : (1) an average of social ratings of house exterior, neighbourhood, living space, family accommodation, and special equipment ; (2) education of the parents ; (3) father's occupation ; and (4) family income. These items were taken as contributing 30, 17, 20, and 33 per cent respectively to the total rating.

Honzik discovered that the socio-economic rating of the home showed only a negligible relation to the children's test scores at 21 months ; but at $3\frac{1}{2}$ years there was a statistic--ally significant positive relationship between test score and socio-economic rating, which increased steadily up to 8 years. The correlations between score and rating at 21 months, $3\frac{1}{2}$ years and 8 years were respectively .04, .24, and .41. There was a specially marked increase in correlation between 3 and $3\frac{1}{2}$ years.

Honzik/

/Honzik also computed mean IQs for children in seven occupational groups at ages 6, 7, and 8. She pointed to a tendency for the IQs of the children of professional men and business executives to increase steadily from 6 to 8, and an opposite tendency for the IQs of the children of unskilled labourers to decrease.

One study which points a way for further investigation is that of Irwin (1948). He showed that infants reared in professional and white-collar families differed from those reared in working-class homes in the number of different types of sound they used and the frequency with which they used them, but that these speech and language differences became statistically significant only after the age of 18 months.

As a result of the investigations reported here, the way in which social environment and test score are associated in the early years of life appears to be well established, although experimenters may differ in detail. The consensus of opinion is that no clear sign of association of the two variables is evident until the second half of the second year, but that thereafter association increases steadily in quantitative extent up to the age of school entry ; similarly differences in mean IQ between groups of children belonging to different occupational categories are negligible in early infancy;/

/Honzik also computed mean IQs for children in seven occupational groups at ages 6, 7, and 8. She pointed to a tendency for the IQs of the children of professional men and business executives to increase steadily from 6 to 8, and an opposite tendency for the IQs of the children of unskilled labourers to decrease.

One study which points a way to further investigation is that of Irwin (1948). He showed that infants reared in professional and white-collar families differed from those reared in working-class homes in the number of different types of sound they used and the frequency with which they used them, but that these speech and language differences became statistically significant only after the age of 18 months.

As a result of the investigations reported here, the way in which social environment and test score are associated in the early years of life appears to be well established, although experimenters may differ in detail. The consensus of opinion is that no clear sign of association of the two variables is evident until the second half of the second year, but that thereafter association increases steadily in quantitative extent up to the age of school entry ; similarly differences in mean IQ between groups of children belonging to different occupational categories are negligible in early infancy./

/infancy, but increase noticeably as the children become older. The fact that different experimenters have obtained quantitatively different results is probably due to the non-comparability of their investigations. Representative samples of pre-school children are difficult to obtain ; tests employed at the very young ages are often inadequately standardized, often do not correlate well with each other, and are poor predictors of performance in later life. The testing of infants presents especially great difficulties, since testers can never be sure that their subjects are co-operating properly, and may find themselves passing general judgements of ability on the basis of subjective impressions.

The lack of a proven association between social environment and test score in the first 18 months of life is undoubtedly a weakness in the hereditarian case. Investigators of the so-called Iowa school have sought to prove that when children of this age group are placed in foster homes their test scores come to be more closely associated with their new social environment than with the social environment into which they were born. As has already been stated, (Chapter III., p. 14) much doubt has been thrown on their research techniques. The hereditarians simply say that the early lack of association springs from the inadequacy of infant tests and/or the immaturity of infant subjects.

(ii) Researches on children of school age.

Early studies in this field have already been reviewed (Chapter VI.) Many investigations have been carried out since the First World War. To a considerable extent they tend to duplicate one another. Twenty-seven references have been chosen for discussion because they seem necessary to the presentation of a complete picture.

One of the first researches to compare the test scores of children of school age with the occupational ratings of their parents was that of Ronsch (1921, mentioned Burks, 1928b). She administered either the Stanford-Binet or the Terman Group Test to groups of children of various ages in California. She obtained Barr occupational ratings for the fathers or for the mothers if the fathers' occupations were unknown. (The Barr Scale had been devised in the following manner. A list of 121 representative occupations had been drawn up, and these had been rated by a panel of twenty expert judges according to the degree of intelligence required by the work. The judgements of the experts had then been employed to construct a hierarchical occupational scale, in which each occupation received a rating between 0 and 20.) She found an average correlation of .55 between the IQs of the children and the Barr ratings of their parents.

It/

/It will be noticed that Rensch's investigation is the first so far noticed which used a group test to measure the intelligence of children. By modern standards her method is faulty, in that she combined in single correlations IQs based upon two different tests. Moreover, she took no account of the possible effect on her results of the existence of racial differences in her experimental group ; many of her subjects were Mexican children. Her study is of historical importance, but little reliance can be placed on her findings.

At this same period in England Thomson was engaged in the development of group intelligence tests. He reported a study in which subjects included 64 pupils in a poor suburb, and 111 pupils in a well-to-do suburb, of Newcastle. (Thomson, 1921) In each case subjects were taken from two different schools. For the two schools in the poor suburb the percentages of subjects above 100 IQ were 30 and 35 ; the comparable percentages for the two schools in the well-to-do suburb were 55 and 56. This study is reminiscent of the work of Binet's collaborators reviewed in (Chapter VI. pp. 45-48), in that the social gradings which it employs are very broad and not applied to individuals. However, in 1923 a much more ambitious investigation was reported (Duff and Thomson, 1923). The Northumberland Mental Test was given to all elementary-school children in Northumberland (except those in Newcastle/

/Newcastle and Tynemouth) who were over 8 and under 15 years of age, and also to many secondary-school children between those ages. Parents' occupations were obtained for 15, 419 of the subjects. 137 children were classified as having fathers in the professions ; for this group the mean IQ was 112. For the 1814 children whose fathers were unskilled labourers the mean IQ was 96. When occupational groups were subdivided in finer detail it was found that the highest mean IQ of 121 had been obtained by the children of clergy--men ; hawkers' and chimney sweeps' children had a mean IQ of 91 ; the children of insane and criminal parents had the lowest mean IQ of 88.

This latter study has formed the prototype of a great many others which have set out to investigate occupational levels of intelligence. Such a detailed occupational division has been characteristic, however, of research on adults rather than on children.

Another early, but unfortunately undetailed, reference to relevant research is contained in an article by Elderton (1922). This reported data collected in Edinburgh in which rather low correlations of .10 for boys and .16 for girls were obtained between test scores of children and the economic condition of their homes.

An investigation which deals with an older group of children/

/children is that of Book (1922). His findings do not show quite the same differences between occupational groups as those of other investigators ; this is perhaps due to the method of assessing differences which he employed. Having tested a large population of high-school seniors in Indiana he tabulated the percentages in various occupational groups scoring above the median of the group as a whole. 60 per cent of the children of professional parents scored above this median, compared with 47 per cent of the children of day labourers. But the proportion for the children of clerical workers was also 60 per cent. And that for the children of artisans was 55 per cent, higher than the 54 per cent for the children of fathers in posts of executive responsibility.

Dexter (1923) tested 2703 American children between the ages of 6 and 13 with two early group tests (The Dearborn Test and The National Test). The mean IQ of children from the professional classes was 115 ; that for children from the unskilled labouring classes was 89 . When the average IQs of children of fathers of different occupations were compared with the average Army Alpha scores of men of the same occupations from three army camps, rank order correlations were obtained of .74, .79, and .79 . These correlations are unusually high, but must be accepted with caution, since they are based on data from group averages.

Isserlis/

/Iserlis (1925, mentioned Burks,1928b) found that correlations of the order of .50 to .55 existed between children's intelligence, as measured by group tests and teachers' ratings, and various measures of environment, including social status. He concluded that a progressive improvement in home conditions might be expected to raise the intelligence of school children. As might be expected, this conclusion was forcibly criticised by Pearson (1925), who claimed that Iserlis' work suffered from crippling defects in statistical method. However, rather similar results were obtained in the United States a little later by Chapman and Wiggins (1925) . They administered the National Test to 632 children of over 10 years of age, and rated the social level of their parents by the Chapman-Sims Scale. They found a correlation of .52 between IQs and social ratings. On the basis of a review of this and other studies Lorimer and Osborn (1934) have concluded that the relationship between intelligence and social levels is probably best represented by a correlation of .50 . Of course, some experimenters have obtained much higher figures. Fukuda (1925) applied Stanford-Binet tests to 257 elementary-school pupils and graded their social environment by the Whittier Home and Neighbourhood Scale. (This scale is interest--ing as being one of the earliest to attempt to measure socio-economic/

/socio-economic status by non-occupational criteria). The score of a child here was the total of the scores of five factors (necessities in home, neatness of home, size of home, general parental conditions, and parental supervision,) each graded from 1 to 5 after a visit to the home). He found a correlation of .53 between IQ and social environment. The size of this correlation may be explained by the presence of a large number of first-generation American children in his samples.

The earliest investigation carried out in Canada seems to have been that of Sandiford (1926) . Since some of his subjects were university students, it is a debatable point whether his study should be referred to here or in the following section. From the point of view of sheer number of subjects tested it compares favourably with better/known investigations. He had an American Army Alpha test somewhat modified for the benefit of the younger age groups administered to more than 5000 high-school pupils, normal school trainees, and university students in British Columbia. When subjects were classified by paternal occupation, differences from one occupational level to another in median IQ made their appearance, but these were not very great. At the top 659 subjects whose fathers were professional men had a median IQ of 105.1 ; at the bottom 456 subjects whose fathers were unskilled labourers had a median/

/median IQ of 100.8 . There is no easy explanation of this relative lack of differentiation. Several possibilities may be canvassed. If it is assumed that individuals in a frontier society such as British Columbia are highly socially mobile, then the occupational level of fathers at the time of testing may not have represented their final occupational level. More prosaic explanations are that Sandiford was dealing with a pre-selected population, since all subjects were at least high school pupils, or that the test which he employed included too few difficult questions, thus giving little chance for differences to emerge.

The study of Jones and Carr-Saunders (1927) appears to have been the first large-scale investigation of the test performance of children in orphanages. They used the Simplex Group Test to measure the intelligence of approximately 800 children in eight English orphanages and industrial schools. Children varied in age from 9 to 15 years. They were classi--fied into five groups according to the occupations in which their fathers had been engaged.(Classification followed very closely the model of the five-category Taussig Scale, which is divided up as follows :- (1) Well-to-do, professional and business men and managers of industry ; (2) lower-middle-class, clerical and semi-intellectual occupations ; (3) skilled workmen ; (4) unskilled workers ; (5) day labourers.)

They/

/They were then further classified into two sections according to period of residence at orphanage or school. For the children who had been in residence up to three years the mean IQ of the two top occupational groups taken together was 107.5 , the mean IQ of the middle occupational group was 97 , and the mean IQ of the two bottom groups taken together was 94.5 . For the children who had been in residence over three years the occupational differences were slightly less ; the mean IQ of the two top groups was slightly lower and the mean IQs of the middle and two bottom groups were slightly higher.

A more comprehensive study of orphanage children was made by Lawrence (1931),and this is probably the best place to discuss those parts of it which are relevant to the discussion. She studied data made available to her by a single English orphanage which admitted only first-born illegitimate children less than one year of age, and which insisted among other conditions of entrance that the names and occupations of fathers should always be given. At about the age of 10½ years all the children took a Stanford-Binet test.

The occupations of fathers and the Stanford-Binet IQs were thus known for 160 boys and 109 girls. Occupations were graded according to a fivefold classification----peculiar in/

/in that professional people and all types of farmers were combined in a single category———— and correlation ratios were computed between IQ and occupational class. These were found to be .22 for the boys and .26 for the girls.

Lawrence then gave the older children the Simplex Group Test of Intelligence in order that the correlations with occupational class might be compared with those found for London elementary-school children living in their own homes. On this test the correlation ratio between occupational class and test score was .25 for 105 boys and .25 for 82 girls. For 211 elementary-school boys living with their own parents the correlation ratio was .27 ; for 236 elementary-school girls it was .32 . These results seemed to show that, as Lawrence herself put it (p.70) "the correlation (between the intelligence of children and the social class of their parents) is not mainly due to the direct social influence of the home, but to a genuine biological fact".

The method of classification of occupations and the correlation techniques used make this study difficult to compare with others, but there is no doubt that it further substantiates the association between social environment and IQ . If the findings of the "Iowa School" are to be accepted, however, Lawrence's results have no validity outside the group which she studied; they have brought evidence to show that the test/

/test scores of deprived children placed in good foster-homes early in life tend to become associated with their new social environments, and to bear little relation to those environments into which they were born.

Stoke (1927) chose for experimental study a group of 508 Boston school children. He classified fathers' occupation into five categories according to the Taussig Scale. Having tested the children by means of the Stanford-Binet test and the Dearborn Series IA test he also classified their IQs into five groupings : below 80, between 80 and 89, between 90 and 109, between 110 and 119, and 120 and above. He obtained a correlation of .30 between IQ and occupational level, which compares closely with that obtained in other studies. Dissatisfied with the appropriateness of product-moment correlation methods for his data, he also worked out the coefficient of mean square contingency between IQ and occupational level ; this time he obtained a coefficient of .557, the maximum possible being .898 . He concluded that the degree of real association between intelligence and social status was somewhat higher than had appeared from previous studies.

Stroud (1928) tested 1037 school children aged between five and eighteen in two counties of Georgia by means of the Pressey/

/Pressey Classification Test. His measure of economic status was a novel one. Since differences of occupational level within his group appeared to be very small, he classified children according to the amount of tax at which their fathers were assessed. He found a correlation of .35 between IQ and tax assessment. This is apparently the only case in the literature in which tax assessment has been used to define socio-economic position. It is probably best compared with those studies which have correlated IQs of children with family incomes. In that case the correlation appears rather high, and may be explained by the fact that subjects are not occupationally representative of the population as a whole but tended to belong to one occupat-ional group, that of poor farmers.

A study which is of special interest because it derives from work done in the Soviet Union in the period before the abolition of intelligence testing is that of Sirkin (1929). His subjects were 2580 elementary-school pupils in Kharkov. As a measure of intelligence he used total score on five tests : reading and comprehension, mixed sentences, analogies, opposites and similarities. He measured the home environment of the children in terms of the occupation and education of the fathers and the education of the mothers, putting each child/

/child into one of six categories on the basis of a sum of all three factors. He found that as home environment improved, test score consistently improved also. Within each of the three school grades to which the children belonged, the children from the highest social category were approximately one standard deviation above the mean score for their grade, while those in the lowest social category were about half a standard deviation below the mean score for their grade. For the three grades correl--ation ratios between test score and home environment were respectively .39, .43, and .40. No marked change in size of coefficient resulted from a computation of product-moment correlations.

As in the case of Dubnoff's investigation, which has already been reviewed (p. 64), the degree of association found between test score and social background is higher than that typical of many comparable studies in the United States.

The study in the literature which shows the greatest differences in test score between children of different occupational levels is that of Hildreth(1934), who tested 608 Oklahoma school children in Grades I to VIII with the Stanford-Binet. She employed the Taussig Scale to divide the children into five occupational groups(for definition of groups see p. 78). The median IQs for the different occupational classifications were as follows: 1, 113; 2,107.5; 3, 97.5/

/3, 97.5 ; 4, 84 ; 5, 75.7 . There was a sharp difference in mean IQ between adjacent occupational levels, and the difference between the children of professional men and the children of unskilled labourers amounted to 57.3 points of IQ, which appears very large when compared with the differences obtained by, for instance, Sandiford (see pp.77-78). Hildreth's experimental group does not appear to have been a representative one, since only 20 per cent fell into the two lowest Taussig categories ; there is thus a possibility that children in these two categories, for one reason or another, formed a sample biased towards low scoring on the test.

Byrns and Henmon (1936) have reported the testing results of an extremely large sample of 100, 820 high school seniors, examined between 1929 and 1933 in Wisconsin. The measures of intelligence employed were the Ohio State University Psycho--logical Test in 1929, the American Council Psychological Examination in 1930 and 1931, and the Henmon-Nelson Test of Mental Ability in 1932 and 1933. Scores on the three tests were made comparable by means of percentiles. Fathers' occupations were classified into the following groups, in order of the median intelligence of the children : professional, executive, business, clerical, miscellaneous, skilled manual, semi-skilled manual, unemployed, unskilled manual, farmers. The correlation between the scores of students and the ranks of/

/of fathers' occupations was .18 . This is unusually low, but may reflect the use of percentile scores.

The data assembled by Terman and Merrill (1937) has already been mentioned(p. 68) in connection with the review of researches on pre-school children. Their results were tabulated for four age groupings : 2 to 5½ years, 6 to 9 years, 10 to 14 years, and 15 to 18 years. Fathers' occupations were put into the seven categories of the Minnesota Scale for Occupational Classification. Differences in test performance between occupational groups at the three older age levels did not differ materially from those appearing at pre-school ages (already shown, p. 68). In each case children of professional men had the highest mean IQ, and the children of farmers and day labourers the lowest mean IQ. Only about 10 per cent of the day labourers' children exceeded the mean of the professional men's children and only about 10 per cent of the professional men's children fell below the mean of the whole standardization group.

The chief value of the results of Terman and Merrill is that they are based on the individual testing of the most representative sample of American children ever selected. But both psychologists have acknowledged that the sample was inadequate so far as rural occupations were concerned and slightly biased in the direction of superiority of occupat--ional status/

/status.

It will be noted that up to now no studies have been mentioned which studied directly the relation of test performance to social class. Most investigators have been content to define broadly socio-economic levels in a commun--ity or to group occupations into categories. In recent years, however, the writings of Warner have become preferred reading among American psychologists and a number of invest--igations have been reported which accept his anthropological approach to social class. Warner has suggested two methods of assessing the social class of experimental subjects (Warner, Meeker and Eells, 1949). The basic method, which he calls Evaluated Participation, involves the interviewing of large numbers of persons in a community. On the basis of the information which these persons give about the social position of other individuals in the community, a model of the social structure of the community is devised and its members fitted into a hierarchy of groups, each group consisting of people who participate, or may participate , in intimate social relationships with one another, but who do not associate freely with the groups that are socially defined as "above" or "below" them. The second method, much easier of application, requires the construction of what is called an Index of Status Characteristics, in which four objective criteria of social position are assessed, given weightings,/

/weightings, and combined to form a socio-economic scale. These four criteria are : (1) Occupation ; (2) Source of Income ; (3) House Type ; (4) Dwelling Area. In practice Warner found a correlation of .97 between Evaluated Participation and Status Characteristics indices when both methods were applied to the same group of subjects.

The first investigations into differences in test performance between social groups to employ Warner's Index of Status Characteristics were those of Havighurst and Janke (1944) and Janke and Havighurst (1945). Batteries of tests were given to ten-year-olds and sixteen-year-olds in a small city which they called Midwest. For the ten-year-olds the battery included the Stanford-Binet, the Cornell Coxe(a performance test), the Iowa Silent Reading, the Minnesota Paper Form Board, two mechanical assembly tests, the Porteus Mazes, and the Goodenough Drawing Test. For the sixteen-year-olds the battery included the Stanford-Binet, the performance tests from the Wechsler-Bellevue Scale, the Minnesota Paper Form Board, The Iowa Silent Reading, and the same two mechanical assembly tests.

There are one or two puzzling discrepancies in the details of this investigation which are not explained by Havighurst and Janke. They decided that the community in which they were working could be divided into five social classes/

/classes or social status groups, A, B, C, D, and E. Nearly all ten-year-olds and approximately 80 per cent of sixteen-year-olds in the community were tested. Thus it would be expected that the distribution of subjects among social classes would be much the same for both age groups. This is not so, however. There are no subjects in Classes A and B for the younger age group, while 8 per cent of the sixteen-year-olds fall into these two categories. Approx--imately 62 per cent of the ten-year-olds belong to Class D, compared with only 43 per cent of the sixteen-year-olds. The general picture seems to be of a ten-year-old sample biased in the direction of the lower social groups, compared with a sixteen-year-old sample biased in the direction of the upper social groups.

Havighurst and Janke obtained results comparable with those of earlier investigators. For both age groups there was a general tendency for test score to decline as social status went down. Differences between the two lowest social classes, D and E, were most significant for the ten-year-olds and showed up as high for this comparison of classes on every test in the battery except the Minnesota Paper Form Board (which measures judgement of spatial relationships) and one of the mechanical assembly tests (that designed for boys). For the sixteen-year-olds differences between the two highest social classes, A and B, and the others were the most significant,/

/significant, showing up as high on every test except the mechanical assembly test for boys. In fact on this test boys in Classes A and B obtained the lowest, and boys in Class E the highest, mean score, a result which Havighurst and Janke explained by the greater familiarity of lower-class children with mechanical tasks. But on the whole there was no tendency for differences between social classes to be greater on verbal than on non-verbal tests.

It is difficult to know how much weight to place on individual findings in this study, in view of the questionable representativeness of the social class sampling. Nevertheless the wide variety of tests administered makes it one of the most informative of recent investigations.

A later study in the same community by Schulman and Havighurst(1947) may be mentioned at this point. Using a test that enabled them to estimate the total vocabulary of subjects to whom it was administered, they again found significant differences between social classes. The average estimated vocabulary for Class B subjects was 45, 000 words, for Class E subjects 28, 800 words. This is a line of investigation which could be followed up with advantage; information of value to test constructors would certainly be provided. For instance, it would be useful to find what words are in common use among all social groups, and what words/

/words tend to be restricted in their use.

Although the researches reviewed so far have invariably dealt with social differences in test performance within a white population, it seems clear that the same results emerge from researches with coloured subjects. At a later stage (Chapter IX), a study will be considered which is concerned purely with social class differences among negro children. Here it suffices to notice the study of Robinson and Meenes (1947), who reported that the IQs of eight-and nine-year-old negro children tended to follow the order of the occupations of their fathers. Similarly Livesay(1944) found that in Hawaii, where the population is racially very mixed, the average test scores of high-school seniors were related to the income levels of their parents. The relative lack of data on this aspect of the problem of social differences in test performance is probably due to the fact that the vast majority of researches have been carried out in Britain and the United States, where the minority of coloured people tend to be employed in lower-level occupat--ions, so that there is difficulty in obtaining subjects from the higher social levels to compare with their less fortunate compatriots. There are of course a great many studies which simply compare the performance of American negroes with that of American whites, or compare the performance of negroes

who/

/who have emigrated to the Northern States with that of those who have remained in the South. But these are not directly relevant to the discussion.

The most comprehensive investigation of occupational differences among children in test score formed part of the Scottish Mental Survey of 1947. It has been reported in Social Implications of the 1947 Scottish Mental Survey (1953). Since the present study is based on data collected in the Survey, much of the relevant data is referred to at a later point (Chapter XI) and there is no need to consider it here also. The following bare facts, however, must be mentioned. In 1947 nearly all eleven-year-old children in Scotland were tested with a Moray House group intelligence test (Appendix I, pp.488-496). For 6687 children both test score and father's occupation were available. When these children were divided into occupational classes, and the mean test score for each class was computed, it was found that the order of the mean scores followed very closely the social order of the occupational classes. The possible score on the test was 76 ; the mean score of the children of professional men and large employers of labour was 51.8, while that of the children of unskilled manual labourers was 31.1 . High scorers were defined as children scoring over 60 ; 30 per cent of the children of professional men and large employers were high scorers, compared with only 2 per cent of the children/

/children of unskilled manual labourers. On the other hand less than one per cent of the first group scored less than 80, compared with 26.8 per cent of the second group. These figures illustrate as well as any that have been mentioned the extent of the differences in performance between social groups. They must not, however, be permitted to obscure the fact that many children in the lower social groups still obtain high or fairly high scores.

MacDonald (1958) examined social differences in test performance on the actual group intelligence test used in one of the "calibration" examinations in Glasgow in 1955. (The "calibration" examination is the term employed in Glasgow for the 11-plus examination which determines the type of secondary education which children are to receive ; Glasgow Education Authority devises its own group intell--igence and achievement tests.)

He divided schools in Glasgow into three categories, on the basis of the quality of housing accommodation available to children in the surrounding areas of the city. In terms of socio-economic level, this meant an approximate division of Glasgow into slum and near-slum areas, inhabited mainly by unskilled and semi-skilled workers, good working class areas, inhabited by skilled industrial workers, artisans and lower-level "white-collar" workers, and so-called "residential areas" inhabited by professional and managerial men/

/men and women and those with private incomes. The mean IQ for 185 children in schools in residential areas was 114.8 ; that for 207 children in schools in good working class areas was 106 ; that for 164 children in schools in slum areas was 99. These figures are specially significant in view of the attitude of the Education Authority in Glasgow that children with an IQ of 105 have a reasonable chance of tackling an academic-style secondary curriculum successfully. Clearly the majority of the children of unskilled and semi-skilled workers did not reach this minimum standard. In this investigation 77.8 per cent of the children in the schools in residential areas obtained IQs of 105 or over, compared with 52.2 per cent of the children in the schools in good working class areas, and 24.4 per cent of the children in schools in slum areas.

The combined evidence of these twenty-seven studies leaves little scope for general argument. They make the relationship of IQ or test score to socio-economic level, social class or occupational level among children of school age one of the best documented facts in the whole field of mental testing. However, as has already been indicated, the meaning of this relationship is still far from certain, and there has been much disagreement here. No further investigati--ons can contribute anything of scientific value which do not attempt/

/attempt a much more detailed analysis of relevant variables than has hitherto been typical psychological practice.

(III) Researches on students at establishments of higher education.

Few investigations have been carried out in this field. It has been recognised that the lower social classes are under-represented in the student population, but there are no certain indications as to whether this is due to inferior intellectual endowment or to inferior social and educational opportunity. Normal adult intelligence tests are not really suitable for administration to such a highly selected group of subjects. Certain special factors have also to be considered in assessing test results which assume great importance at this level of educational attainment ; these are illustrated in the findings of Lorge (1945) and Husen (1951). Lorge obtained the testing records of a group of 131 boys who had been given a number of psychological tests in 1921 - 22. Twenty years later he gave them the Otis adult intelligence test. He found that for subjects whose initial intelligence level had been the same scores on the Otis test varied consistently according to the amount of schooling/

/schooling they had received. Those who had received a university education averaged considerably higher than those who had gone no further than high school. Husen found that in Sweden adults who had received a full university education had a 12-point advantage over others, of the same initial IQ, who had left school at 15.

Nevertheless there are a small number of studies which show a similar relationship between intelligence and social background among university students to that obtained among other groups in the community. They all originate in the United States, where it must be that the university and college populations are more representative of the populations as a whole than in Great Britain.

Cuff(1933) tested 758 college freshmen by means of the American Council Psychological Examination. He rated their family background by the Sims Score Card for Socio-economic Status, a development of the earlier Chapman-Sims Scale (discussed on p. 63). He also obtained their examination ratings from the college registrar. He found that Sims socio-economic ratings correlated .24 with test score and .10 with examination performance. Although these correlat--ions are significant, they are low, and, as Cuff showed, meant that socio-economic level was little better than chance as a predictor either of intelligence level or of scholarship/.

/scholarship.

Glass (1936) did not employ correlation techniques in his study of 118 freshmen, to whom he administered the Thurstone Psychological Examination for College Freshmen, nor was his socio-economic classification so careful as that of Cuff. He divided up the freshmen into four occupational groups, according to whether their fathers belonged to the professional classes, the commercial classes, the skilled labour classes, or the agricultural classes ; no subject had a father who was an unskilled labourer. Then he computed mean percentile scores for each group. The mean score was 65 for those students from the professional classes; 58 for those from the commercial classes; 54 for those from the skilled labour classes ; and 45 for those from the agricultural classes. His results are not of much value to later investigators, since he does not test the significance of differences ; his use of mean percentile scores would in any case make the normal critical ratio test inapplicable. He does seem to have established in a general way, however, the existence of some occupational difference in test performance.

Canady (1936) investigated the performance of 441 negro freshmen of different occupational groups at West Virginia State College. He found that the occupational distribution of his subjects was very different from that usually obtained with/

/with white students; 78 per cent of his cases came from skilled or unskilled labour groups. This obviously reflected the generally depressed occupational status of negroes in West Virginia. He divided up his subjects into five occupational categories, according to whether their fathers were professional men, engaged in commercial undertakings, artisans, skilled workers, or unskilled workers. The method of classification is unusual ; most people would classify artisans as skilled workers.

The median test score for the 441 students on the American Council Psychological Examination was 84.5 . The occupational groups fell in the following order : professional, with a median score of 99 ; commercial, 95 ; artisan, 94 ; skilled labour, 88 ; unskilled labour, 75 . Again the general pattern of occupational differences is meaningful, but it is difficult to draw detailed conclusions from the actual figures obtained.

The most comprehensive investigation in this field is that of Haught (1938), although even it cannot compete in carefulness of planning with several of the investigations reviewed in the last two sections. He accumulated information of 3414 freshmen entering the University of New Mexico between 1921 and 1936. The Army Alpha test had been administ--ered to those entering between 1921 and 1927 ; entrants between/

between 1928 and 1936 had taken the American Council Psychological Examination. Scores for the two tests were made comparable by turning them into percentiles. Occupations of fathers were classified as professional, business and clerical, skilled labour, semi-skilled labour, and unskilled labour. The mean percentile score for the students with fathers in the professions was 60, compared with 57 for the business and clrical group, 50 for the skilled labour group, 45 for the semi-skilled labour group, and 35 for the unskilled labour group. Haught rather unwisely applied the critical ratio test to the differences between mean percentiles, finding that all differences between groups were significant above the 1 per cent level. This is a questionable procedure. But it seems unlikely, for the size of his sample, that chance can account for the differences which he obtained. It therefore appears that, at least within a university popul--ation which is not too highly selected, the same kind of differences in test score between occupational categories emerges as with children of school age. It is uncertain whether or not these differences disappear in the small and stringently selected student populations characteristic of the countries of Western Europe.

(IV) Researches on adults.

Adult subjects are not so readily available for testing as are school children. This partly explains why the attack on the problem of social differences in intelligence test performance has in the main proceeded indirectly by the comparison of children's test scores with parents' occupations or social level.

Three large-scale studies have been reviewed by Johnson (1948). The first was carried out by American psychologists testing large groups of soldiers with the Army Alpha test during the First World War. The second was carried out by Cattell in the mid-thirties in order to determine occupational norms for his own group intelligence tests. The third derived again from testing of American soldiers, this time with the Army General Classification Test during the Second World War.

In his original report of the Army Alpha testing Yerkes (1921) included data based upon the average scores of men in 114 occupations. It was shown that there was a pronounced decrease in test score as the social level of occupations declined, engineers and accountants obtaining the highest and casual labourers the lowest mean scores. Johnson(1948) classified these occupations into seven categories and made results comparable with those in other investigations by using/

/using the means and standard deviations to put the distributions on the same scale as the Stanford-Binet. Thus he obtained a mean IQ for the different occupational categories as follows ; Professional 123 ; Semi-professional and Managerial 119 ; Clerical, Skilled Trades and Retail 108 ; Rural Owners and Farmers 97 ; Semi-skilled and Minor Clerical 101 ; Slightly Skilled 98 ; and Day Labourers 93 . This conforms very closely to the occupational decline which has already been shown with children. There appears to be some tendency for professional men to score more highly than their children ; this could be explained as regression to the mean on the part of the children.

Some early analyses of the Army Alpha results have employed statistical methods which would now be regarded as erroneous. Alexander (1922) correlated the rank order of American States in average Army Alpha score with their rank orders in a number of socio-economic variables. The correlation of test score was .85 with average wage for farm labour ; .62 with per cent urban population in 1910 ; .64 with white literacy; .03 with per cent owning homes in States having over 50 per cent rural population. Bagley (1925) worked out correlations between the rank order of states by Army Alpha score and their rank order on such variables as proportion of population in Who's Who 1850-1880, per capita income, per cent capita circulation of/

/of magazines, etc., obtaining r's ranging from .73 to .91. The correlation of group averages in the manner of these two studies has been long recognized as greatly inflating the coefficients obtained. There would seem to be some relation between Army Alpha test score and the socio-economic variables selected for study, but it is impossible to give any definite meaning to correlations which have been so doubtfully ascertained.

The investigation of Cattell (1934) suffers from the defect that his subjects were not occupationally representative of the population at large, including no unskilled workers or casual labourers, and no farmers or farm workers. Testing with Scale III of the Cattell Group Intelligence Tests, he reported median IQs ranging from 151 to 78, with secondary-school teachers, physicians and surgeons at the top and hairdressers, upholsterers, welders and factory packers and sorters at the bottom. Scaling Cattell's results on the Terman-Merrill, Johnson (1948) estimated a mean IQ for the professional men in his sample of 132, which is considerably higher than that obtained in any other similar investigation. Conceivably the professional men in Cattell's group were more highly selected than in other studies.

American findings on the Army General Classification Test (AGCT) in the Second World War were reported by Harrell/

/Harrell and Harrell (1945) and Stewart (1947). They have provided extremely detailed occupational tables, in which the mean standard score, the median standard score, the standard deviation, and the range of score are given for a large variety of occupations. The following occupational groups all obtained mean scores between 120 and 128: accountants, lawyers, engineers, public-relations men, auditors, chemists, reporters, chief clerks, teachers, draughtsmen, stenographers, pharmacists, tabulating-machine operators, and bookkeepers. Teamsters, miners, farmhands, farmers, lumberjacks, barbers, labourers, truck drivers, weavers, cooks and bakers, crane-hoist operators, general painters, and tractor drivers all obtained mean scores between 87 and 100. (The mean of the AGCT is 100 and the standard deviation 20).

The figures for the range of score are specially interesting. No engineer, public relations man, chemist or reporter obtained a score of less than 100. On the other hand at least one teacher tested at a score of 76, and one stenographer at a score of 66. Scores for teamsters ranged from 46 to 145, for labourers from 26 to 145, and for truck drivers from 16 to 149.

Two facts appear to emerge from these figures. The first/

/first is that some high-grade occupations have a very definite lower limit of intelligence (if test score is taken as measuring intelligence) while others do not. The second is that occupation is no sure indication of intellectual level ; it is quite possible for an extremely intelligent man to be doing unskilled work and for a stupid man to be doing work which is usually regarded as being highly skilled. If all the men of high intelligence were engaged in occupations which made demands on their ability, and all the men of low intelligence were working on simple tasks, the differences between occupational means would be much greater than has been found to be the case. Bingham (1946) estimated that of all the men classified Grade 1 on the AGCT (scoring 130 or above) only about a quarter were college graduates and 5000 had not even finished grade school.

The findings on the group tests employed in the investigations reviewed above have been confirmed by other researchers. Although the 8000 subjects of whom Simon and Levitt (1950) kept testing records on the Wechsler-Bellevue were not occupationally representative of the population as a whole, a consistent occupational hierarchy/

/hierarchy of test score was obtained on both verbal and performance scales. Vernon (1947) recorded scores on the Progressive Matrices Test for all candidates for entrance to the Royal Navy over a period of four war years. This test is a non-verbal one, as free as possible from educational influences. Nevertheless, when mean scores for occupations were calculated, large differences were found between various occupational groups. For instance, the mean score for clerks was 42.96, compared with 36.19 for woodworkers, and 32.43 for labourers (the possible score being 60).

It is thus clear that researches among adults, although relatively few in number, show the same kind of social differences in test score as the more frequent researches with children.

(V) Researches on intellectual deviates
(a) superior deviates.

For the purpose of this review superior deviates are regarded as constituting that 1 per cent of the population testing at or above 130 IQ (Hollingworth 1940, p.58).

The major weight of information on superior deviates has been contributed by one massive study, carried out in California by Terman and his Associates over a period of some thirty years.

Terman's initial sample consisted of approximately 1000 pre-school and elementary school children and 300 high school pupils selected from a school population in California of about 250,000 on the basis of intelligence tests and teachers' ratings because they appeared to have IQs of 140 or over. (Terman et al., 1925). The whole group averaged 3.2 standard deviations above test means for the population as a whole. (Assessments of intelligence in later life showed that there had been some decline. Thorndike (1948) estimated that by the late 'forties the group averaged only 1.7 standard deviations above the population mean).

One of the most striking characteristics of this group/

/group of highly intelligent children was its occupational distribution. Fathers' occupations were obtained for 660 of the children and categorized on the Taussig Scale. 31.4 per cent of the fathers were professional men, 50 per cent worked in semi-professional or commercial occupations, 11.8 per cent were skilled labourers, 6.6 per cent were semi-skilled labourers, and only 0.15 per cent were unskilled labourers. (Note:- this last percentage would appear to mean that only one father was an unskilled labourer). There were more than ten times as many children of professional fathers as would have been expected had the test performance of all occupational groups been alike. When the social background of children was assessed by family income the same picture emerged. The median family income was 3333 dollars per annum, between two and three times as large as that for the general population in California in 1922-23.

Follow-up studies by Terman and Oden (1940a, 1947) have shown that the superior intellectual status of these children tended to be reflected both in their scholastic records and in their personal and occupational adjustments to society. Not all were successful, however/

/however. By the period of the 1947 report, 71 per cent of the males were engaged in professional or higher business fields, compared with less than 14 per cent of California males as a whole. But the remaining 29 per cent were distributed through a variety of occupations ranging down to unskilled labour. And occupational level and income was much lower for females than for males, although this would be expected.

In 1938 an attempt was made to discover the factors responsible for differences in level of adult achievement within this sample of gifted children. (Terman and Oden, 1940b). 600 of the males who were 25 years of age or older were classified into three groups on the basis of rated success. 167 fell into Group A, the most successful group ; 146 fell into Group C, the least successful group.

When Groups A and C were compared with respect to measured intelligence no reliable difference between them was found. On all the tests at various ages on which records have been kept the A Group had only a slight advantage. And no factor so strongly differentiated the two groups as initial social background. 50.7 per cent of the A fathers had been professional men, as compared with 21.1 per cent of the C fathers. 59.4 per cent of the A fathers had been engaged in occupations below the semi-professional and higher business levels, as compared with 58.7 per cent of the C fathers.

These/

These findings of Terman and his co-workers were largely confirmed at another age level by Lewis (1941). Her study formed part of a large research project in which 50,000 grade-school children throughout the United States were tested on the Kuhlmann-Anderson intelligence test, and both the 10 per cent scoring highest and the 10 per cent scoring lowest were singled out for special investigation. She contrasted the high ability children whose scores on school achievement tests averaged at least a year ahead of their mental ages with the high-ability children whose scores on such tests averaged at least a year lower than their mental ages. She found that a larger number of the over-achievers came from the professional classes, and a smaller number from the semi-skilled and unskilled occupational classes.

Other workers in this field have also reported that children of high intelligence tended to belong to the upper social and occupational groups in the community. Jenkins (1935) (reported Witty, 1940a) isolated 103 negro children possessing Stanford-Binet IQs of 120 or above in a survey of seven Chicago public schools enrolling approximately 8000 negro children. 29 of these children tested at IQ 140 or above; the mean IQ for the group was 134.2. 53.4 per cent of the fathers were well-to-do, professional/

/professional, or business men, and 56.5 per cent were engaged in lower middle-class, clerical and semi-intellectual occupations; only 6.5 per cent were day labourers.

Witty (1940b) reported results of a study of fifty gifted children with Stanford-Binet IQs of 140 or above in Missouri. The mean IQ of the group on first testing was 155, but later tests showed a tendency for intelligence level to fall. 64 per cent of the fathers were business men; 34 per cent, professional men. The average yearly income of the fathers in 1924-25 was 5,300 dollars compared with the average yearly income of 2,000 dollars of the fathers of a control group with IQs ranging from 90 to 110.

The general picture presented by the studies in this section is similar to that which emerged from investigations carried out with unselected groups of subjects. Parents from the upper social levels appear to contribute far more than their share of gifted children to the community. It would seem, moreover, that even when children of high intelligence are born to parents of low occupational status they tend to be unable to take full advantage of their ability.

(V)/

(V) Researches on intellectual deviates
(b) Inferior deviates.

Inferior deviates are defined here as that proportion of the population testing at or below 70 IQ. Different estimates have been put forward as to their true numbers. These are undoubtedly swollen by the existence of people who owe their condition to some clearly demonstrable pathological defect or birth injury.

Only a few investigations are relevant to the purposes of this review. Paterson and Rundquist (1933) studied the social origins of 823 residents and 516 applicants for admission at the Minnesota School for the Feebleminded in Faribault. Occupations were classified according to the Minnesota Scale of Occupational Intelligence. Either Stanford-Binet or Kuhlmann-Binet testing results were available for all the subjects. They found that 87.4 per cent of the fathers of the subjects belonged to the skilled, semi-skilled or unskilled labour groups, whereas professional, business, clerical and technical occupations contributed a very small number of cases relative to their numerical strength in the population at large. (This finding must be interpreted/

/interpreted subject to the proviso that wealthier parents more often send their defective children to private institutions). When subjects were categorized by IQ, however, an interestingly different occupational distribution was obtained for the different categories. For instance, 17.5 per cent of those with IQs between I and 20 fell into the occupational class of unskilled labourers, compared with 20.7 per cent of those with IQs between 21 and 40, 45.5 per cent of those with IQs between 41 and 60, and 65.2 per cent of those with IQs of 61 and above. Nearly all the morons (in American parlance mental defectives with IQs between 51 and 70) were the children of semi-skilled or unskilled labourers, while idiots and imbeciles tended to be distributed more or less at random among the various occupational classes. Paterson and Rundquist interpreted these results to mean that higher-grade feeblemindedness was far more a characteristic of the lower classes in the community, due to their inferior heredity, than of the upper classes, whereas pathological defect had an equal chance of occurring among any social group.

Bradway (1935) compared data obtained at Vineland Training School with the results of Paterson and Rundquist. A medical diagnosis of primary or secondary aetiology had/

/had been made on all her 439 subjects. In 116 cases there was no evidence of pathological defect, and aetiology was classified as primary. In 123 cases there was clear evidence of pathological defect, and aetiology was classified as secondary. For the remaining 200 subjects aetiology was so mixed that no classification could be made. Bradway found that the distribution of paternal occupations for the "primary" cases piled up towards the lowest occupational class, while that for the "secondary" cases approximated to a symmetrical curve. Her data thus supported the findings of Paterson and Rundquist.

Two interesting follow-up studies have been made of 560 morons in Nebraska who were given special adjustment training in school. In the first follow-up Baller(1936) was able to locate 95 per cent of the group. He compared their social adjustment records with those of a normal control group with IQs between 100 and 120. Adjustment was inferior in the case of the morons, as had been expected, but the large majority were living under perfectly normal circumstances. Of more immediate interest here are the details of the second follow-up study, reported by Charles (1953), made when the subjects had reached an average age of forty-two. Most were in steady/

/steady employment, and their jobs covered a wide range of skill and salary levels; they were by no means all in the lower occupational brackets.

This is not the place to discuss the studies of Kephart (1939) and Schmidt (1946) who reported striking improvements in the IQs of feeble-minded children following placing in a specially stimulating environment. If they are accepted, however, it will be necessary to set a new interpretation on the findings of Peterson and Rundquist and Bradway which would greatly minimise the importance of heredity.

Sixty-four studies conceived as having some direct or indirect relevance to the problem of social differences in intelligence test performance have been reviewed in this chapter. Almost without exception, no matter according to what principle subjects are selected, they agree in confirming the existence of an association between test score or IQ and occupation, socio-economic background, or social class after the early months of life. They contribute little, however, to a theory of causation. By making appropriate assumptions they can be quoted with equal justification by hereditarians and environmentalists. It may be that the information necessary/

/necessary to come to firm conclusions will never be available in the field of psychology itself and that psychologists will have to wait on future research by geneticists, physiologists and neurologists. However, there are still lines of investigation which look promising and in which surprisingly little has been done. In the first place social environment can be analysed in detail to discover which aspects of it are most significant in causing differences in test performance. In the second place the tests themselves can be analysed in detail to identify the items which are most influenced by differences in social invironment. Studies in which these methods have been employed are reviewed in Chapters VIII and IX.

CHAPTER VIII.

STUDIES OF THE RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND INDIVIDUAL FACTORS OF THE SOCIAL ENVIRONMENT.

Theoretical arguments may be put forward against the differentiation of social environment into its various components for the purpose of assessing particular relationships to test score. In the first place the major influence of social environment on test performance may emerge from the broad constellation of environmental differences which are known to exist; the analysis of these into detailed component parts may destroy the association which it is sought to explain. Secondly, even if this possibility is not accepted, the evidence of studies in which one variable is correlated with many others is difficult to interpret; for instance, where several/

/several environmental factors are found to correlate positively with test score this may well reflect, not so much the individual importance of these factors, but their common dependence upon another, more fundamental, factor.

One of the most important studies in this field is that of Burks (1928a). It has been mentioned before (p. 18) and attention has been drawn to criticism it has received. Nevertheless her findings are extremely interesting, whether or not her interpretations are accepted.

The immediate purpose of Burks' investigation was to compare foster-parent foster-child resemblance in intelligence with true-parent true-child resemblance. She therefore collected data concerning 214 foster-children and 105 children living in their own homes. Parents and children were tested by means of the Stanford-Binet test and home information sheets were completed for each child.

In this review attention will be directed only to one aspect of this study, the relationships found by Burks to exist between the test score of children living in their own homes and various environmental factors.

When/

When Burks computed product-moment or biserial correlations between children's IQs and nineteen environmental factors she found that private tutoring in music and dancing (for girls) was the most significant factor, with a correlation of .52, followed closely by father's vocabulary (.47), general cultural rating of the home (.44) mother's vocabulary (.43), tutoring in music and dancing for boys (.43) and Whittier Index rating of general home conditions (.42). Of mediocre importance was the number of books in the home and in the child's library (.34). Income of parents appeared to be relatively slightly associated with IQ (.24), while amount of home instruction before the age of six seemed of no importance whatever.

It is easy to think of many other variables than those considered by Burks, and also of ways of further subdividing those variables selected; for that reason her ordering of environmental factors by size of correlation with IQ can have no final validity. Nor is there any guarantee that in another investigation of a comparable group of subjects the size of some correlations might not be increased and that of others decreased, so that, for instance, mother's vocabulary correlated/

/correlated more highly with IQ than father's vocabulary. Nevertheless Burks does seem to have isolated, though not with any real clarity, some of the aspects of social environment which are mainly responsible for the found association with test performance.

It was decided not to review at any length here the investigations of Freeman, Holzinger and Mitchell (1928) and Leahy (1935), which bear a considerable resemblance to that of Burks.

The authors of the Freeman study selected for investigation a group of 401 foster-children in Illinois. Intelligence was measured by means of the Stanford-Binet test. Foster-home ratings were made by field workers who gave equal rating to six factors; material environment, evidence of culture, occupation of foster-father, education of foster-father, education of foster-mother, social activity of foster-parents. The order of environmental factors by size of correlation of IQ was as follows: general home rating (.48), mid-parent education (.48), father's occupation (.37), mother's vocabulary (.37), mid-parent vocabulary (.36), and father's vocabulary (.27). Since these findings refer to children in foster-homes, they are not quite relevant here; any discussion immediately/

/immediately merges in the heredity-environment controversy. Freeman argued that they showed that the intelligence of children depended to a large extent not on heredity, but on home conditions; his critics suggested that selective placement, or the matching of child to foster family, had taken place.

Leahy compared a group of 194 foster-children with a group of 194 children living in their own homes. All the children took the Stanford-Binet test. Environmental status score and cultural index of home were based on special questionnaires. Father's occupation was rated on the Minnesota Scale for Occupational Classification. The following order of environmental factors by size of correlation with IQ was obtained for children living with true parents; mid-parent vocabulary (.56), mid-parent education (.54), cultural index of home (.51), mother's education (.50), mother's vocabulary (.49), father's education (.48), father's vocabulary (.47), father's occupation (.45), and economic status of home (.37). Correlations for foster-children were much smaller, as was also the case in Burks' study, but followed much the same order.

It will be noticed that, where the two sets of figures/

/figures are comparable, Leahy's correlations are generally higher than those of Burks, sometimes considerably so. This may be explained by the fact that Leahy's children were less well distributed occupationally; there was an excess of foster- and true fathers in the higher occupational categories. On the other hand, it may also reflect the inevitable differences between samples when so many environmental factors have to be taken into account.

The findings from these three investigations reproduced here have been chosen for comparison with other studies and do not represent the full contribution of Burks, Freeman and Leahy to heredity-environment research. It was felt that a fuller treatment would have exceeded the limits of discussion as defined in Chapter VII (p.61).

A study which investigates the relationship between intelligence and environmental variables for pre-school children is that of Van Alstyne (1929). She obtained a sample group of 75 children within three months of their third birthday, all alike in that they lived with their own parents in unbroken homes in an urban environment. Their intelligence was measured by the Kuhlmann-Binet test. Socio-economic level was assessed by the Chapin Scale/

/Scale for Rating Living-Room Equipment and by a questionnaire including a large number of environmental items selected by experts as probably related to intelligence. The correlations which she obtained between intelligence and various environmental factors are higher than those reported by any other investigator. The correlation of child's mental age with questionnaire rating was .61; with mother's education .60; with Chapin Scale Rating .59; with mother's vocabulary .54; with father's education .51; and with father's occupation .50.

This investigation is most readily compared with one which has already been noticed (p.65), that of Bayley and Jones (1937). They obtained the following correlations of intelligence-test scores with environmental factors at 2½ years of age; test score with mother's education .48; with father's occupation .35; with California Socio-Economic Index rating .30; with home and neighbourhood rating .19; and with family income .08. They also computed correlations for children of six years of age; they discovered that while in every case the size of coefficients had increased, this was more marked for some environmental factors than for others. These correlations/

/correlations were: test score with mother's education, .58; with California Socio-Economic Index rating .41; with father's occupation .39; with family income .33; and with home & neighbourhood rating .29.

These are the only two comprehensive studies so far carried out of the relationship between particular environmental factors and intelligence in pre-school children. They are not altogether in agreement. They do not tell us very much other than that some environmental variables are more closely associated with intelligence test performance than others. It does seem, however, that parental education, and especially mother's education, has been established as an important factor, as also emerges from a glance at the results of Freeman and Leahy. Unfortunately, parental education cannot be considered an environmental variable in any true sense of the word, since it reflects to some extent parental intelligence. This kind of stricture also applies to the other environmental variables which have been isolated, even to father's occupation, but probably in a lesser degree. So the general pattern of results, in which objective indices of socio-economic level, such as parental income and occupation, correlate on the whole less/

/less highly with child intelligence than 'cultural' indices like parental education, parental vocabulary, and cultural rating of home, may simply be explained as showing a close association between child and parental intelligence, and a lesser association between child intelligence and objective economic variables. In that case, all that is learnt from investigations of this sort is that some aspects of the home life of children are more dependent on parental intelligence than others.

Two studies, those of Cuff (1933,1934) and Osborn (1943), have tackled the problem of the relationship between intelligence and individual social variables in a population of university students. Cuff's investigation has already been mentioned; it will be discussed here in more detail.

Cuff's subjects were 758 college freshmen who had been tested by the American Council Psychological Examination and given score ratings on the Sims Score Card for Socio-Economic Status. He divided them into quartiles by test score and calculated the percentages in the highest and lowest quartiles "possessing" each of 23 items in the Sims scale. Then he subtracted for each item the percentage in the lowest quartile from the percentage/

/percentage in the highest quartile, obtaining a series of percentage differences which he used to rank the items in order of importance as predictors of intelligence.

There are several flaws in this experimental method. Firstly, several items in the Sims scale do not demand a direct yes-or-no answer; in such cases it is not clear how Cuff determined the percentages "possessing" the item. Secondly, it seems unlikely that the distribution of the traits underlying the Sims items is rectangular; percentage differences do not therefore mean the same thing at different levels. It is impossible to decide how far this error has invalidated Cuff's findings; since at no point does he state the original percentages from which differences between quartiles were derived.

The following table reproduces the order obtained by Cuff for the 25 Sims items by size of percentage difference between top and bottom quartiles:

Table II.

ORDERING OF 25 SIMS ITEMS BY DEGREE OF ASSOCIATION WITH INTELLIGENCE TEST PERFORMANCE (Cuff 1934).

Sims Items	Rank Order.
About how many books are in your home?	1
Did your mother go to high school?	2

Do/

Sims Items	Rank Order.
/Do you belong to any organizations or clubs where you have to pay dues?	3
Did your father go to high school?	4
Where do you regularly spend your Summers?	6
Did your mother go to college?	6
Have you a telephone in your home?	6
Does your mother belong to any clubs or organizations of which you know?	8
How many servants do you have in your home?	9
Do you have a bathroom that is used by your family alone?	10.5
Did your father go to college?	10.5
How many magazines are regularly taken in your home?	12
Is your home heated by a furnace in the basement?	13
Write your father's occupation on this line.	14.5
How many rooms does your family occupy? How many persons occupy those rooms?	14.5
Does your family attend concerts?	16.5
Do you take private lessons in music?	19.
Do you take private lessons in dancing?	19
Does your mother, or the lady of the house in which you live, regularly attend any lecture courses of which you know?	19
How often do you have dental work done?	19
Does your family own an auto which is not a truck?	21.5

Do/

/Do you have a bank account in your own name? 21.5
Does your home have a room in which to study? 23.

Even accepting Cuff's method of arriving at this rank order of variables, it is difficult to decide which items are to be regarded as discriminating significantly between the top and bottom quartiles. If it is assumed that items in the first half of the order probably do discriminate significantly, then there is some evidence that the results of this investigation conform in general to those obtained by other researchers, since again the variables which are most closely associated with intelligence test performance are 'cultural' rather than 'economic', mother's education taking a high place. On the other hand, the importance of private tuition in music and dancing seems to be considerably lower than that obtained in Burks' study; but it is probably unjustifiable to compare a population of university students with a population of schoolchildren. Final interpretation of the findings is rendered difficult by the undoubted existence of intercorrelations between variables. And as has been said in comment on other studies, some variables are more weighted with parental intelligence/

/intelligence than others; if a high correlation is assumed between child intelligence and parental intelligence, then the order in which Cuff places the variables may simply reflect the extent to which they are associated with parental intelligence. Arguments on the basis of experiments such as this one are thus liable to become circular. If there was some way of 'partialling out' the relationship between child and parental intelligence, it might be that the 'economic' variables would be found to take precedence over the 'cultural' ones.

The apparently low association obtained by Cuff of Child intelligence with occupancy rate of home is rather surprising in view of the different results arrived at by other investigators, as, for instance, Fleming (1943). Having measured the intelligence of 771 Scottish schoolchildren aged between 8 and 13 years, by means of a special test, she related their test performance to the number of rooms in their homes. Dividing homes into categories according to whether they contained 1 room, 2 rooms, 3 rooms, or 5 and more rooms, she computed the percentages above the 75th percentile in test score for each category. 9 per cent of children living in one-room homes scored above this level/

/level, compared with 18 per cent of those living in two rooms, 34 per cent of those living in three rooms, and 68 per cent of those living in five or more rooms. Although she did not subject data to further analysis, there appeared to be a marked association between test performance and number of rooms in home; but again it might be argued that this simply meant that the more intelligent parents lived in better homes.

The only other major piece of research on the relationship between a number of environmental factors and test performance which the present investigator has been able to track down is that of Osborn (1948). Again his subjects were university students, 880 freshmen in three American Midwestern colleges. The measures of intelligence were different for each college, being the American Council Psychological Examination for 1938 and 1940, and the Ohio College Association Examination; scores were regarded as comparable. Data were gathered on age, sex, birthplace, place of residence, extent of travel, high-school background, means of support in college, vocational plans, age of parents, country of parents' birth, schooling of parents, employment of parents, income of parents, type of house, availability of modern conveniences, reading/

/reading materials in home, church attendance, and family mobility.

Osborn divided the freshmen into farm and town students within each college and also within subcurricular divisions of colleges. He computed mean percentile rankings on the intelligence tests for both categories. He found that the rankings of town students were invariably higher than those of farm students. Unfortunately, however, he did not apply tests of significance; the significance of differences between percentile ranks is in any case difficult to assess by any normal statistical procedure.

Of more interest is his method of establishing relationships between detailed environmental factors and test performance. He divided students within each college and within subcurricular divisions of colleges into quintiles by test score. Then he compared the distribution of various socio-economic and cultural traits among the quintiles. Although he does not give his findings in any detail, he stated that there was a general tendency for the students superior in test performance to have better-educated parents with larger incomes, and to belong to smaller families which owned more and newer modern conveniences, enjoyed more literary life, and travelled more/

/more extensively.

For three groups of students (town students in three colleges) Osborn gave a slightly more detailed analysis of the relation between the distribution of individual test scores into quintiles and the distribution among the students of fifteen different environmental factors, one of which was occupation of father. In one college, 29 per cent of the fathers of students in the highest quintile were professional men, compared with 19 per cent of the fathers in the lowest quintile. Figures for the other two colleges were similar. Osborn apparently did not attempt to combine subjects from all three colleges to obtain combined totals.

The other fourteen factors which Osborn listed as showing the greatest differences between quintiles were parents' use of public library, employment of students while at high school, median income of parents, frequency of attendance at movies, lapse of time between high school and college, employment of mothers, time devoted to daily newspaper, degree of economic support by parents, number of magazines read, number of times parents had moved house, vocation planned in profession, vocation planned in teaching, number of books in family library, and/

/and academic standing in high-school classes. He stated that only small differences existed on other factors, and did not detail them.

This investigation is too badly planned to be considered a really important contribution to the elucidation of the relationship between intelligence test performance and factors of the social environment. A major defect is the lack of tests of significance. But the most serious criticism is that it does not advance knowledge, that it does not show a move forward from the position reached in earlier researches. Admittedly it lists some additional environmental factors which appear to be positively associated with intelligence. Researchers could go on making up lists of this kind for a long time, and debating why one list conflicted with another. Such activity is unprofitable in default of a more general agreement as to how results are to be interpreted. There are too many points in Osborn's work on which differences of opinion can arise; it seems clear that his conception of the problem to be investigated was not sufficiently well-defined.

It will be seen that the studies reviewed in this chapter throw little fresh light on the relationship between/

/between measured intelligence and social background which the investigations discussed in Chapters VI and VII have shown to exist. The analysis of social environment into its component parts, and the estimation of their individual relationships to intelligence, has seemed a promising procedure, but has not led to any very definite results. As much as anything this has been due to the sociological naivete of psychologists. They have not realized the importance of firm definition of the major dimensions of social environment. Hence they have been unable to set their findings in a proper theoretical framework. Although there is still much room for research in this field, future investigators would be well advised to work within a consistent sociological frame of reference.

CHAPTER IX.

ITEM-ANALYSIS STUDIES OF THE RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT.

Although item analysis is now recognised as a major tool of the test constructor, only in a few cases has it been employed as a method of approach to the present problem. Thirteen studies are reviewed in this chapter where some form of analysis of items has been used in an attempt to particularize differences between social groups in test performance.

The early experiments of Binet, in which, in collaboration with Thevenot and Morle, he carried out a number of investigations in the schools of Paris into social differences in performance in his 1908 Scale, have already been discussed in Chapter VI. He had been led in the first place to consider the possible effect of social conditions on test performance by the work of Decroly and Degand ; they had applied the 1908 Scale to/

/to children in a private school in Brussels, obtaining norms for every age group above those set by himself. Re-analysing data sent to him by Decroly and Degand, he found that on the average the Belgian children were a year and a half in advance of his Parisian subjects; he inter--preted this advantage in terms of the higher social position of their parents.

In order to determine "which aptitudes are most favoured in the education of a rich child", he divided sub-tests in the Scale into two classes (Kite,1916, pp.319-321). In the first class he put the sub-tests where the Belgian children had an advantage of more than a year and a half. Among these were the repetition of 5 figures, the naming of 4 colours, the knowledge of pieces of money, the arrangement of weights, the naming of months, the repetition of long sentences, the naming of the days of the week, and the definition of abstract words. The second class included the sub-tests where the Belgian children had an advantage of less than a year and a half. Among these were the copying of sentences, counting backwards, writing from dictation, copying a diamond, giving change from 20 sous, finding rhymes, and putting missing words into sentences.

Binet interpreted this dual classification to show that/

/that children of the upper classes, by virtue of their home training, were especially superior in the sub-tests demanding verbal facility, but that their advantage was less on the sub-tests closely associated with school activities. In other words, he admitted that what nowadays would be called "cultural" factors exerted an influence on test performance. He did not regard his tests as thereby invalidated; it must be remembered that their original purpose was merely to find those children who were incapable of profiting by normal school instruction. He probably did not foresee a time when intelligence tests would be used to assist selection for secondary education. Nevertheless, in his 1911 Revision of the 1908 Scale, the alterations which he made suggest that he attributed some importance to the information derived from the Brussels study. He says(Kite,1916,p276)."There are tests which require a knowledge outside the intelligence of a child. To know his age, count his fingers, recite the days of the week indicate that he has learned these little facts from his parents or friends ; we have thought it well to suppress these three tests." And again,"There are tests too exclusively scholastic, as that of reading and retaining a given number of memories of what has been read, or copying a written model, or writing from dictation. We suppress these."

It/

/It is perhaps unfortunate that Binet did not live long enough to find that, in spite of his amendments, the 1911 Revision discriminated between children of different social levels in very much the same way as the 1908 Scale had done. In that event intelligence testing might have followed a different path; few of his immediate successors appeared at all conscious of the importance of social environment.

Stern's Breslau investigation has also been referred to in Chapter VI. He found that nine-year-old children in a Vorschule(school for upper-and middle-class children) scored as well on the tests in the Binet-Simon series as ten-year-old children on a Volkschule(school for working-class children).(Whipple,1914). Carrying the investigation into more detail, he compared performances in the two types of school both on the sub-tests appropriate to the age level of the children and on those designed for administration to eleven- and twelve-year-old subjects. Results were surprising. Nine-year-old Vorschule pupils passed 77 per cent of the sub-tests "at age" and 64 per cent of the sub-tests "above age"; nine-year-old Volkschule pupils passed 81 per cent of the sub-tests "at age" but only 34 per cent of the sub-tests "above age". The ten-year-old Volkschule children passed 86 per cent of the sub-tests "at age", compared with 46 per cent/

/cent of the sub-tests"above age". Clearly the superiority of the upper-class children was not evident in their performance of tasks appropriate to their age but in their mastery of tasks above their age level. The more difficult the sub-tests, the more marked the advantage of the Vorschule group.

This is not a finding which has been borne out by the results of other investigations, as will be seen when the studies of Eells(1948) and Macdonald (1953) are discussed. Macdonald found little association between the difficulty level of items and their tendency to discriminate between social groups. Eells found a general tendency for more difficult items to discriminate less between social groups than easier items.

Weintrob and Weintrob(1912), the results of whose inquiry have been detailed in Chapter VII (pp. 50-51), also made some attempt to analyse differences in the way in which different types of item were answered by children of different social classes. They divided items in the Goddard Revision of the Binet Scale into three categories : (1) language items ; (2) items testing ability to reason ; (3) items involving observation, sense discrimination, and counting or reckoning abilities. They discovered that on the language items children from the Hebrew Sheltering Orphan/

/Orphan Asylum and belonging to a low-status social group were superior to children from the Speyer School (catering for middle-class children) and children from the Horace Mann School (catering for upper-class children). On items testing ability to reason the Horace Mann School was only slightly ahead of the Orphan Asylum, and the Speyer School was behind both. On items involving observation, etc. there were no clear differences between the three institutions. So far as the language items are concerned, their results are directly opposed to those of Binet. It should be remembered, however, that the study of Weintrob and Weintrob is one of the few in the literature which does not show the inferiority of lower-class children. The data for categories of item merely confirm the general findings of the main inquiry.

Yerkes and Anderson (1915), reporting standardization data for the Yerkes Revision of the Binet Scales, included an analysis of responses to 20 individual test items by children from "favoured" and "unfavoured" schools. On 14 items both boys and girls in the "favoured" school were superior in performance ; on one item boys and girls in the "unfavoured" school were superior ; and on 5 items sex differences rather than social differences seemed to emerge.

In/

/In view of the small numbers of subjects tested, differences in performance between the two groups of children are probably not significant. The one item where the "unfavoured" children did appear to have an advantage was regarded by Yerkes and Anderson as testing "aesthetic judgment" ; subjects were asked to pass judgment on the prettiness of a picture of a face. However, this item was the first to be administered to both groups ; an explanation could therefore be put forward in terms of the greater timidity which the investigators gained the impression was characteristic of the "favoured" children.

A somewhat more detailed inquiry is reported by Bridges and Coler (1917). They administered the Yerkes-Bridges test to 301 children in the first three grades of two schools in Columbus, Ohio. One was a "favoured" school in a neighbourhood of high social status ; the other was an "unfavoured" school in a factory district and near a railroad, mainly attended by the children of labourers.

Bridges and Coler presented data to show that age norms for all the Columbus children taken together were superior to those for the Massachusetts children on whom the test had been standardized between the ages of 6 and 9 years, but that they fell below the Massachusetts norms for the later ages. Although the experimenters do not seem to have recognized it, this difference in results was obviously/

/obviously due to the fact that the older children in their sample were almost bound to be backward; Yerkes-Bridges norms had been established with a group of children ranging over the first eight grades.

They estimated that the children in the "favoured" school were from 21 to 32 per cent superior in performance to children in the "unfavoured" school, the actual amount varying with chronological age. In the absence of any satisfactory zero point for the test scale, this estimate in terms of percentages tells us very little about the real size of the difference between the two schools. Expressing data by Mental Age, they found a difference of two years in favour of the upper-and middle-class children at Grade I, and a similar difference of one and a half years at Grade III.

When the 20 separate sub-tests of the Scale were examined, differences in perfomance were always in favour of the "favoured" school, but they were sometimes very small. Differences were expressed as differences between mean scores; no tests of significance were applied. The sub-tests on which the "favoured" children showed the greatest superiority were absurd statements, comprehension of questions, comparison of familiar objects, concrete definitions, and counting backwards from 20 to 1. Those on which they showed least superiority were arranging weights, aesthetic judgment, copying/

/copying a square or a diamond, comparison of lines of different lengths, and drawing designs from memory.

Bridges and Coler offered special explanations for differences in performance on three sub-tests. They asserted that the superiority of the children from the "favoured" school in counting backwards from 20 to 1 reflected their greater familiarity with games in which counting backwards played a part. They suggested that differences between the two groups on the sub-test involving abstract definitions would have been much more marked had the "unfavoured" children not found "charity" so easy to define. And they pointed out that the sub-test requiring "aesthetic judgment" was so simple that it was seldom answered wrongly by children from either school.

They summed up their results by saying that the sub-tests showing the greatest differences between social groups invariably involved "analysis" and "abstraction", while those showing the smallest differences involved "motor co-ordination" and "kinaesthetic judgment". They claimed that this was in accordance with Thorndike's view that individuals differed least in sensory-motor functions and most in analysis and abstraction. Since in fact what they had shown was the lack of large <u>group</u> differences in sensory-motor tests, the reference to Thorndike/

/Thorndike is irrelevant. Variations in individual score on those tests may have been small, but they did not investigate this point.

One of the few item-analysis studies of this problem to be carried out on children in Great Britain is that of English (1917). His subjects were two groups of children aged from 12 to 14 years attending two English schools. The first group of 37 children belonged to the professional and upper-middle class; the second group of 31 children belonged to the lower-middle class, and included many sons and daughters of tradesmen, small shopkeepers, etc. He administered to both groups ten tests specially devised by himself. Many of these tests measured sensory-motor or kinaesthetic functions, but several items were included which he believed to measure "intellectual" functions such as ability to distribute attention, ability to sustain effort, analogical reasoning, and ability to apprehend spatial relationships.

On each test, mean scores for the two school groups were compared. In addition, pupils in each group were divided into sextiles on the basis of their performance on each test, and inter-school comparisons were made between equivalent sextiles. The statistical treatment of data is comprehensive, and in fact more detailed than the small numbers of subjects justify. Many simple, partial/

/partial, and multiple rank-order correlations were computed.

English came to three major conclusions. The first was that all children from the higher social class were strikingly superior on all tests except those involving rapid movement. The second was that on a number of tests differences between the groups were much smaller when top sextiles were compared than when any other comparison was made. The third was that the degree of correlation of test score with estimated intelligence sometimes varied widely for the two groups; he thus concluded "that a test may exercise one function in one group, another in another group, is clear enough when the two groups are quite different in respect to training and native ability". (English, 1917, p.299) In attempting to explain differences between groups in performance on individual tests, he had recourse to subjective impressions gained while the testing was being conducted. He stated that they could mainly be explained as reflecting basic differences in motivation, speed of working, and test attitude. Certainly this may be accepted as a <u>general</u> explanation for the existence of differences between groups, but variations between items, or tests, in the amount of difference shown cannot be accounted for in this manner unless it is assumed that motivation/

/motivation, speed, and attitude vary from task to task.

This was the first item-analysis study to employ correlation procedures. They were hardly appropriate in an investigation where so few subjects were available. But that does not detract from the fact that for the first time in this field something more elaborate than simple percentages and arithmetical means was used to bring meaning to results.

The first study reviewed in this chapter to measure intelligence by means of a group test is that of Pressey (1920). Working in collaboration with Ralston, she had found earlier a definite relationship between parental occupation and child IQ among children aged from 10 to 14 years. (Pressey and Ralston,1919). She thought that the real reason for this relationship might be differences in cultural home opportunities. She therefore undertook in 1920 a further study. In order to lessen the opportunity for home influences to operate she chose this time a group of younger subjects, aged from 6 to 8 years, dividing them into categories by occupation of father. She administered to these children the Pressey Primer Test, which she considered was less likely to be influenced by cultural environment than the Pressey Group Point Scale which had been used in the earlier investigation. This test included four/

/four sub-tests. Two of these she believed to be subject to some influence by home environment. One required a child to select the one of three pictures of objects which was not like the other two ; the other involved the crossing out of errors in a series of pictures. The remaining two subjects were included as being less influenced by home environment. The first involved the recognition of patterns of dots, the second the fitting together of pieces of geometrical figures.

Pressey found that substantially the same differences existed between occupational groups for all four sub-tests, being just as great for the supposedly non-cultural tests as for the others. She reasoned that these differences must therefore be "true differences in a fundamental, mainly general intelligence or native endowment" (Pressey, 1920, p.95). Here she seems guilty of thinking in a circle, since the differentiation between "cultural" and "non-cultural" sub-tests had not been made originally on the basis of experiment, but by the use of subjective judgment. All that can be taken out of the results is that her subjective judgments may have been wrong.

Nevertheless her study is a valuable contribution to the/

/the field, not least because she was the first experimenter here to measure social level by a scale applied to individual subjects, rather than by treating whole school populations as homogeneous social groups.

Burt reports a study carried out on school-children in London (1922, pp 190-99). Actually two studies are reported, carried out with different subjects, the one concerned with comparing the total scores of two different social groups on the Binet-Simon test (modified for English usage), the other concerned with an analysis of their performance on individual items. The first is not of importance to the discussion, since the difference found between the two groups was of the same order as in many studies reviewed in Chapter VII. The groups consisted of children aged from 7 to 14 years attending two schools in a London borough, the first an "occupationally good" school with parents employed as post office sorters, railway inspectors, etc., the second an "occupationally poor" school with parents employed in such occupations as hawker, roadsweeper and railway porter.

In the second study children from two pairs of schools were chosen as subjects. Burt gives no information/

/information about the social classification of the schools, the number of children involved, or their age and grade levels. He does say that the schools were picked to represent the extremes of the social scale, probably by the same method as in the first study.

Burt's experimental procedure was as follows. He administered the sub-tests in the Binet-Simon series to both groups; having calculated the percentage of pupils in each group passing each sub-test, he ranked the 65 sub-tests in order of their difficulty for each group. He took the difference between the two rankings as a measure of the social difference in each sub-test. Unfortunately he reports his findings in slight detail, giving only differences between rankings. Had he recorded the rankings themselves and the percentages by which they were calculated, it would have been possible to determine whether subjects in the lower social group excelled on any sub-tests.

Burt concluded that certain types of sub-test were easier/

/easier for children in the upper social group. These were : (I) tests requiring linguistic facility, particularly those dependent on a wide vocabulary; (2) scholastic tests, especially tests in literary subjects; (3) memory tests, such as repetition of sentences; and (4) tests depending upon items of information imparted during early life in a cultured home: e.g. surname, age, number of fingers, names of weekdays. For poorer children the easier tests were : (I) tests dependent upon familiarity with money; (2) tests perceptual rather than conceptual in character, especially manual drawing tests; (3) tests of a practical nature; and (4) tests depending upon practical shrewdness, such as the noting of absurdities.

Burt's analysis is an extremely interesting one. But it is evident he did not fully appreciate the difficulties inherent in his experimental procedure. At several points he comes dangerously near to asserting that tests which were relatively easier for the lower social group were absolutely easier as well. Of course the fact that a test varies in position in an order of difficulty from one group to another does not tell us which group finds it easier or more difficult. The further criticism may be made/

/be made that there seems to be considerable agreement between his two rankings. Although most sub-tests show some shift in rank between the two orderings, only 3 shift more than 5 points. There is no guarantee that changes in rank do not fall within the chance expectation for random errors.

Stoke (1927) carried out an investigation designed either to confirm or modify Burt's findings. Although he followed Burt's techniques, his approach is superior in several important respects. Children in the first three school grades were classified into two social groups, not by a general estimate of type of school attended, but by individual placing on the Taussig scale, those whose fathers fell into the two top categories being contrasted with those whose parents belonged to the two bottom categories. The groups contained 70 and 75 children respectively. Intelligence was measured by the Stanford-Binet test.

Stoke ranked items by order of difficulty for each group, and immediately discovered that the rank of a few items varied considerably. Out of 46 items, 12 showed a shift/

/ shift of more than 10 points, and 5 of 5 or more points. This is a much greater difference in ranking than that found by Burt. This may be attributable in part to the fact that Stoke's groups were socially further apart. Although Burt talked of his subjects as belonging to extreme social groups, there were no representatives in his upper group of the professional or higher business classes.

Stoke found that the items showing the largest variation in ranking were either very easy or very difficult. To avoid problems of interpretation he eliminated the 21 easiest and most difficult items and re-computed rankings for the 25 items within the middle range of difficulty for both groups. The highest shift of rankings was only 2.5 points. On the basis of this finding he claimed that Burt's methods were statistically unsound.

Unlike Burt, Stoke reported the percentage of each group passing each item. On 1 item the lower group excelled, though only by a margin of 4 per cent. On another item the two groups tied in performance. On the remaining/

/remaining 83 items the percentage of children in the upper group who passed exceeded the percentage of children in the lower group by an average of 10 per cent, the largest difference on any item being 38 per cent. Stoke offers the following explanations of certain of the differences. Of 8 items showing differences of 15 per cent or more in favour of the upper group, he says, "These are quite linguistic. They do not however demand answers of the rote meaning type, but require some original thinking" (Stoke 1927, p.30). Of other items he says, "There are other linguistic tests which seem about as easy for the low group as for the high................ Evidently the children of the low group have almost as much linguistic ability as have the children of the high, where rote memory is concerned; but where more than rote memory is required in the use of the words, the children of the high group excel" (Ibid, pp. 30-31).

Although the way in which percentages are employed is open to question, this is a study of great merit, which shows an unusual awareness of the dangers inherent in the use/

/use of inappropriate statistical techniques. It throws considerable doubt on Burt's findings, although these are still quoted in textbook accounts of research on this problem.

The next relevant study in historical order is that of Long (1935) who administered a battery of intelligence and achievement tests to third-grade negro children in Washington, D.C. Subjects were divided into two social groups. The first group included practically all children in certain schools in relatively under-privileged neighbourhoods. The second group consisted of a sample of children from certain schools in more favoured neighbourhoods; the sample was taken on the basis of assessments by principals and class teachers of the quality of children's homes. Social case histories for children in both groups were gathered by trained social workers and used to provide detailed descriptions of socio-economic background. For example, 19 per cent of the fathers of the lower group were illiterate, 72 per cent were unskilled labourers, and only 1 per cent were engaged in professional occupation. Of the fathers of the/

/the upper group 1 per cent were illiterate, 57 per cent were unskilled labourers, and 11 per cent were professional men. Socio-economic differences were thus considerable, but not so great as would be expected with white subjects.

Long analysed sub-tests in the Kuhlmann-Binet test for both groups. For each sub-test he reports mean mental ages, the difference between the two means, and the critical ratio of the difference.

He found that on four sub-tests (completion of sentences, likenesses in pictures, counting taps, and finding words not matching with others) the pupils in the upper social group were significantly superior. On two sub-tests (finding similar forms among a group of varied forms and substituting numbers for letters) the pupils in the lower social group were significantly superior. On the remaining four sub-tests there was no evidence of significant difference, although on three the lower group showed slight superiority.

A defect in this study is that no indication is given/

/given of differences in performance on the individual items comprising the sub-tests, and there is little attempt at interpretation of results. Nevertheless Long's investigation is important for three reasons. It is the first study to analyse in detail for negro pupils the relationship between socio-economic level and test score; it includes a very careful socio-economic description of the background of subjects; and it gives a relatively complete presentation of statistical data.

An investigation the value of which is seriously affected by inadequate statistical analysis is that of Saltzman (1940). She took as her subjects 354 first-grade pupils in two schools in New York. The schools were chosen to represent two social and economic extremes. Each child was given a Stanford-Binet test and the Goodenough Draw-A-Man test. On the Stanford-Binet pupils in the upper-class school obtained a mean IQ of 115, compared with that of 103 for pupils in the lower-class school. Mean IQs for the Goodenough test were 108 and 105 respectively.

Saltzman divided the pupils in each school into "inferior"/

/"inferior", "average" and "superior" groups by Goodenough IQ. Then she carried out an analysis of Stanford-Binet items from Age 5 to Age 10, not only for the total populations of the two schools, but also for each of the six Goodenough groups seperately. Her method of analysis was simple. Both for the social groups and the Goodenough groups into which they were divided she computed the percentages of subjects passing items, and used the difference between percentages as a measure of difference on individual items. Unfortunately she reports the standard errors of percentage differences only for the comparison of social groups.

She states that large differences seemed to occur with items of middling difficulty. Although she failed to recognise the fact, this is almost certainly a statistical artefact produced by the use of a percentage scale. A test of significance must always be applied to the difference between two percentages before conclusions can be derived as to its meaning.

Her interpretation of apparent social differences in performance/

/performance is equally unsound. It is based on the division of items into three categories, by the size of percentage differences between equated Goodenough groups. 10 items appeared to show differences in favour of the lower social group, 15 to show no difference, and 20 to show differences in favour of the upper social group. Of the first category of items she says, "It will be noted that the tests in which Group A (the low-status pupils) showed superiority involved the types of ability which children of a poor environment probably have more opportunities to acquire : counting and handling money, rote memory, and sensory discrimination". (Saltzman 1940, p.78). It is not at all clear that this explanation fits all the ten items. And since no estimates of significance are reported, and groups are small, it may well be that most of the differences fall within the bounds of chance variation.

In discussing the fifteen items which show no difference she does not point out that thirteen are either exceptionally easy or exceptionally difficult for both social groups; in nine every child marks a correct answer. There/

/There are only two items where lack of difference may have real meaning, and they are not discussed individually.

Of the twenty items showing differences in favour of the upper social group she says, "Their greatest superiority (Group B) was on the tests involving vocabulary, verbal comprehension of everyday situations, and rhymes. They are also superior in motor control. In the environment of these children we can readily see that there would be opportunities to converse with adults, develop motor control, and acquire information and accepted moral concepts" (Saltzman 1940, p.79). The same kind of criticism applies to this judgment as has already been made of her explanation for the items favouring the lower group. Her statements are plausible, and may well contain some truth, but they are not justified by her data.

The most important item-analysis studies of the relationship between intelligence and social environment so far made are those of Stone (1946), Murray (1947) and Eells (1948, 1951). Their researches form part of a large/

/large University of Chicago project designed to investigate "cultural learning". The project has been strongly influenced by the ideas on social class of Lloyd Warner (for some description of these ideas see pp. 25-27).

Stone's research is not directly concerned with intelligence test items as such. What he attempted to do was to measure differences in the knowledge by various social-class groups of the words used in standard group intelligence tests. His purpose was to determine the extent to which social class differences in test score might be due to basic differences with respect to familiarity with the vocabulary in which intelligence test items were expressed.

To that end he constructed a vocabulary test of 90 multiple-choice items, more than 90 per cent of which consisted of stimulus words in common use in group intelligence tests. He divided the test into two parts. The first part contained 50 items, one half based on stimulus words that were "picturable", the other half based/

/based on stimulus words that were "non-picturable". Distractors for each item were related to the stimulus word according to a definite predetermined plan. The second part contained 80 items, each offering to subjects a choice of two correct answers and four distractors. Subjects were required to circle what they considered to be the "very best" answer, and to underline the "next to best" answer. In this way Stone was able to identify the knowledge of some meaning for a word, the range of knowledge covering two possible meanings, and patterns of preference for one or other of the two correct meanings.

He classified his subjects into social groups by a rather more complicated procedure than has been normal in these studies. Warner's Index of Status Characteristics (see p. 87) was used to define social class levels in Rockford, Illinois; nine and ten-year-olds and thirteen and fourteen-year-olds in the schools were categorised by Index score as of high status, of medium status, and of low status. However, the children were also classified ethnically/

/ethnically according to whether they were native white American, of Scandinavian origin, or of foreign origin other than Scandinavian. A nine-fold division of pupils was thus produced. Only in three "cells", however, were numbers large enough to warrant experimental study. The groups in these cells are called in the investigation Middle Old American, Lower Old American, and Lower Italian Ethnic.

Most of the data which Stone reported concerns only the two American groups. Working in terms of percentages, he found that on 48 to 88 per cent of the items in his vocabulary test, the exact proportion varying with age level, there were differences in favour of the medium-status (Middle Old American) group significant at least at the .05 level. There were ten items showing differences in favour of the low-status group, but only two of these were significant. On the second part of the test, there were about 7 per cent of the items, at both age levels, showing differences small enough not to be significant even at the .5 level of significance.

When the words known best by each status group were listed, a considerable agreement was evident. The/

/ The degree of correspondence between the two lists, when the best known quarter of the words for the middle-status group was compared with the best known quarter for the low-status group, varied from 60 per cent to 77 per cent. The use of percentages is clumsy here; rank-order correlations would have been more easily interpreted.

When preferences between the two correct answers in the second half of the test were analysed at both age levels, Stone found that for 77 per cent of the items there were marked changes in preference by one or both of the two status groups as children grow older. For 47 per cent of the items these changes were in the same direction for both groups. For 30 per cent of the items the middle-status group lost preference for a specific meaning, while the lower-status group increased their concentration of preference. Stone's interpretation of this finding is not altogether satisfactory . He asserts that the data showed that early word preference depends on word knowledge and reading ability. This is probably correct, but does not explain differences between the status groups at the older age level——— unless it is assumed that the older children in the low-status group have less word knowledge than the younger/

/younger children.

In the first part of the test both groups found the "picturable" items easier than the "non-picturable" items. But the superiority of the middle-status group was still marked on both types of item. Among thirteen and fourteen-year-olds the middle-status group exceeded the lower-status group by an average of 4.3 per cent on "picturable" items, and by an average of 24 per cent on "non-picturable" items. For the nine and ten-year olds this relationship was reversed , the middle-status group exceeding the lower-status group by an average of 30.6 per cent on "picturable" items , and by an average of 19.9 per cent on "non-picturable" items.

Stone explains the differing results for the two age levels in the following manner. He hypothesises that the Middle Old Americans have become acquainted with a wider range of words representing "things" at the younger age than the Lower Old Americans. But, he says, as the Lower Old Americans grow older they begin to approach the Middle Old Americans in this kind of knowledge; meanwhile the Middle Old Americans have gone ahead in their knowledge of words descriptive of abstract concepts.

Stone also analysed in some detail wrong answers by b oth status groups. No social differences were in evidence here. All children tended to be distracted most/

/most frequently by wrong answers which were similar in appearance to stimulus words or related to them in a more general or abstract way than the correct answers, with the former tendency predominating for the younger pupils and the latter tendency for the older pupils.

This is a study which has extreme importance for the constructors of group intelligence tests, and deserves more recognition than it has so far obtained. Admittedly Stone's analysis of data is not always statistically sound; the averaging of percentages over a wide range of items is a dubious technique. Nevertheless his errors of method do not, in the opinion of the writer, affect the general validity of his conclusions. He has demonstrated clearly that the vocabulary in which many group intelligence tests are couched is subject to significant social class differentials. It follows that social class differences in performance on test items need not necessarily be attributed to differences in ability to solve the problems posed by the items.

The second important investigation in the Chicago project is that of Murray(1947). He administered a series of group intelligence tests to ten and fourteen-year-old negro children in Gary, Indiana. The Otis Alpha Verbal, the Henmon-Nelson and the Kuhlmann-Anderson were administered/

/administered to the ten-year-olds; the fourteen-year -olds sat the Henmon-Nelson, the Otis Beta, and the Chicago Primary Mental Abilities Test. Children were classified into three social-class groups on the basis of data obtained from interviews with parents. The groups are defined in the investigation as middle-class, upper-lower and lower-lower(the last two being working class groups).

Murray found that on every test except the Otis Alpha there were significant differences in mean IQ between the social-class groups, the advantage in any comparison always being in favour of the upper groups. The non--significance of differences on the Otis Alpha may have been accidental; only pupils in the first four school grades sat the test, and thus 62 ten-year-olds in higher grades were not tested.

For his item-analysis Murray omitted the upper-lower class, comparing directly performance by the middle-class and lower-lower-class children. He included in the analysis only items attempted by substantially all subjects in both groups, working with a total of 119 items from three tests. He does not state whether he dealt with the two age groups separately (one of the tests from which items were selected for analysis was the Henmon-Nelson, which/

/which had been administered to both ten-year-olds and fourteen-year-olds).

For each of his 119 items he calculated the percen--tage passing in both social class groups, the difference between the two percentages, and the critical ratio of the difference. Only 5 items showed differences in favour of the lower group; in one case the difference was significant at the .05 level. 16 per cent of all items in the Kuhlmann-Anderson, and 64 per cent of all items in the Henmon-Nelson, showed differences in favour of the upper group significant at the .01 level.

Then Murray divided up the items into categories e.g. number series, word series, analogies, artificial language, geometric design, opposites, syllogisms, arithmetic and social reasoning. He determined the proportion of items in each category which showed significant differences between the two groups. His findings were not very conclusive, but there did seem to be a tendency for the geometric-design items to discriminate less than items in the more highly verbal categories. He suggested that these items mainly demanded abilities of perception and form, wheras verbal items demanded abilities dependent on school and extra-scholastic training and were thus more likely to be influenced by social environment.

Murray then proceeded to make an analysis of wrong responses/

/responses for 82 of the items. He computed the percentage of pupils in each group marking each wrong response on each item. 82 items showed significant differences between groups. He concluded that all these differences could be accounted for in terms of (a) differences in social experience, (b) differences in school achievement, (c) differences in mental associational patterns, and (d) structural peculiarities of the items.

Although results are not as conclusive as might be expected, this is a study of considerable merit, distinguished by its systematic approach. It is the first to attempt to classify items in a consistent and exhaustive manner in order to derive generalisations as to social class differences. Of particular interest and novelty is the analysis of wrong responses, although the criticism may be made that this seems to have been carried out in such a way as to identify class differences in items as a whole, rather than so as to isolate differences in response to specific choices.

The investigation of Eells (1948, 1951) is the most thorough and detailed of all item-analysis studies of the problem. It contains a wealth of statistical material, some of which however, is only indirectly relevant/

/relevant, and tends to crowd out of his thesis reviews of previous work in the field.

The children selected for study were the same as in Stone's investigation (See p.158), belonging to three social class groups and two age levels. Eells defined the groups in a slightly different way from Stone, calling them high-status Old Americans, low-status Old Americans, and low-status ethnics. Most of the analysis is in terms of the differences between the high-status Old Americans and the low-status Old Americans. There were 467 children at the two age levels in the high-status group, compared with 706 in the low-status group. The Henmon-Nelson, the Otis Alpha Verbal, the Kuhlmann-Anderson, and the Otis Alpha Nonverbal were administered to all nine- and ten-year-olds; the Terman-McNemar, the Otis Beta, the California Mental Maturity, the Thurstone Reasoning, and the Thurstone Spatial were administered to the thirteen- and fourteen-year-olds. Substantial differences were found in average IQ for the two status groups, always in favour of the high-status children. The IQ difference varied from 8 IQ points in the Otis Alpha Nonverbal test to 23 IQ points in the Terman-McNemar Test. Differences for the tests scored in percentile ranks varied from 17 to 33 percentile-rank points. All differences were significant/

/significant, the smallest being seven times its standard error.

The nine tests provided nearly a thousand items for analysis, one-third of which were excluded because they had not been reached by a substantial proportion of subjects in either or both status groups. Very difficult and very easy items were also eliminated. The basic statistical measure used in determining the status differentials from the remaining items was the difference between two percentages —— the percentage of one status group answering an item correctly, and the percentage of the other status group doing so. For certain purposes these percentages were converted into a normalized index, in order to avoid the basic defects of a percentage scale.

Eells' most important findings may be summarized as follows; About half of the items in the test for nine and ten-year-olds and about 85 per cent of the items in the tests for the thirteen and fourteen-year-olds showed differences in favour of the high-status group large enough to be significant at the .01 level. On the other hand, more than a third of the items from the tests for the younger pupils, and about 10 per cent of those from the tests for the older pupils, showed status differences too small to be significant even at the .05 level. 27

of/

/of these items showing insignificant differences were at neither extreme of difficulty for either group.

Mean status differences were largest for verbal items and smallest for picture, geometric-design and stylized-drawing items. Practically all the items showing large status differences were verbal in symbolism, while items showing small status differences were usually non-verbal or involved simple everyday words in common use.

When the distribution of wrong answers was compared for the two status groups, 75 items were found on which differences in the pattern of error were significant at the .01 level. Eells felt that on a number of these items the differences could be accounted for in terms of status differences in opportunity for familiarity with certain objects, words or processes.

When performance at the different age levels was examined, it was found that the proportion of items showing significant differences was larger in the tests for the thirteen- and fourteen-year-old pupils than it was in the tests for nine- and ten-year-old pupils. This age differential appeared to be due in part to the presence of a larger proportion of verbal items in the tests for the older pupils.

When/

/When the relation between status difference and item difficulty was investigated, it was found that the size of status difference on any item tended to vary markedly with the difficulty of the item, being largest for the easiest items and smallest for the hardest items. In the case of non-verbal items, however, the relationship varied according to the age level. For the older pupils, differences were still most marked for the easier items; but for the nine- and ten-year-old pupils the easier items showed the smaller status differences.

In general Eells explains his results in terms of an environmental hypothesis. He claims that the status differences shown by items can be shown to be related to differences in social opportunity, nature and extent of vocabulary, and life experience between the two status groups. He admits, however, that differences on some items cannot be interpreted in this fashion, but suggests that his own lack of insight may be responsible (Eells et al., 1951, p.357). The major defect of his work is an unwillingness to take cognizance of the work of professional colleagues who assert that differences in test performance between social groups reflect real differences in innate ability. He does sometimes refer to/

/to "genetic" hypotheses, but only cursorily, and without specific relation to his experimental findings. Nevertheless, his investigation deserves praise as the most articulate and statistically adequate so far carried out on the problem of social differences in test performance.

Macdonald's study has already been noticed. (Chapter VII, p. 92). It was carried out in Glasgow on three groups of children belonging to different socio-economic levels. The basis of social categorization was the quality of housing accomodation available to children in various Glasgow schools. The measure of intelligence was Glasgow Group Mental Test No. 5, which pupils sat as a routine part of their 'calibration' examination for secondary education.

In the item-analysis section of this study the main findings were: (I) that differences in performance favouring the upper social groups were uniformly large on verbal items, but items involving the use of number also showed considerable differences; (2) that differences in performance favouring the upper social groups were uniformly large on types of problem such as the logical ordering of five given words and the identification/

/identification of opposites to given words and uniformly small on the identification of missing components and the completion of analogies; (3) that there was no consistent relationship between the difficulty level of items and the extent to which they discriminated between social groups. Results are not altogether in agreement with those of Eells, but this may be explained by reference to differences in the method of social selection and in the measure of the intelligence employed; it is also possible that the performance of American and Scottish children may not be directly comparable.

The thirteen studies reviewed in this chapter comprise, so far as the writer is aware, all the investigations of social differences in test performance in which the technique of item-analysis has been used. Clearly much remains to be done. Of only the latter two studies, those of Eells and Macdonald, can it be said that they suffer from no noticeable statistical defects. Macdonald's mode of social classification is open to criticism, however.

The deficiencies in the majority of the studies makes it difficult to derive satisfactory generalizations, a/

/a difficulty that is emphasized by their inconsistency of findings. The most that can be said is that items do vary markedly in the extent to which they show differences in performance between social groups; and that highly verbal and linguistic items seem to favour the upper social groups more than "practical" items. In itself this is insufficient to weaken seriously hereditarian interpretations of social differences in performance. Yet it is true that Stone's work on test vocabulary, if his findings were substantiated by further research, would provide strong arguments at least for a remodelling of the language in which test problems are expressed. Obviously there is much scope for further investigation. There is a definite possibility that a sustained attack with item-analysis procedures on this problem of social differences on test performance might produce a solution of the heredity-environment controversy which has agitated psychologists over so many years.

CHAPTER X.

THE SOCIAL AND EDUCATIONAL RELEVANCE OF RESEARCH ON THE RELATIONSHIP BETWEEN INTELLIGENCE TEST PERFORMANCE AND SOCIAL ENVIRONMENT.

The analysis and interpretation of social differences in test performance is not to be valued merely as an academic exercise. If , as Zangwill suggests, psychology has given birth to a technology in intelligence testing, then its practitioners, like their colleagues in other fields of practical science , will find that their activit--ies are of social relevance. Informed discussion of psychometrics may be confined to professional journals; but knowledge of the common facts of testing is now wide--spread, and these have crept more and more into political debate.

The immediate social relevance of intelligence tests is that they are part of the machinery of educational selection, which means that they help to determine which individuals in a generation of school children will be socially/

/socially mobile. No arguments that different types of secondary education should have parity of esteem can disguise the fact that the decisive event in the careers of the vast majority of children whose parents cannot afford to send them to independent schools is their selection or rejection for a grammar or senior secondary school at the age of 11. Under these circumstances it is understandable that parents apprehensive of the chances of their sons and daughters in the struggle for better jobs and social advancement should claim that intelligence tests are inaccurate measures of ability.

On the evidence of the last few chapters, it would be expected that the children of working-class parents, and especially the children of unskilled manual workers, would be under-represented in the populations of secondary schools offering academic-type curricula. Experimental findings have confirmed this expectation. Gray and Moshinsky (1938), in their study of a representative sample of London children attending private, preparatory, and public elementary schools in 1933-34, showed a striking positive relationship between the social origins of able children and their chances of obtaining entry to a grammar school or a "major" or "minor" public school. They found that while nearly all the children of professional and business parents/

/parents who possessed high ability(defined as an IQ of 130 or more on the Otis Advanced Group Intelligence Test) had the opportunity of secondary education, the corres--ponding proportion for the children of clerical and commercial employees was 50 per cent, for the children of skilled manual workers 30 per cent , and for the children of unskilled manual workers 20 per cent. For children of lesser ability this social disproportion was even more marked. Only 1 per cent of the children of unskilled workers with IQs less than 130 entered a secondary school, compared with 50 per cent of the children of business men and 35 per cent of the children of professional men. Gray and Moshinsky pointed out that the inequality of educational opportunity reflected by these figures represented a tremendous waste of natural ability . It may be said here that the machinery of educational selection put into operation after the passing of the 1944 Act was in large part designed to ensure equality of opportunity for all able children, no matter what their social origins. Nevertheless it should not be forgotten that the possession of ability is still not enough in itself to guarantee social advancement, even though it may lead to full secondary education. Some prizes are as yet beyond the reach of working-class children ; Floud(1950) assures us/

/us that many have been immobilised permanently in the lower middle classes.

Nearly a quarter of a century has passed since the Gray-Moshinsky investigation, and it seems clear that the gross waste of working-class ability which they deplored no longer characterizes the educational scene, to whatever extent wealth and privelege still operate to determine possession of a minority of important posts at the top of the social tree. As the result of a survey carried out in Middlesborough and South West Hertfordshire in 1952 and 1953, Floud, Halsey and Martin (1956) concluded that virtually the full quota of boys with the requisite minimum IQ from every social class admitted to grammar schools. There was an almost perfect relationship between "ability", as defined by intelligence test score, and "opportunity", as defined by freedom of access to grammar schools.

But the fact that children of different social classes are now competing on apparently equal terms for places in academic secondary schools does not mean that no inequalities remain. In the same survey Floud, Halsey and Martin calculated the "class-chances" of a grammar school education for boys in different occupational/

/occupational groups. Although the selection procedures were absolutely fair, working-class children not being discriminated against in any way, they found that the sons of manual workers had a chance below the average, and the sons of non-manual workers a chance above the average, of being selected for grammar school. The sons of clerks had at least four times as good a chance as the sons of unskilled manual workers, and two to three times the chance of sons of skilled workers. In Middlesborough the son of a professional or a business man had more than seven times the chance of the son of an unskilled worker, and nearly five times the chance of the son of a skilled worker, while in South West Hertfordshire he had six times the chance of the unskilled worker's and three times the chance of the skilled worker's son. Furthermore, it appeared clear that these differences in proportion of the contribution of different occupational classes could be explained almost entirely in terms of the unequal distribution of measured intelligence.

It would thus appear that equality of educational opportunity has been achieved, at least at the level of grammar school entrance. It has come to mean ,however, as Marshall (1950, p.66) says, "the equal right to be recognised as unequal".

Economic/

/Extreme importance must therefore be attached to fundamental research into the meaning of social differences in test performance. If they can be attributed entirely or in large part to environmental, motivational and attitudinal factors, then it must be obvious that the present methods of educational selection are less than just to very many working-class children. If they can be shown to be due to real innate differences in natural ability, then the selection methods are scient--ifically justified, working-class children are being fairly treated, and complaining parents are simply allowing their social ambitions to override their common sense. Of course what is scientifically justified is not always morally justified; some educationists consider that the use of education as an agent of social selection is normally indefensible, however equitable the techniques employed.

It must be said that there is a noticeable tendency among practising psychometrists to regard the matter as already settled; if working class children are inferior in test performance, then they are of lower intelligence. Their colleagues who have inclinations towards research are usually more guarded, knowing the difficulties which this interpretation raises.

In/

/In the last analysis the decision to employ or not to employ selection procedures, including intelligence tests, for entrance to particular schools or courses is in a democratic community in the hands of the people as a whole, including those sections who perform poorly on intelligence tests. The people as a whole are not always guided by rational considerations, although it is the chief tenet of democracy that in the main and in the long run their judgements are correct. But they have a right to demand that wherever possible the experts in particular fields should provide them with unambiguous information. It is the duty of psychologists by continued research to take the ambiguities out of intelligence testing.

PART II.

AN INVESTIGATION OF THE INTELLIGENCE TEST PERFORMANCE OF 1000 CHILDREN OF FIVE DIFFERENT SOCIAL GROUPS.

CHAPTER XI.

PROBLEMS AND DIFFICULTIES IN SELECTION OF SUBJECTS.

The theoretical difficulties and mass of routine operations involved in even a moderately comprehensive study of the relationships between intelligence test performance and socio-economic background, social environment, or social class are so considerable as to place it without the normal practical range of the solitary research worker. This partly explains why only in mid-twentieth century are we beginning to feel our way to some understanding of these relationships. The researcher may only undertake such a study as a member of a research team or by the chance that already existing material may be adapted to his purposes.

In the present case the data accumulated by the

1947/

/1947 Scottish Mental Survey Committee in order to examine, inter alia, gross differences in test score between occupational groups were also capable of being employed to examine the differential responses of different social groups to individual test items. The Survey Committee therefore consented to the use of its primary data in this way.

It was found, however, that the employment of this material presented a number of special problems which had to be solved before the study could be advanced. A discussion of these problems and the solutions event--ually found will illustrate some of the unusual diffic--ulties faced by the worker in the field of community studies.

As has already been stated, the Survey Committee were interested in the gross differences in test score between occupational classes or, to put it in a more abstract form, in the relationship between intelligence test score and socio-economic status.(Social Implications of the 1947 Scottish Mental Survey, 1953, p.3) The first difficulty arose with regard to the method of measurement of socio-economic status employed. This was as follows. So far as possible, the occupation of parent was obtained for all children tested. All the occupations/

/occupations represented were then classified in accordance with the method employed by the Royal Commission on Population in the Family Census of 1946. Briefly, all occupations were divided up into nine categories which were then defined as occupational groups, (Social Implications, 1953, p.58). These categories were:-

I) Professional and large employers.

2) Small employers.

3) Salaried employees.

4) Non-manual wage-earners.

5) Skilled manual wage-earners.

6) Semi-skilled manual wage-earners.

7) Unskilled manual wage-earners.

8) Farmers.

9) Agricultural workers.

In certain cases it was by no means easy to fit the given occupation of a child's parent to a category. This was especially true where the degree of skill of a manual worker could not readily be assessed. The Survey Committee decided that the most satisfactory method of dealing with the latter problem was to place children in category 6, the occupational class of semi-skilled manual wage-earners/

/earners. (Social Implications, 1953, p.44).

All comparisons between social groups made in the publications of the Survey Committee are based on this division of the community into nine occupat--ional classes.

The main interest in the present inquiry, however, is not in social groups considered as occupational classes, but in social groups considered as social classes. The nature of social class, and its various suggested modes of definition, have already been considered in some detail in Chapter IV.

It is unnecessary to labour the points made. Suffice it to say that the occupational similarities of members of an occupational class do not imply corresponding psychological similarities. It is unlikely that an occupational class considered purely as such will possess psychological unity as a group; its members are unlikely to share more than a few major interests associated with their work. A social class, on the other hand, according to many observers, does possess psychological unity, is a psychologically integrated group. (Linton,1947,pp.39-40) It may there--fore be felt to follow that occupation is not a very satisfactory index of social class . But recent findings seem to show a very close association, under/

/under normal circumstances , between objective indices of social class, of which occupation is a type, and subjective indices such as self-assessment of social class belonging. (Warner, Meeker and Eells,1949) Thus there is considerable justification for assuming that at least some of the occupational classes categorized by the Mental Survey Committee may be taken as practic--ally equivalent to social classes, and regarded as distinguishable and rather well-integrated psychological groups.

The decision that occupation could be taken as, if not a perfect, at least a good practical index of class, generated a further problem. How many social classes were to be included in the study ? In order to arrive at a proper answer to this question, it would be necessary to know how many social classes may readily be differentiated in our society. As has already been pointed out (Chapter IV,p.31) there is so far insuffic--ient evidence available to sociologists and social psychologists to enable them to determine this point with any accuracy. It follows that any proximate solution is to some extent tainted by the subjective viewpoint/

/viewpoint of the experimenter.

For the purposes of the present study, it was felt that five distinct social classes of some size could usefully be differentiated. These were (1) an upper-middle-class corresponding to occupational class 1, that of professional workers and large employers, in the Mental Survey data, (2) a middle-middle-class corresponding to occupational class 3, that of salaried employees, (3) a lower-middle-class corresponding to occupational class 4, that of non-manual wage-earners, (4) an upper-working-class corresponding to occupational class 5, that of skilled manual wage-earners, and (5) a lower-working-class corresponding to occupational classes 6 and 7, the semi-skilled and unskilled manual wage-earners. There may be some quarrel with the terminology adopted. The names given to the five social classes are certainly chosen fairly arbitrarily. But they are of importance mainly in so far as they convey the notion of a hierarchy. Another set of names equally descriptive of the hierarchical principle of social class would be equally acceptable. However, defining terms which contain the phrases "working class" and "middle class" have the merit of being easily assimilated to common usage.

It/

/It will be noted that the selection of five social classes for study has involved the exclusion from consideration of three of the occupational classes categorized by the Mental Survey, namely the farmers, the agricultural labourers, and the small employers. This step reflects to some extent the difficulties inherent in the choice of occupation as an index of social class. It is difficult to subsume farmers and agricultural labourers under the normal class headings. For a time the possibility was considered of assimilating farmers to the upper-middle-class and regarding agricultural labourers as belonging to the lower-working-class. But on mature consideration this procedure appeared so questionable that it was abandoned. In any case it seemed likely that the introduction of the two groups into the study would produce a rural-urban dichotomy which would seriously complicate the experimental investigation.

The category of small employers was left out for the following reasons. In an era of full employment, with a steady increase in both production and consumption, it is probable that there exists a strong trend of upward social mobility in this group. It was also considered that/

/that the method of differentiating small employers from large employers used by the Mental Survey (small employers were defined as employers who employed less than ten people, large employers as employers who employed ten or more people), while valuable as a principle of occupational classification, is of little merit as a means of social classification, since it obscures the very large differences in income and standard of living which may exist within the groups. This does not matter so much with large employers, of whom even the less prosperous are still comparatively wealthy compared with the rest of the community. But individuals with incomes barely, if at all, superior to those of skilled manual workers may be placed in the occupational category of small employers beside other individuals whose yearly incomes, despite the small number of workers they employ, are considerably larger than those of many large employers.

While three occupational classes were excluded, two, the semi-skilled and the unskilled manual wage-earners, were merged to represent the lower-working-class. The justification for this procedure is as follows. Firstly, it/

/it is often very difficult to decide whether a particular occupation should be classified as unskilled or semi-skilled, especially when the classification is carried out by middle-class people with little or no experience of manual labour. This difficulty finds voice in the instructions for coding of occupations issued by the Survey Committee, where it is set out that the occupational class of semi-skilled manual wage-earners should be regarded as a "residual class" into which all manual wage-earners not obviously falling in the other two classes of manual wage-earners should be placed (Social Implications, 1953, p. 72). Similar advice is proffered for no other coding problem. It obviously creates the danger that the individuals placed in this class will form only an amorphous group of "rejects".

G. D. H. Cole makes something the same point in discussing the conclusions regarding British class structure to be drawn from the British Census of 1951. The Census statisticians had divided the community into "social classes" by occupation in such a way that skilled manual wage-earners fell into Class III, semi-skilled/

/semi-skilled manual wage-earners into Class IV, and unskilled manual wage-earners into Class V. Of this classification, and of the decisions by which occupations were assigned to various Classes, Cole remarks, "...... the realistic information that can be derived from the Census classification is very limited as far as the distinctions between Class III and Class IV are concerned. Nor is the situation any better in respect of the distinctions between Class IV and Class V. These indeed hardly appear to rest on any principle at all"(Cole, 1955, p.184).

In the senond place, some legitimate doubt may be expressed as to whether manual wage-earnors themselves visualize a tripartite division of their jobs in terms of skill. There may be a logical superimposition of inapplicable categories on manual occupation by middle-class intellectuals. Perhaps the most knowledgeable of all psychologists who have worked in this field in Britain is Ferdynand Zweig. Admittedly many of his insights are not easily translatable into statistical language. But while he sees manual workers as being divided into a very large number of horizontal and vertical/

/vertical groupings, whenever he has occasion to talk in a general way about working class characteristics, he distinguishes between two basic groups, the craftsmen and the labourers. (Zweig, 1952, passim).

No further problems of this special type arose in regard to the adaptation of the Mental Survey material to the purposes of the investigation.

CHAPTER XII.

SELECTION OF SUBJECTS AND SECURING OF TEST MATERIAL.

The last chapter has dealt with the way in which the material provided by the Mental Survey Committee was approached, and some methodological difficulties were tackled. The procedure of selection of the basic experimental data from this material is described below. A short account of the material available will show that selection was not necessarily a simple matter.

The Survey Committee attempted to test a complete year group of Scottish children, those born in 1936. In addition schools were requested to complete seven--teen-item sociological schedules for all pupils on their rolls. Altogether 75, 451 were filled up. When incorrectly completed or damaged schedules had been eliminated/

/eliminated, detailed sociological information was available for 75,211 children. Of these children 4406 were absent on the day of the test and 532 had been judged physically and mentally too defective to attempt it. It was decided to include the latter group with the children who sat the test and obtained a zero score. Thus there remained 70,805 children for each of whom test score was known or assumed, and correpspondingly for each of whom a number of important sociological facts had been recorded.

For all children born on the first three days of each month of 1936, a group afterwards known as the Thirty-six-day Sample, a more comprehensive sociological schedule was also completed. There were 7380 of these children: test scores were known for 6857: test scores and sociological information were both available for 6627.

A further sample was also drawn, consisting of all children born on the first days of February, April, June, August, October and December. This group of 1215 children, described as the Six-Day-Sample, were given an individual Binet test (Terman-Merrill Revision, Form L), so that for each of them is recorded full sociological information/

/information, a group test score, and a Terman-Merrill IQ. In practice numbers dwindled to 1208. (Social Implications, 1953, pp 1-2 and passim).

There were thus three aggregates available for experimental use, the total group, the Thirty-Six-Day Sample, and the Six-Day Sample. The main purpose of the experiment was to compare the group test performances of children in the five social classes already defined, (Chapter XI, p. 181). Since the information on parental occupation had only been recorded for the two latter aggregates, there was no means of distinguishing social classes in the total group. Therefore for practical purposes only the material of the Thirty-Six-Day Sample and the Six-Day-Sample could be used in the study.

It was felt that it would be easier to compare the relative performances of the five social classes if the numbers of subjects in the classes were equal. It was also necessary to strike a balance with regard to the numbers selected. Too large numbers would result in a mass of unwieldy data; too small numbers would produce unrepresentative data from which inappropriate generalizations might be drawn. Eventually it was decided that each class should consist of 200 individuals/

/individuals. On that basis selection was made directly from the Thirty-Six-Day Sample : numbers in certain of the occupational classes in the Six-Day Sample were less than 200. Of course by reason of the way in which the sampling was originally made the larger sample included the smaller.

The following table shows the distribution of the Thirty-Six-Day Sample by occupational class where both occupational class and test score are known.

TABLE III.

DISTRIBUTION OF THIRTY-SIX-DAY SAMPLE BY OCCUPATIONAL CLASS.
(Reproduced from Social Implications, 1953, Table XIX, p. 44).

Occupational Class	n.
1	221
2	330
3	236
4	556
5	2392
6	1190
7	1132
8	142
9	428

An/

/An examination of this table shows, as would be expected, that numbers vary considerably from class to class. It follows that the 200 subjects drawn from occupational class 1 to form the upper-middle class for the purpose of this study stand a somewhat better chance of being a representative sample than the 200 subjects drawn from occupational class 5 to form the upper-working class. In the first selection 90.5% of all subjects will be included in the sample; but only 8.4% will be drawn in the second sampling. Nevertheless if sampling is made on a properly random basis any loss in accuracy in the latter case should be insignificant.

The detailed procedure of selection was as follows: All relevant sociological schedules for the Thirty-Six-Day Sample were arranged in five piles corresponding to the five social classes. Arrangement within each pile was done alphabetically, firstly by geographical area, secondly by name of school within area, thirdly by name of child within school. The piles were then counted and divided in groups of ten. Schedules which showed the children for which they were completed as suffering from physical or mental defect too serious for them to be tested were excluded.

Reference/

Reference was then made to Tippett's Tables of Random Numbers. Numbers were taken for each pile until 200 had been written down. These numbers were regarded as referring to the order of schedules in a pile, and the appropriate schedules were selected. The worked test papers for the 200 schedules in each class were then obtained from among the 70,805 papers which represented the worked test material for the entire 1936 year group. Seven papers had disappeared and could not be traced. To replace them other random numbers were drawn until the count was finally complete.

The basic research material, consisting of 1000 worked group test papers divided into five social class categories, had now been accumulated. The 1000 sociological schedules were also retained, to be consulted in case of need.

CHAPTER XIII.

DESCRIPTION OF TEST.

Before proceeding to deal with other matters it is probably best to consider in some detail the group test which had been sat by the 1000 subjects concerned in this investigation. In the course of planning the 1947 Survey the Mental Survey Committee decided to use the same group test as had been employed in the Survey of 1932, in order that the two sets of results might be directly comparable. (The Trend of Scottish Intelligence, 1949, p.10). The test was a typical Moray House test, one of the many produced in the last twenty-five years and used by education authorities all over the country. It was felt that although test construction had improved since 1932, the improvement was not so great as to render a test of that period in any way obsolete.

This/

/This test is divided into two parts; it comprises both a verbal test of the normal type, and a pictorial or non-verbal test. As in 1932, the whole test was attempted by children, but attention has been directed in the Mental Survey investigations to their performance on the verbal part, as also in the present inquiry.

The verbal test is made up of 71 numbered items. Two items are divided into three parts, and one item into two parts; for these items three and two marks of total score are awarded respectively. Thus the possible score on the test is 76, 68 items each counting as I, I item as 2, and 2 items as 3. It will be noticed in succeeding chapters that throughout the inquiry all sub-items for which test instructions demand the award of separate marks have been treated as whole items.

Items included in the test are of many different types, as categorization in later chapters will show. One important fact which stands/

/stands out is the high proportion of multiple--choice items. 40 items are multiple-choice, i.e. 52.6 percent of all items. 30 of the 40 offer a choice of five answers to the subjects, 6 a choice of four, 3 a choice of three, and I a choice of two. The presence of so many multiple-choice items must obviously give some scope for scores on the test by individual children which do not represent their true scores. A subject may be postulated, for instance, who is ignorant of the correct responses to any of the five-choice items. Marking any one purely at random, he has one chance in five of marking the correct response. Marking all 30 at random he will on the average mark 6 correct responses. Similarly he will get I or 2 right of the four--choice items, and I right of the three-choice items, while he has a fifty-fifty chance of marking the correct response to the two-choice item. Furthermore, if we assume a population of subjects marking the items in this random manner, responses/

/responses will distribute themselves in such a way that certain individuals will mark correct responses considerably in excess of the chance total for any one individual.

This criticism is of course put forward to illustrate a slight deficiency in the test, not to invalidate it. Test compilers often find that multiple-choice items are good test items in the sense that they distinguish well between high and low scorers. They have thus justified the use of these items empirically. Also, few subjects attack test items in a random manner; they produce responses on the basis of either information or misinformation.

Nevertheless this defect of multiple-choice items is recognized, and statistical methods have been worked out of correcting assumed mutations from true score. They are not frequently used in this country, but sometimes in the United States. Their use also involves certain assumptions which are not universally accepted by psychometrists (Eells et al,/

/(Eells et al.,1951, p.167).

Another point to which an extremely meticulous critic of this test might draw attention concerns the unequal distribution of correct responses among the various positions of the multiple-choice items. Table IV shows this unequal distribution very clearly.

TABLE IV.

DISTRIBUTION OF CORRECT RESPONSES FOR MULTIPLE-CHOICE ITEMS BY POSITIONAL ORDER.

Position of Correct Response.	Five-Choice Items.	Four-Choice Items.	Three-Choice Items.	Two-Choice Items.
1	6	3	-	1
2	8	2	1	-
3	10	1	2	
4	3	-		
5	3			
Totals	30	6	3	1

The conclusion to be drawn from this table is that the/

/the earlier positions in the order tend to be weighted for correct responses. The danger is thereby renewed that an ignorant subject may secure a score greatly superior to his true score, in this case by marking a particular position in the order throughout the test. It has already been suggested that a subject who marks in a completely random manner will be most likely to score six correct responses on the five-choice items. But if instead he always marks position 3 he will score ten correct responses.

The test constructor would probably not regard this criticism as important. He might point out that in fact the later positions tend to be given less consideration by subjects, and that a concentration of correct responses in the earlier positions is necessary to correct this tendency. Again, he might argue that since he treated each item on its individual merits, he must be given liberty to arrange multiple-choice items individually to produce the best results.

Thirdly/

/Thirdly, he might uphold the principle of random selection for each item of the position of the correct response, for instance by throwing a dice.

Thus, while this test may be fairly criticised along certain lines, a good case can be made for its present form. It is typical of the kind of group test which has been in accepted use for many years, and there need be no lack of confidence in its findings, although of course they may be subject to differing interpretations.

CHAPTER XIV.

SOCIAL CLASS DIFFERENCES IN GENERAL TEST PERFORMANCE.

Once it had been established that the test possessed no special defects it was possible to proceed with the analysis of the test scores of subjects. This was necessary for three reasons, firstly because of the intrinsic value of the information to be gained, secondly in order to compare samples with the "universes" from which they had been drawn, thirdly in order to test that real differences existed between social class groups. Experimental logic demands that analysis under the latter two headings should precede informative discussion.

Table V gives the distribution of test scores/

/scores for the five social classes.

TABLE V.

DISTRIBUTION OF TEST SCORE BY SOCIAL CLASS.

Test Score.	Social Class A	Social Class B	Social Class C	Social Class D	Social Class E
75-	-	-	-	-	-
70-74	10	-	1	-	-
65-69	15	10	6	2	-
60-64	38	24	15	10	8
55-59	40	30	22	10	7
50-54	26	36	30	20	11
45-49	16	32	36	23	19
40-44	21	22	19	32	30
35-39	12	18	21	26	21
30-34	6	8	14	15	21
25-29	9	9	17	20	17
20-24	5	5	11	9	22
15-19	1	2	3	14	19
10-14	1	2	2	3	8
5- 9	-	1	3	7	9
0- 4	-	1	-	9	8
Totals	200	200	200	200	200
Means	51.7	47.8	43.9	37	32.4
SDs	12.8	12.2	13.4	15.7	15.4

/The social classes have been designated alphabet--ically in order to retain the appearance of hierarchy.

Table VI compares the test score means for the samples with those for the "universes" from which they were drawn. The mean test scores for the Thirty-Six-Day Sample occupational classes are reproduced from Social Implications. (1953, Table XIX, p.44).

TABLE VI.

MEAN TEST SCORE BY OCCUPATIONAL CLASS AND SOCIAL CLASS.

Thirty-Six-Day Sample Occupational Classes.	Mean Test Scores	Social Classes	Mean Test Scores
1	51.8	A	51.7
3	47.7	B	47.8
4	43.6	C	43.9
5	37.2	D	37
6 & 7	32.2	E	32.4

The mean test score for the combination of occupational classes 6 and 7 was obtained by multiplying in each case the given score by the given number in the class, summing the

two/

/two products, and dividing by the total number in both classes.

It is obvious by inspection that the differences between the scores for the occupational classes and those for the social classes are too small to be attributed to the operation of any factor other than chance. Confidence may therefore be felt that the samples are representative of the material from which they have been drawn. If this were not so, reference at a later stage of the investigation to the findings of the Mental Survey Committtee would be inhibited.

Once the exactitude of sampling had been established, the next step was to test the differences in mean test score between the five social classes for significance. It was discovered that all differences were significant at above the .01 level, in other words that there was less than one chance in a hundred that any difference was due to chance. The importance of this finding is not so much that it confirms many previous investigations/

/investigations with social groups, but that it lays the foundation for the later part of the inquiry. Investigation of the reasons for apparent differences in test performance between children of different social classes is inadvisable if there is no initial guarantee that the differences really exist.

For convenience the mean scores and standard deviations of score for the five social classes are reproduced in Table VII.

TABLE VII.

MEAN TEST SCORE AND STANDARD DEVIATION OF TEST SCORE BY SOCIAL CLASS.

Social Classes	Mean Test Scores	Standard Deviations
A	51.7	12.8
B	47.8	12.2
C	43.9	13.4
D	37	15.7
E	32.4	15.4

The relationship between the mean test score of the children and the social class to which they belong shows up very clearly in the above table/

/table. The range of mean test score between social classes is very large, amounting to 19.3 points of total score when Class A is compared with Class E. This difference is equivalent to a quarter of the total mark available on the test. It is also larger than the largest standard deviation, that of 15.7 points for Class D. Furthermore, the order of the social classes is exactly the order of the mean test scores of the children.

Of course there are many children with high or relatively high scores in Classes D and E, and conversely many with low or relatively low scores in Classes A, B, and C, as indicated by the size of the standard deviations. A glance at Table V confirms this point more directly. In Class A, for instance, some children score below the mean for Class E, while in Class E itself other children score above the mean for Class A. Details derived from the ungrouped scores are as follows.

In/

/In Class A 20 children, i.e. 10 per cent of the group, score 32 points or less: in Class E 20 children, or again 10 per cent of the group, score 52 points or more. The similarity in numbers is of course coincidental. But the numbers themselves are large enough to indicate that we cannot predict the performance of an individual child on an intelligence test very satisfactorily from a knowledge of his social class. It would probably be possible to predict, however, on the basis of an investigation such as this one, the proportions of children in different social classes to attain a particular score in a particular test or type of test. What does emerge from an examination of Table V is that in any test battery including a group intelligence test on the results of which selection is made for secondary education the children of the lower social classes are likely to do more poorly and be allocated to secondary modern and junior secondary schools. In practice/

/practice the social class composition of the grammar school is likely to be very different from that of the secondary modern school. Two recent investigations have provided some first-hand evidence. While carrying out a study to determine to what extent young adolescents viewed society in terms of adult frames of reference Halsey (1952) experienced great difficulty in obtaining a sample of upper and upper-middle class children from seven secondary modern schools in the Greater London area. Only 23 subjects falling into those class divisions could be located, in spite of aiming at a wide social spread. Differences between schools in social class composition were striking; 46 per cent of the grammar school sample came from middle-class homes, compared with only 19 per cent of the modern school sample. In a similar investigation Gardner (1953), comparing the social composition of four grammar schools and four secondary modern schools from comparable districts, concluded that the latter schools catered largely for/

/for the children of semi-skilled and unskilled manual workers. Only in one grammar school did middle-class pupils form less than half the total. And in one secondary modern school no social class above the lower-middle class was represented.

It may therefore be asserted that from the sociological point of view the fact that children of all social classes obtain high and low scores is of less importance than the fact that in general some social classes obtain less high mean scores than others.

A further interesting fact derives from an examination of Table VII, namely that differences of some size exist between the standard deviations of the social class groups. When this was noticed during the investigation these differences were immediately tested for significance. Findings are contained in Table VIII.

TABLE VIII.

LEVEL OF SIGNIFICANCE OF DIFFERENCES
BETWEEN STANDARD DEVIATIONS OF TEST SCORE.

Comparisons of Social Classes.	Differences in Standard Deviations.	Critical Ratios	Significances.
A-B	0.6	0.7	-
A-C	0.6	0.6	-
A-D	2.9	2.9	.01
A-E	2.6	2.6	.01
B-C	1.2	1.3	-
B-D	3.5	3.5	.01
B-E	3.2	3.3	.01
C-D	2.3	2.2	.05
C-E	2	1.9	-
D-E	0.3	0.2	-

Four differences are significant at above the .01 level, and one is significant at above the .05 level. There is a definite tendency for the standard deviations to be larger for the two working-class groups, and all significances occur when a middle-class group and a working-class group/

/group are compared.

There was a possibility that these differences in variation were due to a differential sex composition of the groups. It is well known that the range of score among a group of boys tends to be wider than among a comparable group of girls. The vagaries of selection might have produced a large preponderance of boys in Classes D and E, and a large preponderance of girls in Classes A,B and C. This had not happened, however. Class A had 100 boys and 100 girls; Class B 97 boys and 103 girls; Class C 99 boys and 101 girls; Class D 102 boys and 98 girls; and Class E 95 boys and 105 girls. Differences in sex distribution are insignificant, and certainly too small, to account for differences in variability of the order found in table VIII.

It seems therefore that the fact of greater variability of working-class groups must be accepted. There is nothing mysterious about this. Close examination/

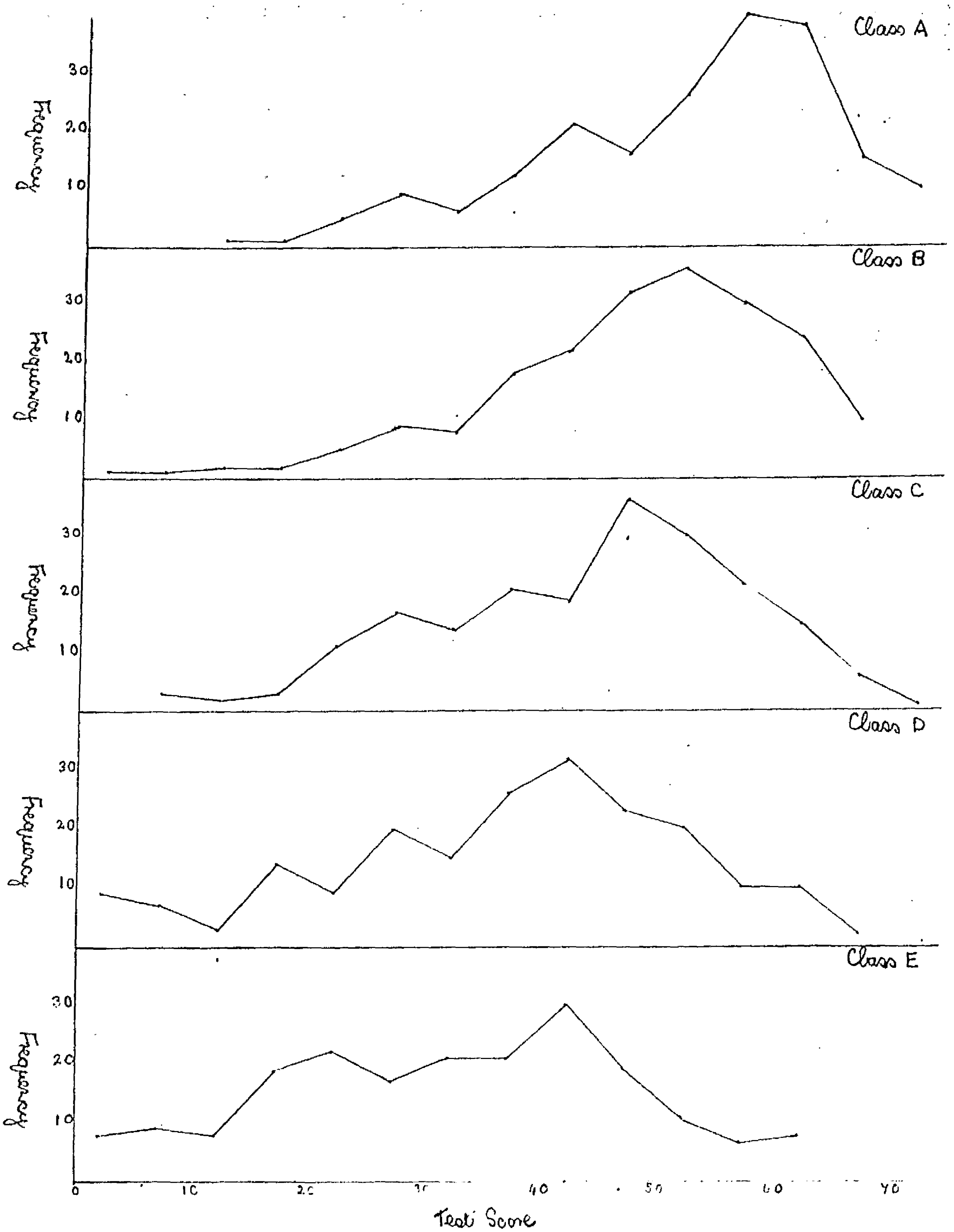

FIGURE 1. FREQUENCY POLYGONS SHOWING DISTRIBUTION OF TEST SCORE FOR FIVE SOCIAL CLASS GROUPS.

/examination of Table V would suggest that the primary cause is the presence within these groups of a comparatively large class of very low scorers. For instance, in Class E 17, and in Class D 16, children have a score of less than 10 points; while for Classes A,B and C the figures are 3,2 and 0 respectively. Admittedly high scorers are also distributed unevenly among Classes, but seemingly less unevenly than low scorers.

Figure 1 illustrates graphically the distribution of test score for the five social class groups. Distributions appear asymmetrical.

In general the findings derived from an examination of the data presented in this chapter are very similar to earlier findings on the social distribution of intelligence (Chapter VIII), although the social divisions imposed on the community are not quite the same as in any former investigation, and previous workers have often relied chiefly on the information to be derived from/

/from a study of group means. Thanks to the co-operativeness of the Mental Survey Committee, however, and also to the method employed in selection of groups, the differences found between social groups are more accurately defined than in most earlier investigations.

CHAPTER XV

VARIATION IN TEST SCORE WITH SOCIAL CLASS, LOCATION OF HOME, AND SIZE OF SCHOOL.

The data provided by the sociological schedules which were available for the 1000 subjects concerned in this investigation are of such a nature that it is possible to classify these subjects other than by the social or occupational class of fathers. Among the facts recorded are the kind of area in which the children's homes are situated and the number of pupils on the rolls of the schools attended by the children. It was therefore decided to investigate how test scores varied with location of home and size of school.

For the purpose of classification of children by/

/by the area in which their homes were situated Scotland was regarded as being divided into four types of area: cities, large towns, small towns, and Other Areas. The cities were Aberdeen, Dundee, Edinburgh and Glasgow, all well over the 100,000 level in total population. The large towns were the eighteen towns whose population exceeded 20,000 at the 1931 Census of Scotland. The small towns were the nineteen burghs which at the 1931 Census had populations less than 20,000 , but more than 10,000. The rest of Scotland was termed the Other Areas. It should not be thought that these Other Areas were necessarily rural in character; besides country districts, they included housing and industrial estates and mining villages (Social Implications, 1953, pp. 172-173)

There has been no attempt in the Mental Survey to use the information provided by the sociological schedules on size of school. The present investigator divided schools into three categories./

/categories. In the first category he placed schools with rolls between 200 and 650; and in the third category schools with rolls of less than 200.

The following tables show how the 1000 children were divided among the four categories of area and the three categories of school.

TABLE IX.

DISTRIBUTION OF ALL SUBJECTS
BY LOCATION OF HOME.

Location of Home.		n
1.	City	478
2.	Large Town	160
3.	Small Town	63
4.	Other Area	299
	Total	1000

TABLE X.

DISTRIBUTION OF ALL SUBJECTS
BY SIZE OF SCHOOL ATTENDED.

Size of School.		n
1.	650+	480
2.	200-649	399
3.	-199	121
	Total	1000

It/

/It is clear that the cities and Other Areas contain most of the subjects, and that only a small proportion attend schools with rolls of less than 200.

The distribution of mean test scores by location of home and size of school follow in Tables XI and XII.

TABLE XI.

MEAN TEST SCORE BY LOCATION OF HOME.

Location of Home.		n	Mean Score.
1.	City.	478	43.7
2.	Large Town.	160	40.4
3.	Small Town.	63	39.2
4.	Other Area.	299	42.4
		1000	

TABLE XII.

MEAN TEST SCORE BY SIZE OF SCHOOL ATTENDED.

Size of School.		n	Mean Score.
1.	650	480	42.7
2.	200-649	399	42.4
3.	-199	121	41.6
		1000	

It/

/It will be seen that mean score is highest for subjects living in cities, and lowest for subjects living in small towns. The difference between the two means is significant at the .05 level. This is a markedly different result from that obtained when the Thirty-Six-Day Sample was categorized by location of home. It was then found that mean score was lowest for subjects living in Other Areas (Social Implications, 1953, p. 174.) It would be unwise, however, to make too much of the variation in findings. Undoubtedly the results in this investigation have been influenced by the way in which subjects were selected so as to equalize numbers for five social class groups; it must be remembered also that the children of farmers and farm workers were excluded.

Mean test scores show a slight, but nowhere significant, tendency to decline as the number of children on school rolls falls. It is difficult to say how far this is a true indication/

/indication of inferiority of test performance on the part of children attending smaller schools. There is the possibility to be considered that children of the upper social groups tend to be taught in larger schools while children of the lower social groups tend to be taught in smaller schools. It will be seen from Table XIV that while the numbers of children from the five social classes attending schools with a roll of 650 and over vary little from class to class, the upper social classes are less well represented in the schools with a roll of between 200 and 650. On the other hand, they are on the whole better represented in the small schools, so no clear picture emerges.

It was decided that the variation of mean test score should be analysed in greater detail. Table XIII shows mean test score by social class by location of home.

TABLE XIII /

TABLE XIII.

MEAN TEST SCORE BY SOCIAL CLASS
BY LOCATION OF HOME.

Social Class.	Cities.	Large Towns.	Small Towns.	Other Areas.
A	54.6(96)	50.6(25)	46.5(17)	49.3(62)
B	48.5(96)	48.9(34)	42.1(9)	47.6(61)
C	43.5(108)	46.4(28)	44.2(10)	42.6(54)
D	37.4(94)	33.3(31)	32.7(14)	37.7(61)
E	32.8(84)	28.3(42)	30.7(13)	35.5(61)

The figures in brackets are the numbers from which the mean scores are derived.

It is clear from Table XIII that for each location of home the relative positions of the five social groups in respect of mean test score remain much the same; Class C subjects in small towns seem to score more highly than Class B subjects, but this may be due to sampling fluctuations. One fact, however, which does emerge is that the size of differences between social groups appears to decrease as we move along the scale from

cities/

/cities to Other Areas. A difference of 21.8 points between Class A and Class E subjects in cities declines to a difference of 15.8 points between Class A and Class E subjects in small towns, and a difference of 13.8 points between Class A and Class E subjects in Other Areas. This decline is too large to be attributed to chance factors; the difference of 8 points between the difference between means for Class A and Class E subjects in cities and the difference between means for Class A and Class E subjects in Other Areas is significant at the .01 level.

If the hypothesis that the four types of area vary in the amount of environmental "stimulation" which they provide for subjects is accepted , then it can be said that the least stimulating environment is that of small towns. Class A , B and D subjects whose homes are situated in small towns tend to return lower scores on the test than subjects from these/

/these groups living in cities, large towns, and Other Areas. The general picture presented in Table XI is thus confirmed to some extent. There are, however, interesting variations between the groups. In Class C the lowest mean score is obtained by subjects living in Other Areas, while in Class E subjects living in large towns make the lowest mean score.

It does seem, therefore, that no general thesis of environmental stimulation can be accepted, since the mechanism of stimulation appears to operate differently from class to class. For instance, in Class A, subjects living in cities are notably superior to subjects living elsewhere; but this is not true of any of the other social groups. If the performance of subjects living in large and small towns is ignored, leaving performance in cities and Other Areas to be considered, the following facts emerge; in Class A subjects in cities score more highly than subjects in Other Areas; in Classes B and C subjects in cities/

/cities score slightly more highly than subjects in Other Areas; in Class D subjects in Other Areas score slightly more highly than subjects in cities; and in Class E subjects in Other Areas score more highly than subjects in cities. In other words, as social class level declines the trend of differences between areas seems to reverse in direction.

Table XIV shows mean test score by social class by size of school attended.

TABLE XIV.

MEAN TEST SCORE BY SOCIAL CLASS
BY SIZE OF SCHOOL ATTENDED.

Social Class.	Large Schools.	Schools of Medium Size.	Small Schools.
A	53.3(96)	51.3(63)	49.2(51)
B	48(96)	47.3(76)	46.8(28)
C	43.8(98)	43.4(88)	44.1(14)
D	37(101)	37.1(85)	30.2(14)
E	30.8(91)	35.8(85)	27.8(24)

For each size of school the relative positions of the five social groups in respect of/

/of test score remain the same. The difference in mean score between Classes A and E is least for schools of medium size.

For Classes A and B mean score declines consistently as schools become smaller, and for all classes except Class C the lowest scores are obtained by subjects in small schools. The low scoring of Class D and Class E subjects in small schools is particularly noticeable. But in most cases differences between mean scores within a social class group are small.

If the hypothesis that different sizes of school provide differentially stimulating environments is accepted, it is then evident that the best environment for subjects in Class A is that of a large school, while for subjects in Class E the best environment is that of a school of medium size. For subjects in Class D the worst environment is that of a small school. For Classes B and C differences between subjects in schools of differing size are too small to justify/

/justify a statement.

As a check on the previous tables, Table XV shows mean test score by location of home by size of school attended.

TABLE XV.

MEAN TEST SCORE BY LOCATION OF HOME BY SIZE OF SCHOOL ATTENDED

Location of Home	Large Schools	Schools of Medium Size	Small Schools
Cities	43.8(302)	43.4(161)	44.7(15)
Large Towns	38.7(96)	43.5(57)	39.3(7)
Small Towns	38.5(24)	36.9(33)	38 (6)
Other Areas	43.2(58)	42.7(148)	41.5(93)

Numbers in some of the cells of this table are too small for much dependence to be placed on the mean scores obtained. Nevertheless, scores obtained by subjects living in small towns appear to be lower for every size of school than scores obtained by subjects living in other types of area.

An analysis of scores in the manner employed/

/employed in this chapter leaves too many variables uncontrolled for definite conclusions to be reached. Yet there does seem a case for asserting that "intelligence" is differentially stimulated by different environments, and that the nature of this differential stimulation varies between social classes.

CHAPTER XVI.

ERRORS IN TEST MARKING.

Gross differences in intelligence test performance between social classes are capable of being interpreted in a number of different ways. The latitude of choice offers sufficient scope for subjective predilection and personal prejudice. It seemed therefore to the experimenter that the logical next step in the investigation was to analyse the performance of the social class groups at a more detailed level than has been the practice in this country. To this end two item-analysis charts were drawn up for each group. The first showed how each subject tackled each item, whether he marked it correctly or incorrectly, omitted it, or had insufficient time to reach it. The second recorded/

/recorded the nature of all his incorrect responses. It was hoped that when completed these charts would provide a complete picture of the performance of every subject in every group on every item.

While carrying out the necessary routine work involved, it was discovered that numbers of residual errors of marking were still present in test papers, although these had been checked when first corrected during the 1947 Mental Survey. The following quotation describes the method of checking them adopted. " It was found that it was essential to have the marking checked. On the first check a small, but not negligible amount of error (mainly omitted items and wrong addition) was found and rectified. A number of scripts were subjected to a second check, and though a few errors were still found, these were sufficiently small and infrequent to have no effect on the final findings of the survey". (The Trend of Scottish Intelligence, 1949, p.p. 71-72). And, later, it is said, "....it can be concluded that any comparisons or measures made from the test results will/

/will not be vitiated to any significant extent by marking error".(ibid. p.73).

In effect, the Mental Survey Committee recognized the probable existence of residual errors, but felt that their presence would have little effect on the scores of a large aggregation, and that the small gain in accuracy provided by their elimination or near-elimination would not compensate for the time and labour spent on this task. However, when responses to individual items are being recorded, these errors are easily noticed, and have to be corrected in order that a true picture of item response may be presented.

Another problem which arose while tabulating item responses was not solved so easily. It was found that the marking of some answers demanded a considerable exercise of judgment on the part of markers such that they differed very noticeably in their assessment of these answers as correct or incorrect. For instance, Item 56(a) runs as follows.

> "Underline the ONE of the four answers to (this) statement which seems to you to be correct:-
> Vitamins are found in --(fresh milk and fruits, lard, dried fruits, stale bread)."

Purely/

/Purely from the point of view of dietetic fact, this does not seem a very satisfactory item. But obviously the expected response is "fresh milk and fruits". Now this response differs from the others not only in that it is the correct response. It is also longer than the other responses, and it involves two categories of food in which vitamins may be found. In addition, it is capable of being understood in two different ways. It may be read as "fresh milk and (simply) fruits" or "fresh milk and (fresh) fruits".

The responses of subjects to Item 56(a) reflected its ambiguity. Of those who responded "fresh milk and fruits" some underlined the whole phrase, others only "fresh milk", and most of the remainder either "fruits" or "and fruits". A few merely placed a tiny line vaguely underneath the phrase. And what one marker passed as correct, another marked as incorrect.

Because of this subjective variation in marking it was decided that on a number of items the/

/the apparent proportion of correct responses was not representative of the true proportion. The experimenter therefore re-marked them, setting up his own standards beforehand of what was to constitute a correct response. Since personal error could not be eliminated from the data, it seemed best to make it a constant factor.

It will be apprehended that a fair number of markings had to be altered because of demonstrable marking error or to conform to the experimenter's standards. A breakdown of error markings for the five social classes is shown in Table XVI.

TABLE XVI.

DISTRIBUTION OF MARKING ERRORS
BY SOCIAL CLASS

Social Class	All Error Markings	Individual Papers in Error	Average Errors per Paper
A	50	36	1.4
B	55	40	1.4
C	141	88	1.6
D	157	104	1.5
E	171	112	1.5

This table has not been included here purely for its intrinsic interest, although it may be of/

/of value to future workers investigating sources of error in test assessments. Its main purpose, and that of Table XVII, is to test a hypothesis which might conceivably be put forward to account for at least a part of the differences in test score between social class groups, namely that errors and subjective decisions in marking are not determined by chance factors but by conscious or unconscious bias in favour of the children of the upper social classes and against children of the lower social classes. The idea behind the hypothesis would fall into the following logical pattern.

1. All people in a class society tend to favour members of their own class.
2. Test papers will usually be marked by middle-class people.
3. Test markers will tend to favour middle-class subjects.

If it were asked how a marker would know whether an individual paper had been completed by a middle-class subject or not, it might be retorted that many of the markers were acquainted with/

/with the children with whose papers they were dealing, and that in any case people living in our society were particularly sensitized to the detection of class differences even with such limited evidence as name of school attended, style of calligraphy etc.

Certainly Table XVI indicates that the number of errors and subjective variations of marking increases steadily as we descend the social ladder from group to group, although by the hypothesis we should expect a steep increase from Class C to Class D rather than from Class B to Class C. But the real disproof of the hypothesis is contained in Table XVII, which shows the actual changes in mean score and standard deviation brought about by the re-marking. It may be noted here that a strong tendency was found for upward and downward revisions of score to cancel each other out.

TABLE XVII./

TABLE XVII.

CORRECTED AND UNCORRECTED MEAN TEST SCORES AND STANDARD DEVIATIONS BY SOCIAL CLASS.

Social Classes	Means	Standard Deviations	Corrected Means	Corrected Standard Deviations.
A	51.7	12.8	51.8	12.8
B	47.8	12.2	47.6	12.6
C	43.9	13.4	43.7	13.5
D	37	15.7	36.6	15.5
E	32.4	15.4	32.6	15.4

The changes in mean score and standard deviations are insignificant. Classes A and B show slight increases of 0.1 and 0.2 points of score respectively, while Classes B, C and D show slight decreases of 0.2 points, 0.2 points and 0.4 points. The effect of correction on the standard deviations is even less than on the mean scores. There is not the slightest evidence of bias on the part of the markers.

Two qualifications may be entered. The possibility is not ruled out that an individual marker here and there may show bias. And there is/

/is no guarantee that the errors and subjective variations discovered are typical of the whole category of error when the test was marked for the first time. But it is clear that so far as the evidence from the groups of subjects included in the present inquiry is concerned no part of the differences in test score between social classes may be attributed to the conscious or unconscious prejudices of markers.

If finest accuracy is regarded as desirable in the assessment of intellectual performance, then the corrected means and standard deviations given in Table XVII should be preferred to the original statistics. In the present instance it matters little which estimates are accepted, since when the significances of differences between means and standard deviations are re-calculated their levels remain unchanged.

The best explanation of the data seems simply to be that errors in marking increase as subject errors increase, the parallelism with social classification being coincidental.

CHAPTER XVII.

METHODS USED IN ITEM ANALYSIS: SOCIAL CLASS DIFFERENCES IN PERFORMANCE ON INDIVIDUAL ITEMS

Once the item-analysis charts had been drawn up it remained to consider the best method of developing the raw data thus made available. It was decided in the first place to concentrate attention on the relative ability of the five groups to answer individual items correctly. As an estimate of the difficulty of an item for a particular group the number of subjects answering the item correctly was expressed as a percentage of the total number of subjects in the group.

In work of this kind the problem has always to be faced that the number of subjects who actually/

/actually know the correct answer to an item is less than the number who mark the correct answer. Guessing, chance responses, slips of the pencil, all go to swell the number of correct responses. Some educational statisticians suggest that obtained percentages should always be adjusted downwards in order to eliminate correct responses from pupils who "do not know" the answer. (Davis, 1946). Certainly it would be very convenient to be sure that all the responses in a test which had been classed as correct had been provided by subjects who were aware they were correct. But the suggested methods of obtaining certainty, although statistically sound, rest on doubtful psychological assumptions. (Eells et al.,1950) Although for practical purposes a distinction can be drawn between the subject who marks a correct response and the subject who marks an incorrect response, that does not mean that one subject is perfectly knowledgeable and the other absolutely ignorant. For any item there exist many/

/many gradations of knowledge and ignorance. A few subjects will be absolutely certain of the correct response, and nothing less than sheer accident, mementary nervous aberration, etc. will prevent them from marking it . At the other end of the continuum a small group will possess no relevant knowledge, and will only mark the cor--rect response by chance. For the majority, less than absolutely certain, less than completely ignorant, a subtle interplay of factors develops, such that highly individual, personal, idiosyncratic motivations mat determine their responses.

Psychologically speaking, the intellectual position of subjects about to answer items should therefore not be dichotomized into knowledge and ignorance, but should be treated as a continuous variable with an infinity of intermediate status.

In addition, false knowledge, or the idea that a response which objectively considered is the wrong response is correct, must be distinguished from ignorance. This may be regarded as equivalent to a subjective mental set, possibly based on environmental determinants. If the set is very powerfully established, it is conceivable that it may block out consideration of a correct response of whose rightness a subject would become intellectually aware if brought to his/

/his attention.

It would thus appear that the statisticians who recommend correction for chance responses are accepting an unduly simplified picture of the test situation. It was therefore decided not to adjust obtained percentages.

Now the use of a percentage scale as a measure of performance creates its own difficulties. A percentage scale does not usually constitute a linear scale with respect to the trait or traits being measured. Differences between two percentages do not have the same meaning at different points along the scale. If the exact distribution of the traits underlying each test item were known this flaw could be repaired. As such information is not available the choice must be made between assuming rectilinear distribution and retaining the percentage scale and assuming normal distribution and converting the percentage scale into a normalized set of scaled values of such a form that equal differences between two measures will represent equal differences in difficulty no matter at what point along the scale they may be taken. During the present investigation it was decided to employ both types of scale, the purpose for which information was required determining in each particular instance which was to be used.

There/

/There are several systems of scaled scores in common use, almost without exception based on the adaptation of the standard deviation of the distri--bution as a unit and a fitting of score values to the normal curve. The particular system selected here is that proposed by F.B.Davis, which he calls a "Difficulty Index". (Davis, 1946, pp. 7-8)

He uses the Kelley-Wood Table of the Normal Probability Integral to transform percentages or proportions into standard-deviation units. Each standard score is then multiplied by a constant---- 21.066-so chosen as to make the largest become 49. 50 is then added algebraically to each product.

The complete system of conversions from percentages to normalized scores is contained in Table XVIII. Where it was necessary to find intermediate values, this was done by simple interpolation.

TABLE XVIII.

DAVIS' "DIFFICULTY INDEX" FOR CONVERSION OF PERCENTAGES TO NORMALIZED SCORES.

Percentage.	Normalized Index Value.
0	0
1	1
2	7

contd.

contd. TABLE XVIII.

Percentage	Normalized Index Value.
3	10
4	13
5	15
6	17
7	19
8	20
9	22
10	23
11	24
12	25
13	26
14	27
15	28
16	29
17	30
18	31
19	32
20	32
21	33
22	34
23	34

contd.

contd. TABLE XVIII.

Percentage	Normalized Index Value.
24	35
25	36
26	36
27	37
28	38
29	38
30	39
31	40
32	40
33	41
34	41
35	42
36	42
37	43
38	44
39	44
40	45
41	45
42	46
43	46
44	47

contd.

contd.

TABLE XVIII.

Percentage	Normalized Index Value.
45	47
46	48
47	48
48	49
49	49
50	50
51	51
52	51
53	52
54	52
55	53
56	53
57	54
58	54
59	55
60	55
61	56
62	56
63	57
64	58
65	58

contd.

contd.

TABLE XVIII.

Percentage	Normalized Index Value.
66	59
67	59
68	60
69	60
70	61
71	62
72	62
73	63
74	64
75	64
76	65
77	66
78	66
79	67
80	68
81	68
82	69
83	70
84	71
85	72
86	73

contd.

contd. TABLE XVIII.

Percentages	Normalized Index Value.
87	74
88	75
89	76
90	77
91	78
92	80
93	81
94	83
95	85
96	87
97	90
98	93
99	99
100	100

When the percentages of subjects within each group correct on each item had been worked out these were tabulated as in Table XIX.

TABLE XIX/

TABLE XIX.

DISTRIBUTION OF TEST ITEMS BY DIFFICULTY OF ITEMS FOR FIVE SOCIAL CLASS GROUPS.

Number of Items Answered Correctly---

	Class A	Class B	Class C	Class D	Class E
by 95-99% of the Group.	9 items	3 items	-	-	-
" 90-94% " " "	7 "	8 "	5 items	3 items	1 item
" 85-89% " " "	7 "	7 "	10 "	-	2 items
" 80-84% " " "	6 "	6 "	6 "	5 "	-
" 75-79% " " "	3 "	2 "	2 "	4 "	4 "
" 70-74% " " "	4 "	7 "	6 "	7 "	8 "
" 65-69% " " "	8 "	2 "	2 "	6 "	2 "
" 60-64% " " "	3 "	4 "	5 "	3 "	5 "
" 55-59% " " "	10 "	7 "	4 "	3 "	4 "
" 50-54% " " "	5 "	5 "	5 "	3 "	2 "
" 45-49% " " "	2 "	7 "	8 "	5 "	3 "

(contd.)

TABLE XIX (contd.)

Number of Items Answered Correctly—

	Class A	Class B	Class C	Class D	Class E
by 40-44% of the Group.	3 items	5 items	6 items	7 items	5 items
" 35-39% " " "	1 item	3 "	4 "	4 "	10 "
" 30-34% " " "	2 items	3 "	4 "	7 "	6 "
" 25-29% " " "	1 item	-	2 "	7 "	9 "
" 20-24% " " "	1 "	3 "	1 item	4 "	3 "
" 15-19% " " "	2 items	1 item	3 items	2 "	6 "
" 10-14% " " "	1 item	1 "	2 "	4 "	3 "
" 5- 9% " " "	1 "	2 items	1 item	2 "	1 item
" 0- 4% " " "	-	-	-	-	2 items
Totals	76 items	76items	76 items	76 items	76 items

That the test as a whole becomes more difficult as the social level of subjects declines makes itself evident when the relative proportions of items passed by 50 per cent or more of the subjects are compared.

TABLE XX.

PROPORTION OF ITEMS PASSED BY 50 PER CENT OR MORE OF SUBJECTS IN FIVE SOCIAL GROUPS.

Social Class	A	B	C	D	E
Number of Items Passed by 50% or more of Subjects.	62	51	45	34	28
Proportion of all Items passed by 50% or more of Subjects.	81.6%	67.1%	59.2%	44.7%	36.8%

Speaking objectively, most of the items would appear very easy for Class A, fairly easy for Classes B and C, and rather difficult for Classes D and E. It is interesting to note that the drop from Class A to Class B is equivalent to that from Class C to Class D, and that both are considerably larger than that from Class B to Class C or Class D to Class E. That the difference between Classes C and D should be large is understandable; reference back to Table VII shows that at this point in the social scale the decrease in mean test score is also most marked. The reason for the size of the drop from Class A to Class B is not so easy to find. Inspection of Table XIX suggests that a major/

/major difference in the distribution of item difficulty for the two classes lies in the strikingly larger proportion of items which appear very easy for subjects in Class A i.e. are answered correctly by 95 per cent or more. An examination of the item-analysis charts shows that the nine items answered best by Class A were also found easy by Class B, although falling below the 95 per cent level ; and that the individuals in Class B who failed to mark correct responses to these items often succeeded on objectively more difficult items.

A hypothesis explaining the data of Tables XIX and XX may therefore be put forward to run as follows. The reactions of subjects answering a test item range through a continuous series of gradations from full knowledge to complete ignorance. When the item performances of groups of varying "intelligence levels" are being compared, sharp changes will be noted when either of two requirements is met; firstly when a group including a sizeable number of individuals who are absolutely certain of the correct answers to a sizeable number of items is compared with another group in which the comparable best individuals are almost, but not quite, certain of the answers; secondly, when a group including a sizeable number of individuals who are absolutely ignorant/

/ignorant of the answers to a sizeable number of items is compared with another group in which the comparable worst individuals are almost, but not quite, ignorant of the answers. Whenever certainty is replaced by near-cert--ainty, knowledge is no longer the sole determining factor in response; non-intellectual factors immediately come into play. Whenever ignorance is replaced by near-but-not-quite ignorance, chance is no longer the sole determining factor in response; the same non-intellectual, and also non-chance, factors begin to operate.

Thus, when Classes A and B are compared, the first requirement of the hypothesis is met; when Classes C and D are compared, the alternative requirement. Striking differences in item performance are not so noticeable when comparison is made between Classes B and C or Classes D and E. In the first case there is probably no sizeable group in either class either completely knowledgeable or completely ignorant, and similar determinants govern response for both classes. In the second case there is probably a sizeable group in both classes which is completely ignorant with regard to some items, and again similar determinants govern response for both classes.

Table XXI records the mean item difficulty and the standard deviation of item difficulty for each class.

TABLE XXI /

TABLE XXI.

MEAN ITEM DIFFICULTY AND STANDARD DEVIATION
OF ITEM DIFFICULTY BY SOCIAL CLASS.

Social Classes.	Mean Item Difficulty.	Standard Deviation of Item Difficulty.
A	67.3	23
B	61.6	23.8
C	57.9	23.5
D	48.5	23.2
E	43.4	22.3

So far as the evidence provided by differences between means is concerned, this table corresponds very closely to Tables V, VI, and VII. As the social scale is descended, the test becomes more and more difficult; mean difficulty for both Classes D and E is below 50. At the same time examination of the standard deviations confirms that for all five classes there is considerable variation in difficulty between the items in the test.

It is noticeable that the size of standard deviation does not appear to vary greatly from class to class. When the significances of differences between standard deviations/

/deviations were computed the highest Critical Ratio (obtained on the comparison of Classes B and E) was only 0.56 which falls far below even the .05 level of significance. A priori it might be expected that variability in some classes would be greater than in others since comparisons of the standard deviations of the children's test scores produced significant results. But this is not so. It appears that there is a normal level of achievement for each group i.e. the test as a whole is found to be of a certain difficulty, and items always distribute themselves around the norm in very much the same way.

It is clear that the examination of the test behaviour of different social classes in terms of item response, while in general confirming information derived from the study of test scores, supplements it in valuable ways and points the way towards a truer understanding of the complex factors underlying differences in performance. In succeeding chapters this examination of item response will be carried out in greater detail.

CHAPTER XVIII.

FAILURE TO REACH ITEMS: SOCIAL CLASS DIFFERENCES

Chapter XVII has illustrated general differences between social class groups in item performance, or, to put it in another way, the general "behaviour" of items in different social contexts. Before going on to consider more specific differences, it is advisable to pause here to examine categories of response which are often ignored in test analysis. There are three different ways in which subjects may fail to answer an item correctly. They may simply answer the item wrongly; they may omit the item; they may fail to reach the item. This chapter/

/chapter and the next offer an analysis of these three categories of incorrect response for the five social class groups. In the text the three types of response will be defined as "wrongly answered", "omitted" and "unreached".

Table XXII gives complete information for all five groups on cases where subjects failed to reach items.

TABLE XXII.

DISTRIBUTION OF "UNREACHED RESPONSES" BY SOCIAL CLASS.

No. of Item.	Number of Subjects not Reaching Item in -- Class A	Class B	Class C	Class D	Class E
8	-	-	-	-	1
9	-	-	-	-	1
10	-	-	-	-	1
11	-	-	-	1	1
12	-	-	-	1	1
13	-	-	-	1	1
14	-	-	1	1	1
15	-	-	1	1	1
16	-	-	1	2	1
17	-	-	1	2	2

(contd..)

TABLE XXII (contd..)

Number of Subjects not Reaching Item in--

No. of Item	Class A	Class B	Class C	Class D	Class E
18	-	-	1	2	2
19	-	-	1	2	2
20	-	-	1	2	2
21	-	-	1	2	3
22	-	1	1	2	3
23	-	1	1	2	3
24	-	1	1	2	3
25	-	1	1	2	3
26	-	1	1	2	3
27	-	1	1	2	3
28	-	1	1	2	4
29	-	1	1	2	4
30	-	1	1	2	4
31	-	1	1	3	4
32	-	1	1	4	5
33	-	1	1	4	5
34	-	1	1	4	5
35	-	2	1	4	6
36	-	2	1	4	6

(contd..)

TABLE XXII (contd..)

Number of Subjects not Reaching Item in--

No. of Item	Class A	Class B	Class C	Class D	Class E
37	-	2	1	4	7
38	-	2	1	5	7
39	-	2	1	5	8
40	-	2	1	5	8
41	-	2	1	5	8
42	-	2	1	5	8
43	-	2	1	5	8
44(a)	-	2	2	5	9
(b)	-	2	2	5	9
(c)	-	2	2	5	9
45	-	2	2	5	10
46	-	2	2	5	12
47	-	2	2	6	14
48	-	2	2	6	14
49	-	3	2	8	14
50	-	3	3	8	14
51	-	3	3	8	15
52	-	3	3	8	16

(contd..)

TABLE XXII (contd..)

No. of Item	Class A	Class B	Class C	Class D	Class E
	Number of Subjects not Reaching Item in--				
53	-	4	3	9	17
54	1	4	4	11	19
55	1	4	4	13	19
56(a)	1	5	4	13	19
(b)	1	5	4	13	19
(c)	1	5	5	14	19
57	2	8	6	20	19
58	3	9	7	20	20
59	4	11	7	21	23
60	4	11	8	21	23
61	4	11	9	21	24
62	5	11	11	22	25
63	8	11	15	27	26
64	8	11	18	27	29
65	8	12	20	27	29
66(a)	9	12	22	35	33
(b)	9	12	22	35	33
67	11	15	27	43	36
68	14	20	27	47	41
69	18	25	34	52	51

(contd..)

TABLE XXII (contd..)

Number of Subjects not Reaching Item in--

No. of Item	Class A	Class B	Class C	Class D	Class E
70	21	31	52	64	56
71	25	35	58	70	67

In 31 cases a subject tackled Item 70 but offered no response for Item 71. There was no means of telling whether or not he had considered the latter item. It was decided to regard these cases as "unreached" rather than "omitted".

In general Table XXII shows an increasing number of subjects not reaching items as the social scale is descended, presumably indicating a declining speed of performance. At least a part of this decline is undoubtedly due to the fact that each group finds the items as a whole more difficult than the groups above. There are, however, a number of variations from the general pattern. At certain points in the test there are more subjects in one group who have failed to reach a particular item than in the group immediately below. For the early items variations of/

/of this kind can probably be traced to the chance inclusion of a few specially slow or specially "stupid" subjects in a group. Whether this interpretation can be applied to the result of a comparison of Classes D and E on the later items in the test is not so evident. More subjects in Class D than in Class E consistently fail to reach a point in the test later than Item 65. It has already been shown that the difference in mean test score between Classes D and E is almost the same as the difference between Class B and C. Yet for the same group of items i.e. these following Item 65 a smaller number of individuals are classified in the "unreached" category for Class B than Class C.

It is tempting to attribute the aberrant pattern which develops when Classes D and E are compared to fundamental class-linked mental characteristics. And such an explanation may conceivably be correct, although the intervening variables between social class membership and not reaching items are so complex as to render its experimental investigation difficult. What is more/

/more likely is that the pattern is a test artefact. It may be assumed that the sum of knowledge at the disposal of Class D is slightly greater than that at the disposal of Class E. Subjects in Class D may therefore spend a little more time worrying out the correct answers to items on which they possess a certain amount of hazy information. Subjects in Class E are more likely to omit these items immediately and hurry through the test, pausing at other items which appear easier. Thus, although in general they are slower in answering the items which they do tackle, a somewhat larger number contrive to reach the final part of the test. On this hypothesis it will be expected that the number of omissions in Class E will be larger than might be anticipated purely from the consideration of mean test score.

Further examination of Table XXII shows that at certain points in the test the number of subjects who fail to reach items increases sharply, and that these points are not always the same for the different groups. Increases from item to/

/to item from Classes A and B are very regular, and little can be said about them. But for Class C there is a noticeably sharp increase from Item 69 to Item 70; for Class D relatively sharp increases occur from Item 56(c) to Item 57, from Item 65 to Item 66(a), from Item 66(b) to Item 67, and again from Item 69 to Item 70; for Class E, where the distribution is again noticeably regular, the sharpest increases are from Item 68 to Item 69 and from Item 70 to Item 71. As would be expected, all the points of sharp increase fall towards the end of the test, only one occurring other than in the last six items. It would be unwise to regard the fact that these sharp increases take place at slightly different points in the test for different groups as being other than a reflection of the differential difficulty of the test for the groups and the probable variations between them in speed of work.

If a subject does not reach an item, he can hardly be considered as having responded to it. To avoid inelegant circumlocutions, however, it was/

/was decided, when comparing the relative distributions of the three types of incorrect answer among the five social classes, to talk occasionally of "unreached responses".

Table XXIII shows that as the social scale is descended from group to group "unreached responses" constitute a steadily increasing proportion of all incorrect responses. In other words the evidence is not only that the lower groups fail to reach a greater number of items, which would be expected, but that failure to reach items increases more rapidly with decline in social level than omission and wrong answering of items.

TABLE XXIII.

CONTRIBUTION OF "UNREACHED RESPONSES" TOTAL INCORRECT RESPONSES FOR FIVE SOCIAL CLASS GROUPS

	Class A	Class B	Class C	Class D	Class E
(1) Total Incorrect Responses	4840	5680	6458	7881	8677
(2) Total "Unreached Responses"	158	329	422	786	918
(3) (2) as % of (1)	3.3%	5.8%	6.5%	10%	10.6%

If the number or proportion of "unreached responses" is to be taken as an indication of the test/

/test speed of a group, then not only is each group slower than its predecessor in the order, but the decrease in speed is greater than would be expected by inference from wrong answers and omissions. Of course, test speed is not independent of test difficulty. Nevertheless, this is an interesting finding.

Differences in speed emerge most clearly when Class C is compared with Class D, and Class A with Class B, while the middle-class groups taken as a whole contrast strongly with the working-class groups.

It is possible to arrive at some kind of estimate of the extent to which differences in mean test score between groups are dependent on the increase in "unreached responses" for the lower groups.

TABLE XXIV /

TABLE XXIV

CONTRIBUTION OF "UNREACHED RESPONSES" TO DIFFERENCES IN MEAN TEST SCORE BETWEEN FIVE SOCIAL CLASS GROUPS (1)

	Classes A–B	Classes B–C	Classes C–D	Classes D–E
(1) Differences in mean test score	4.2	3.9	7.1	4
(2) Differences in total incorrect responses	840	778	1423	796
(3) Differences in total "unreached responses"	171	93	364	132
(4) (3) as % of (2)	20.4%	11.9%	25.6%	16.6%
(5) Parts of differences in mean test score attributable to differences in total "unreached responses"	.9	.5	1.8	.7

The data for this table are derived in the following manner. All incorrect responses and "unreached responses" are totalled seperately for each ~~comparison of~~ group. Differences between these totals are tabulated for each comparison of groups. The differences for the "unreached responses" are then taken as percentages of the differences for/

/ for all incorrect responses. The appropriate percentage is used in each comparison of groups to obtain the amount of difference in mean test score attributable to the greater number of "unreached responses" by subjects in the lower group.

It is clear that in each case where one group is compared with a substantial part of the difference in mean score can be attributed to the increase in "unreached responses" for the lower social group, and this is especially true in the comparison of Classes C and D, the lowest middle-class group and the upper-working-class group. The general picture emerges more clearly in Table XXV, where the social distance between the groups being compared steadily increases.

TABLE XXV.

CONTRIBUTION OF "UNREACHED RESPONSES" TO DIFFERENCES IN MEAN TEST SCORE BETWEEN FIVE SOCIAL CLASS GROUPS (2)

	Classes A-B	Classes A-C	Classes A-D	Classes A-E
(1) Differences in mean test score	4.2	8.1	15.2	19.2
(2) Parts of differences in mean test score attributable to difference in total "unreached responses"	.9	1.3	3.1	3.8

It is evident that if an untimed version of the test were administered the differences in test score between the groups would be lessened, even though subjects in the lower groups were guilty of more wrong answers and omissions on the freshly completed items.

One further question regarding "unreached responses" is due for discussion at this point. For some groups the number of subjects who have not been able to reach the latter items on the test constitutes a high proportion of all subjects. It is probable that the proportion of correct responses given to these items is an underestimation of the "real" ability of subjects to answer them correctly. Is the inclusion of data in later chapters derived from the analysis of these items justified? Faced with a similar problem, Eells (1948) determined to exclude from consideration any item not reached by more than 5 per cent of the subjects in either of two groups being compared. This is too drastic a procedure to apply in the present investigation. It would mean that all items/

/items after Item 45 would have to be excluded in any comparison in which Class E was involved. Out of 760 possible comparisons of item responses for the five groups, 264, or 34.7%, would have to be ignored. This is too high a price to pay for reliability, especially since items are not being excluded at random, but always towards the end of the test.

It will be found that in the present study all items have been analysed for all comparisons of groups, but the evidence of items which many subjects failed to reach has been treated with caution.

CHAPTER XIX.

OMISSION OF ITEMS: SOCIAL CLASS DIFFERENCES.

The three types of incorrect response have been defined as "wrongly answered", "omitted" and "unreached". This chapter deals with omissions in the same general way as Chapter XVIII dealt with "unreached responses". It is assumed throughout that wherever an item has been omitted by a subject he has considered it but been unable to decide on a response. There must undoubtedly be cases where an omission takes place either because a subject fails to notice an item or, having noticed an item and determined his response, fails to record it/

/it. But these should constitute only a small part of the total category of omissions.

Table XXVI provides complete information on omissions for all five groups.

TABLE XXVI

DISTRIBUTION OF OMISSIONS
BY SOCIAL CLASS

No. of Item	Number of Subjects Omitting Item in-- Class A	Class B	Class C	Class D	Class E
1	6	10	9	13	15
2	2	1	2	5	4
3	2	3	6	10	18
4	7	6	10	16	19
5	6	3	8	13	11
6	9	8	14	13	22
7	3	4	7	14	16
8	7	8	8	15	16
9	14	27	35	51	67
10	5	4	7	18	13
11	4	2	10	13	12
12	4	2	5	11	15
13	1	1	2	9	9

(contd..)

TABLE XXVI (contd..)

No. of Item.	Number of Subjects Omitting Item in-- Class A	Class B	Class C	Class D	Class E
14	7	6	8	17	17
15	4	6	5	16	20
16	11	11	11	14	23
17	11	14	25	42	54
18	14	17	27	40	49
19	6	2	8	15	13
20	4	2	5	11	20
21	3	3	2	6	10
22	12	20	19	32	38
23	4	5	13	28	32
24	7	8	13	29	35
25	6	8	10	29	33
26	3	6	4	13	12
27	7	9	5	11	13
28	52	52	69	62	79
29	13	24	33	56	61
30	18	19	28	30	39
31	7	4	8	8	15

(contd..)

TABLE XXVI (contd..)

No. of Item.	Class A	Class B	Class C	Class D	Class E
	Number of Subjects Omitting Item in--				
32	4	7	5	21	23
33	1	4	7	17	21
34	12	8	13	30	25
35	55	87	92	116	125
36	8	3	5	29	17
37	7	10	7	26	8
38	14	11	19	31	24
39	11	15	20	24	35
40	25	35	35	46	45
41	1	4	5	16	18
42	2	3	4	14	21
43	1	3	4	11	18
44(a)	11	11	22	46	52
(b)	18	34	53	77	85
(c)	9	18	33	46	55
45	9	6	5	15	14
46	12	11	15	19	28
47	5	6	20	35	33
48	16	7	21	38	40

(contd..)

TABLE XXVI (contd..)

No. of Item.	Class A	Class B	Class C	Class D	Class E
	Number of Subjects Omitting Item in—				
49	2	1	4	10	11
50	7	6	8	19	16
51	18	19	30	40	40
52	3	3	3	10	10
53	10	13	14	30	26
54	12	11	13	34	34
55	21	31	43	52	50
56(a)	13	11	17	36	35
(b)	7	7	14	24	27
(c)	13	19	27	34	37
57	42	60	67	87	81
58	22	17	33	42	38
59	34	39	60	67	77
60	3	3	0	4	9
61	5	7	8	5	10
62	3	3	1	4	8
63	21	30	30	31	38
64	71	76	89	63	68
65	68	79	75	50	50

(contd..)

TABLE XXVI (contd..)

No. of Item	Number of Subjects Omitting Item in — Class A	Class B	Class C	Class D	Class E
66(a)	65	90	99	106	115
(b)	65	91	99	106	115
67	10	20	18	39	46
68	25	19	22	25	24
69	10	11	8	12	21
70	38	46	42	49	50

In all cases where no response was recorded for Item 71 a subject was regarded as not reaching the item.

It will be noted that, while there is a tendency for the number of omissions to increase towards the end of the test, as would be expected, it is not very marked. Certain items and certain groups of contiguous items stand out as eliciting no response from subjects. Among these are Items 9, 17 and 18, 22, 28, 29 and 30, 35,38,39 and 40, 44(a), (b) and (c), 48, 51, 53, 54, and 55, 57, 58, and 59. Among the items towards the end, most of/

/of which tend to be omitted by many subjects, Items 66(a) and 66(b) are noteworthy. It must be remembered that the numbers of omissions for items in the latter half of the test, especially with Classes D and E, would probably have been increased had more time been given.

Reference to the basic analysis data regarding these items indicates that in general the explanation of their frequent omission is simply that subjects find them difficult. This is not always so, however. Although often omitted, Items 29, 44(a) and 54 are not particularly difficult, relatively speaking, for subjects in any group. The best explanation here seems to be that they are associated in the test with other items which are found difficult. Item 29, an easy item, lies between Items 28 and 30, which are difficult; Item 44(a) is next to Items 44(b) and 44(c), which are difficult; and Item 54 lies between items 53 and 55, which are likewise difficult.

There is a general pattern of increase in omissions/

/missions for each item, as social level declines, although there are minor exceptions here and there. But two items, Items 64 and 65, offer a contrast to this pattern which is surely too striking to be accidental. Instead of the normal increase from Classes C to D, there is a sharp decrease in omissions. Admittedly there are fewer subjects in Class D left to omit, but that applies to succeeding items as well and they do not show the same trend.

This is a baffling anomaly. The two items are reproduced below

> Look at the word in front of the bracket, and in the bracket find one word which is either nearly the same, or nearly the opposite. Underline it if it is the same, cross it out with an X if it is opposite.
>
> 64. legislature..............(executive, municipal, parliament, court, palace)
>
> 65. oscillate...............(bring, swing, king, sing, bright)

Both items are very difficult. The correct response for Item 64 is provided by 12 subjects in Class A, 15 in Class B, 11 in Class C, 19 in Class D, and 20 in Class E. The comparable figures for Item 65 are 21, 15, 21, 21, 18. In Item 64 Classes/

/Classes D and E do better, but not significantly, than Class A; and on Item 65 Class D does as well as Class A.

It could be said that two items had been found which were not influenced by the factors which make for differences in test score between social groups. But the real explanation is probably a more pedestrian one. Just as very easy items do not distinguish well between subjects, because practically everyone knows the answer, so in this case too difficult items do not distinguish well between social groups, because practically no one knows the answer: the small differences which are found are due to chance.

Even when every possibility has been taken into account, however, it still remains true that Classes D and E do not omit the two items as often as they would be expected to do. The answer may lie in the fact that the items are the last multiple-choice items in the test; if the instructions are understood they can be answered very quickly without any relevant knowledge and with some chance of success/

/success. Desperately pressed for time, and with the remaining items looking hopelessly formidable, some subjects in Classes D and E might fall back on random marking of the last questions which could be treated in this way.

The two items which in general are omitted most often (although subjects in Class A omit a few other items to a greater extent) are Items 35, 66(a) and 66(b). They are reproduced below.

> In a certain secret writing
> lzqkcofu fttr yggr means-
> STARVING, NEED FOOD
>
> 35. In the same secret writing you find
> this. Write below it what it means: -
>
> yoct kgctkl rtqr.
>
> The next question is written in the
> secret writing you have already seen
> in question 35.
> Write down what it means <u>and answer it</u>.
> You can get most of the letters from the
> explanation in front of question 35, but
> there are some letters you will have to
> guess.
> 66(a) ol zgrqn Dgfrqn.......(   )
> (b)   answer (   )

It is obvious that the content of these items is responsible for the large number of omissions.

Item/

/Item 66 is, of course, the one referred to in Chapter ~~Three~~ XIII for which two marks of total score were awarded, and which it was decided to split into two parts. The second part will almost invariably not be answered correctly unless the correct solution is found to the code in the first part.

No other items involving the translation of a code are present in the test. It looks as if subjects tend to omit this type of problem automatically, perhaps less from its intrinsic difficulty than because of the length of time required to read the instructions and provide the solution. A larger number of subjects in Classes D and E actually omit the earlier and easier item.

However these findings concerning omissions are regarded, it is plain that as social level declines this type of response constitutes a steadily increasing proportion of all incorrect responses. Table XXVII brings out this point

TABLE XXVII/(cont...)

TABLE XXVII.

CONTRIBUTIONS OF OMISSIONS TO TOTAL INCORRECT RESPONSES FOR FIVE SOCIAL CLASS GROUPS

	Class A	Class B	Class C	Class D	Class E
(1) Total Incorrect Responses	4840	5680	6458	7881	8677
(2) Total Wrong Answers/Omissions	4682	5351	6036	7095	7759
(3) Total Omissions	1068	1290	1625	2296	2543
(4) (3) as % of (1)	22.1%	22.9%	25.2%	29.1%	29.3%
(5) (3) as % of (2)	22.8%	24.1%	26.9%	32.4%	32.8%

The general tendency is not perhaps so strong as in the case of "unreached responses", but it exists nevertheless, and is perfectly correlated with social class. It is shown most strongly in (5); the inclusion of "unreached responses" in the total of incorrect responses tends to mask the upward trend.

It is now possible to arrive at an estimate of the extent to which the differences between groups in mean test score are caused by the greater frequency of omissions by subjects in lower groups.

TABLE XXVIII (contd..)

TABLE XXVIII.

CONTRIBUTION OF OMISSIONS TO DIFFERENCES IN MEAN TEST SCORE BETWEEN FIVE SOCIAL CLASS GROUPS (1)

	Classes A–B	Classes B–C	Classes C–D	Classes D–E
(1) Differences in mean Test score	4.2	3.9	7.1	4
(2) Differences in total incorrect responses	840	778	1423	796
(3) Differences in total "omitted" responses	222	335	671	247
(4) (3) as % of (2)	26.4%	43.1%	47.1%	31%
(5) Parts of differences in mean test score attributable to differences in total "omitted" responses	1.1	1.7	3.3	1.2

This table is constructed in the following way. All incorrect responses and omissions are totalled separately for each group. Differences between these totals are tabulated for each comparison of groups. The differences for the "omitted" responses are then taken as percentages of the differences for all incorrect responses. The appropriate per centage is used in each comparison of groups to obtain the amount of difference in mean test score attributable to the greater/

/greater number of omissions by subjects in the lower group.

In each case where one group is compared with another a very considerable part of the difference in mean test score is attributable to the increase in omissions by the lower group, and this is particularly in evidence where the lowest middle-class group is compared with the upper-working-class group. If a table is arranged so as to include some of the extreme comparisons, the effect is even more marked.

TABLE XXIX

CONTRIBUTION OF OMISSIONS TO DIFFERENCES IN MEAN TEST SCORE BETWEEN FIVE SOCIAL CLASS GROUPS (2)

	Classes A-B	Classes A-C	Classes A-D	Classes A-E
(1) Differences in mean test score	4.2	8.1	15.2	19.2
(2) Parts of differences in mean test score attributable to differences in total "omitted" responses	1.1	2.8	6.1	7.4

The more frequent omissions of Class E are responsible for a difference in mean score between Classes A and E equivalent to nearly 10 per cent of the/

/the total score available on the test.

Tables XXX and XXXI show complete breakdowns of differences in mean test score for seven comparisons of groups.

TABLE XXX

CONTRIBUTION OF WRONG ANSWERS, OMISSIONS AND "UNREACHED RESPONSES" TO DIFFERENCES IN MEAN TEST SCORE BETWEEN FIVE SOCIAL CLASS GROUPS (1)

	Classes A-B	Classes B-C	Classes C-D	Classes D-E
(1) Differences in mean test score	4.2	3.9	7.1	4
(2) Parts of differences in mean test score attributable to differences in "wrong answer" responses	2.2	1.7	2	2.1
(3) Parts of differences in mean test score attributable to differences in "omitted" responses	1.1	1.7	3.3	1.2
(4) Parts of differences in mean test score attributable to differences in "unreached" responses	.9	.5	1.8	.7

TABLE XXXI/(contd..)

TABLE XXXI

CONTRIBUTION OF WRONG ANSWERS, OMISSIONS AND "UNREACHED RESPONSES" TO DIFFERENCES IN MEAN TEST SCORE BETWEEN FIVE SOCIAL CLASS GROUPS (2)

	Classes A-B	Classes A-C	Classes A-D	Classes A-E
(1) Differences in mean test score	4.2	8.1	15.2	19.2
(2) Parts of differences in mean test score attributable to differences in "wrong answer" responses	2.2	4	6	8
(3) Parts of differences in mean test score attributable to differences in "omitted" responses	1.1	2.8	6.1	7.4
(4) Parts of differences in mean test score attributable to differences in "unreached" responses	.9	1.3	3.1	3.8

Slight discrepancies between the two tables are caused by the rounding off of figures to the first decimal place. Their most striking feature is the regularity of the difference from group to group when wrong answers alone are considered. It is clear, for instance/

/instance, that the fact that the difference in mean test score between Classes C and D is so much larger than any other such difference between contiguous groups is due to the sharp increase in "omitted" and "unreached" responses for Class D.

Table XXXII presents the data for the three categories of incorrect response in another way. The mean error score for the 200 subjects in each group on each type of response has been worked out, and the standard deviation of the distribution has also been calculated. It can thus be said of Class A, for instance, that the mean number of "wrong answer" responses is 18.1, the mean number of omissions is 5.3, and the mean number of "unreached" responses is 0.8.

TABLE XXXII

MEAN NUMBER OF WRONG ANSWERS, OMISSIONS AND "UNREACHED RESPONSES" FOR FIVE SOCIAL CLASS GROUPS

Social Class	"Wrong Answer" Responses		"Omitted" Responses		"Unreached" Responses	
	Mean	SD	Mean	SD	Mean	SD
A	18.1	8.8	5.3	6.9	0.8	2.6
B	20.3	10.3	6.5	5.1	1.6	6.1
C	22.1	11.3	8.1	8.1	2.1	5.8
D	24	10.2	11.5	11.2	3.9	7.6
E	26.1	10.5	12.7	11.4	4.6	11

This/

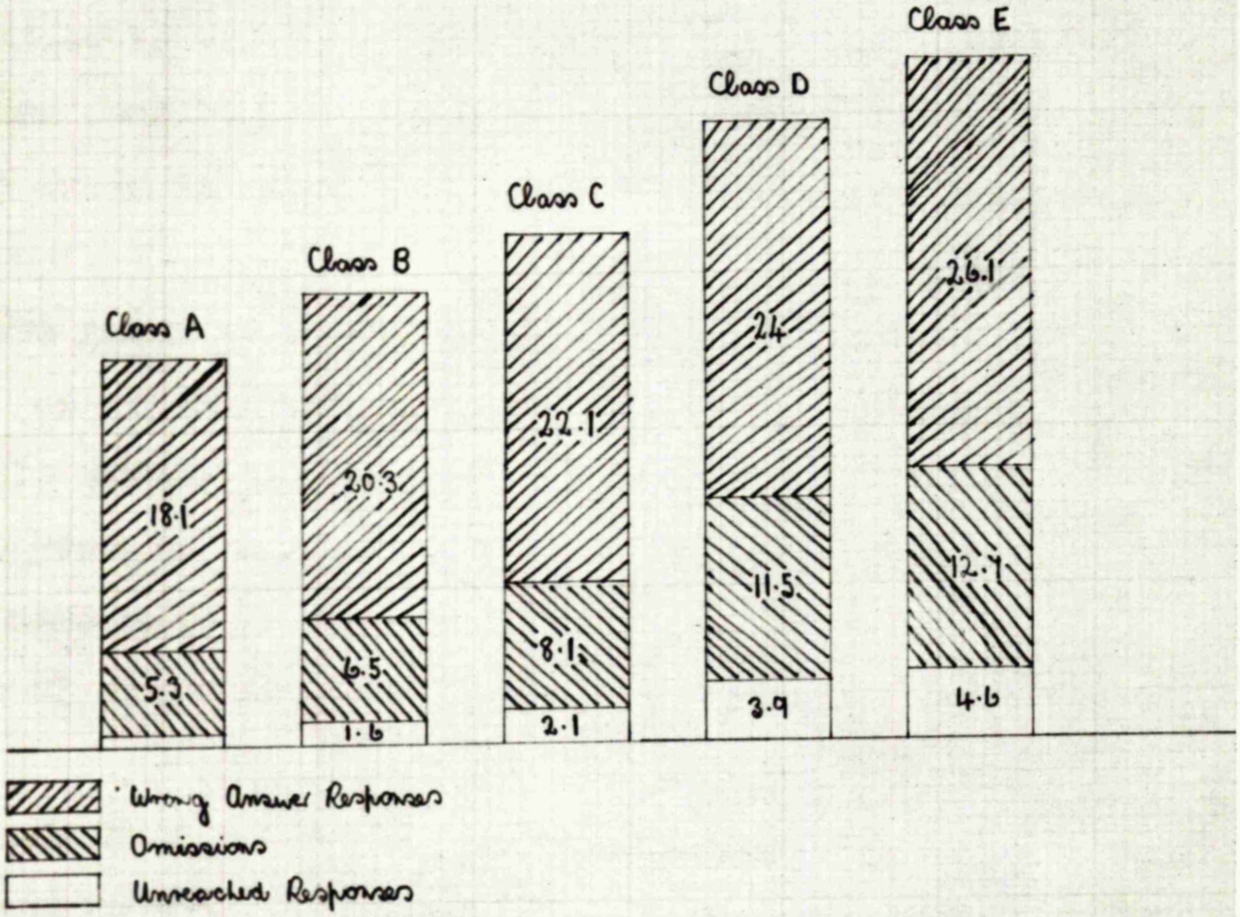

FIGURE 2. DISTRIBUTION OF INCORRECT RESPONSES FOR SOCIAL CLASS GROUPS
(Source: Data in Table XXXII)

/This table shows the same kind of relationship as has been commented on already, omissions and failure to reach items increasing steadily in importance as sources of error as the social scale is descended.

Data were originally cast in this form in the hope that an analysis of variance technique might be applied, and thus a formal statistical assessment of the significance of the findings obtained. Unfortunately no analysis of variance could be made, since it will be noted that standard deviations tend to increase in size as means become bigger.

Figure 2 was therefore drawn to bring out the differences between the groups as clearly as possible. Each bar represents the number of incorrect responses made by a "mean subject" in one of the five groups; the contribution to that number of each type of incorrect response is also shown.

The argument may be put forward that the distribution of the different types of incorrect response/

/response among the groups simply reflects their varying levels of intelligence. One way to check this possibility is to compare the performance of subjects from each of the groups who have been matched for test score. In this instance it was decided to pick one subject with a score of 52 from each of the groups and tabulate his incorrect responses. (52 was chosen as approximating to the mean score for Class A). The Class A subject had 18 wrong answers, 5 omissions and 1 "unreached response". The comparable figures for the Class B, C, D and E subjects were; 17, 5 and 2; 16,6 and 2; 14,7 and 3; and 14,8 and 2. It seems clear that even when there is no question of differences in intelligence between subjects from different groups the tendency still exists for the children in the lower groups to omit more items and to work more slowly.

The statistical evidence of the last two chapters, undoubtedly of a novel kind, poses problems in interpretation. Most psychometrists would anticipate a general increase in omissions and/

/and decline in speed consequent on the lowering of the social level in subjects; they would not anticipate that omitting and not reaching items would become proportionately more characteristic of lower-class performance.

If all kinds of error response do not increase from group to group at the same rate, then it seems that the found differences in mean test score between social groups must be exaggerated by the conditions under which testing takes place. It may be assumed that these differences would be whittled down, perhaps quite considerably in the case of a comparison between a middle-class group and a working-class group, if untimed tests were administered and if instructions were given that no items were to be omitted. Normally most psychologists would regard such a procedure as a waste of time, believing that omitting or not reaching an item is as valid an indication of failure on the item as answering it wrongly. But if item omission and test speed are to some extent socially determined, then the timing of/

/of tests and the allowance of omissions places working-class children at an unfair disadvantage over and above any presumed disadvantage due to inferior intelligence.

How are the statistical indications to be interpreted in psychological terms? It would appear that with decline in social level concentration and confidence deteriorate, and habitual speed of work lessens. It is also possible that the natural incentive to do well on a test is not seen as an incentive by working-class pupils. Alternatively, it may be that working-class children tend to be more cautious about setting down answers when they are less than absolutely certain of their accuracy.

Whatever their detailed interpretation, it may be concluded that intelligence tests show some differences between social classes which are not associated with intelligence, but which emerge from different psychological approaches towards the tests, and are dependent on a fundamental separation of experience and training.

CHAPTER XX.

SIZE AND STATISTICAL SIGNIFICANCE OF SOCIAL CLASS DIFFERENCES FOR INDIVIDUAL ITEMS.

So far attention has been devoted almost entirely to general differences in test performance between the five social classes. In this and succeeding chapters there will be concentration on analysis at a more detailed level of their differing performance on individual items.

Data regarding the number of items showing varying degrees of difference for the various comparisons of groups(ten in all) follow in Tables XXXIII - XLII. Each table is divided into two sections, as for instance XXXIII(1) and XXXIII(2). The first section shows the level of significance of the differences in item responses between the two groups being compared, and at the same time gives the actual size of differences in terms of percentages;/

/percentages; the second section shows the size of these differences in terms of the normalized index described in Chapter XVII.

Significance was estimated in each case of the comparison of two groups on an item by finding the standard error of the difference between the two percent--ages of correct responses and obtaining a Critical Ratio.

The thesis has been put forward (Eells,1948) that experimenters, when using this method of assessing differences in test performance between groups, should ignore items which more than 92 per cent or less than 8 per cent of either of the two groups gets right, because of the lessened stability of response.(The index values of these two percentages are 80 and 20 respectively). If adopted here, this method would mean the exclusion of 119 comparisons of item, or 15.7 per cent of all possible comparisons; by no means a negligible proportion. Moreover, the two categories of item which would be excluded, the very difficult and the very easy, might prove very enlightening to study. They will thus be included in the present investigation, their lesser stability being noticed wherever relevant.

TABLE XXXIII (1) /

TABLE XXXIII (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS A AND CLASS B RESPONSES.

Size of % Difference.	Significant at .01 level	Significant at .05 level	Not Significant	Total
15-19	5 items	-	-	5 items.
10-14	2 "	8 items	-	10 "
5- 9	-	1 item	23 items	24 "
0- 4	-	-	30 "	30 "
(-.5)-(-4)	-	-	7 "	7 "
TOTAL	7 items	9 items	60 items	76 items

TABLE XXXIII (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS A AND CLASS B RESPONSES.

Size of Index Difference	Number of Items.	
10-14	1	
5- 9	33	Mean Index Difference 3.8
0- 4	35	Standard Deviation 3.3
(-1)-(-4)	7	
TOTAL	76	

TABLE XXXIV (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS B AND CLASS C RESPONSES.

Size of % Difference.	Significant at .01 level	Significant at .05 level	Not Significant	Total
15-19	2 items	-	-	2 items
10-14	3 "	3 items	-	6 "
5 - 9	1 item	4 "	23 items	28 "
0- 4	-	-	31 "	31 "
(-.5)-(-4)	-	-	9 "	9 "
TOTAL	6 items	7 items	63 items	76 items

TABLE XXXIV (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS B AND CLASS C RESPONSES

Size of Index Difference	Number of Items.
10-14	2
5- 9	23
0- 4	43
(-1)-(-4)	8
TOTAL	76

Mean Index Difference 3.3

Standard Deviation 3.3

TABLE XXXV (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS C AND CLASS D RESPONSES.

Size of % Difference.	Significant at .01 level	Significant at .05 level	Not Significant	Total
20-24	2 items	-	-	2 items
15-19	10 "	-	-	10 "
10-14	12 "	12 items	-	24 "
5- 9	-	7 "	17 items	24 "
0- 4	-	-	13 "	13 "
(-.5)-(-4)	-	-	3 "	3 "
TOTAL	24 items	19 items	33 items	76 items.

TABLE XXXV (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS C AND CLASS D RESPONSES.

Size of Index Difference	Number of Items.
10-14	12
5 - 9	40
0 - 4	21
(-1)-(-4)	2
(-5)-(-9)	1
TOTAL	76

Mean Index Difference 6

Standard Deviation 4

TABLE XXXVI (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS D AND CLASS E RESPONSES.

Size of % Difference	Significant at .01 level	Significant at .05 level	Not Significant	Total
10-14	3 items	12 items	-	15 items
5- 9	-	5 "	19 items	24 "
0- 4	-	-	33 "	33 "
(-.5)-(-4)	-	-	4 "	4 "
TOTAL.	3 items	17 items	56 items	76 items.

TABLE XXXVI (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS D AND CLASS E RESPONSES.

Size of Index Difference	Number of Items.	
10-14	2	
5- 9	23	Mean Index Difference 3.5
0- 4	47	Standard Deviation 3.1
(-1)-(-4)	4	
TOTAL	76	

TABLE XXXVII (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS A AND CLASS C RESPONSES.

Size of % Difference	Significant at .01 level	Significant at .05 level	Not Significant	Total
25-29	1 item	-	-	1 item
20-24	8 items	-	-	8 items
15-19	8 "	-	-	8 "
10-14	13 "	6 items	-	19 "
5- 9	3 "	8 "	13 items	24 "
0- 4	-	-	15 "	15 "
(-.5)-(-4)	-	-	1 item	1 item
TOTAL.	33 items	14 items	29 items	76 items.

TABLE XXXVII (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS A AND CLASS C RESPONSES

Size of Index Difference	Number of Items
15-19	2
10-14	18
5- 9	39
0- 4	16
(-1)-(-4)	1
TOTAL	76

Mean Index Difference 7.3

Standard Deviation 3.9

TABLE XXXVIII (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS B AND CLASS D RESPONSES.

Size of % Difference	Significant at .01 level	Significant at .05 level	Not Significant	Total
25-29	3 items	-	-	3 items
20-24	10 "	-	-	10 items
15-19	26 "	-	-	26 "
10-14	13 "	3 items	-	16 "
5- 9	1 item	5 "	7 items	13 "
0-4	-	-	5 "	5 "
(-.5)-(-4)	-	-	3 "	3 "
TOTAL	53 items	8 items	15 items	76 items.

TABLE XXXVIII. (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS B AND CLASS D RESPONSES

Size of Index Difference	Number of Items
15-19	12
10-14	27
5- 9	27
0- 4	7
(-1)-(-4)	3
TOTAL	76

Mean Index Difference 9.5

Standard Deviation 5

TABLE XXXIX (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS C AND CLASS E RESPONSES.

Size of % Difference	Significant at .01 level	Significant at .05 level	Not Significant	Total
30-34	1 item	-	-	1 item
25-29	4 items	-	-	4 items
20-24	11 "	-	-	11 "
15-19	20 "	-	-	20 "
10-14	14 "	9 items	-	23 "
5- 9	1 item	5 "	7 items	13 "
0- 4	-	-	3 "	3 "
(-.5)-(-4)	-	-	1 item	1 item
TOTAL	51 items	14 items	11 items	76 items

TABLE XXXIV (2)/

TABLE XXXIX (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS C AND CLASS E RESPONSES.

Size of Index Difference	Number of Items	
15-19	8	
10-14	29	Mean Index Difference 9.5
5- 9	30	Standard Deviation 4.6
0- 4	8	
(-1)-(-4)	-	
(-5)-(-9)	1	
TOTAL	76	

TABLE XL (1) /

TABLE XL (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS A AND CLASS D RESPONSES.

Size of % Difference	Significant at .01 level	Significant at .05 level	Not Significant	Total
35-39	3 items	-	-	3 items
30-34	6 "	-	-	6 "
25-29	10 "	-	-	10 "
20-24	17 "	-	-	17 "
15-19	20 "	-	-	20 "
10-14	9 "	1 item	-	10 "
5- 9	-	2 items	5 items	7 "
0- 4	-	-	2 "	2 "
(-.5)-(-4)	-	-	1 item	1 item
TOTAL	65 items	3 items	8 items	76 items

TABLE XL (2) /

TABLE XL (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS A AND CLASS D RESPONSES.

Size of Index Difference	Number of Items.
25-29	1
20-24	10
15-19	14
10-14	34
5- 9	13
0- 4	3
(-1)-(-4)	-
(-5)-(-9)	1
TOTAL	76

Mean Index Difference 12.9

Standard Deviation 5.8

TABLE XLI (1) /

TABLE XLI (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS B AND CLASS E RESPONSES.

Size of % Difference.	Significant at .01 level	Significant at .05 level	Not Significant	Total
35-39	1 item	-	-	1 item
30-34	5 items	-	-	5 items
25-29	15 "	-	-	15 "
20-24	21 "	-	-	21 "
15-19	11 "	-	-	11 "
10-14	10 "	1 item	-	11 "
5- 9	1 item	3 items	6 items	10 "
0- 4	-	-	-	-
(-.5)-(-4)	-	-	2 items	2 items
TOTAL	64 items	4 items	8 items	76 items.

TABLE XLI (2) /

TABLE XLI (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS B AND CLASS E RESPONSES.

Size of Index Difference	Number of Items.
20-24	8
15-19	22
10-14	25
5- 9	15
0- 4	4
(-1)-(-4)	2
TOTAL	76

Mean Index Difference 12.6

Standard Deviation 5.8

TABLE XLII (1) /

TABLE XLII (1)

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES BETWEEN CLASS A AND CLASS E RESPONSES.

Size of % Difference	Significant at .01 level	Significant at .05 level	Not Significant	Total
45-49	1 item	-	-	1 item
40-44	3 items	-	-	3 items
35-39	5 "	-	-	5 "
30-34	16 "	-	-	16 "
25-29	17 "	-	-	17 "
20-24	15 "	-	-	15 "
15-19	6 "	-	-	6 "
10-14	6 "	1 item	-	7 "
5- 9	2 "	2 items	-	4 "
0- 4	-	-	1 item	1 item
(-.5)-(-4)	-	-	1 "	1 "
TOTAL	71 items	3 items	2 items	76 items.

TABLE XLII (2) /

TABLE XLII (2)

DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN CLASS A AND CLASS E RESPONSES.

Size of Index Difference	Number of Items
30-34	1
25-29	5
20-24	17
15-19	27
10-14	17
5- 9	7
0- 4	1
(-1)-(-4)	-
(-5)-(-9)	1
TOTAL	76

Mean Index Difference 16.4

Standard Deviation 6.4

In/

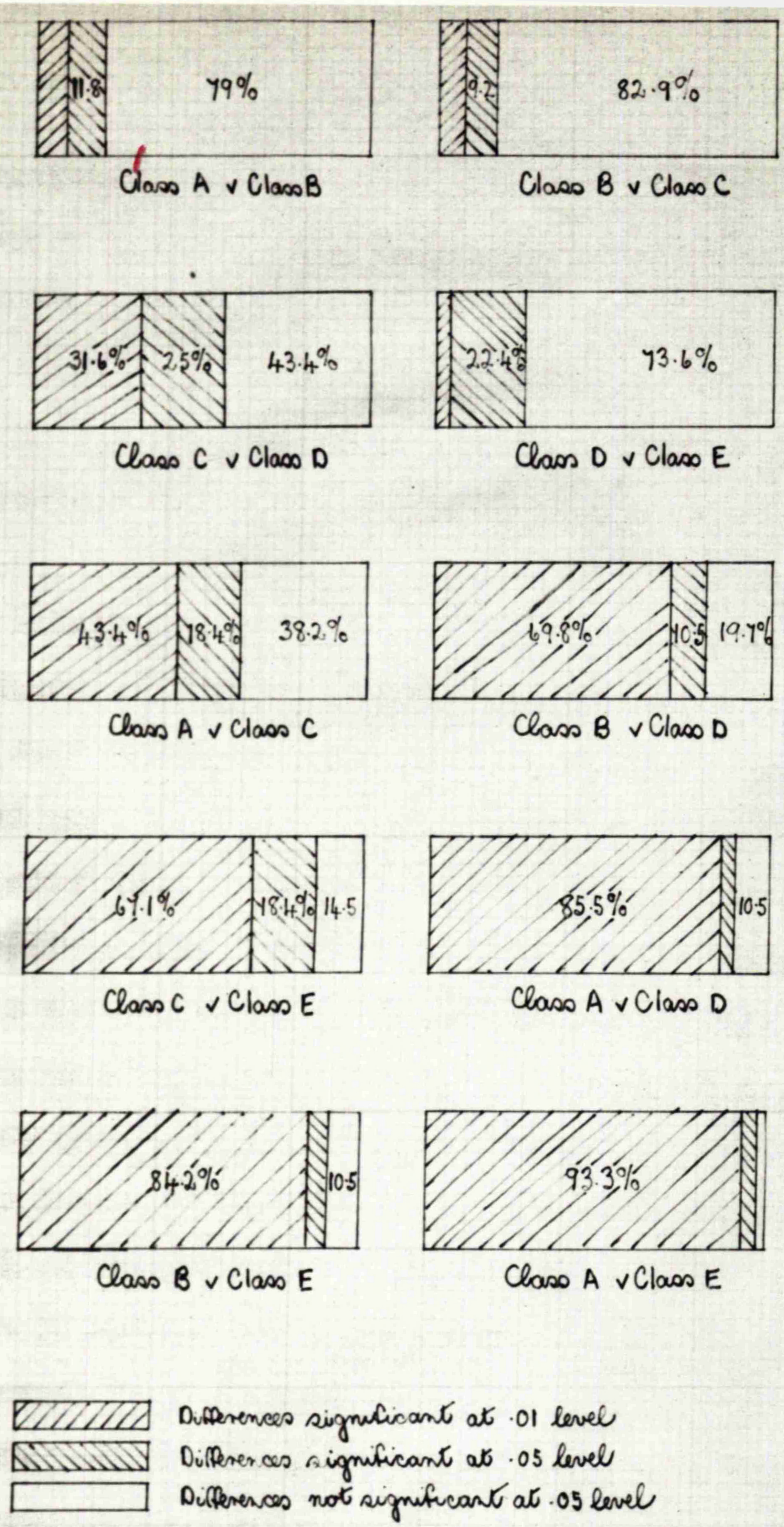

FIGURE 3. DISTRIBUTION OF TEST ITEMS BY LEVEL OF STATISTICAL SIGNIFICANCE OF DIFFERENCES BETWEEN PERFORMANCE OF SOCIAL CLASS GROUPS

/In order to render the task of describing these tables easier, when a significantly higher proportion of one of two groups provides the correct response to an item that item will be referred to as distinguishing, or discriminating between the two groups. The item will be regarded as particularly "sensitive" to whatever differences exist between the two groups.

Whatever comparison of groups is chosen for examination it is clear that items vary considerably in their capacity to distinguish ; also that the likeli--hood of an individual item distinguishing increases with the social distance between groups being compared, whether or not there is a direct causal relationship. The differing proportions of items showing significant discrimination for the various comparisons are illustrated graphically in Figure 3. It will be seen that, if the 5 per cent level of significance is accepted as a standard, only 16.8 per cent of items distinguish in favour of Class B when Classes B and C are compared; while 97.3 per cent of items distinguish in favour of Class A when Classes A and E are compared. Among the primary comparisons, i.e. when contiguous groups are compared, that of Classes/

/classes C and D, as must by now be expected, shows the highest proportion of distinguishing items. One interesting point elucidated is the large proportion of items in the comparison of Classes D and E which distinguish at the 5 per cent , but not at the 1 per cent level. This suggests a greater general similarity of response on the part of these two groups towards the test as a whole i.e. a tendency for differences to be reflected in performance on a comparatively large number of items rather than concentrated into performance on a few. Table XXXVI (2) confirms that the standard deviation of index differences in item performance between Classes D and E is lower than that for any other comparison, although the mean index difference is higher than that for the comparison of Classes B and C.

It must not be thought that because when two groups are compared numbers of items are found for which differences are not large enough to be significant it can be said that these differences are not real ones, but due to sampling errors. The small differences are too unequivocally in favour of the upper group for their existence to be attributed to chance. There is always the possibility that any single difference is due to the/

/the operation of chance factors, but the general picture presented by the mass of small differences is almost as strong evidence of the superior performance of the groups higher up the social scale as that presented by the smaller number of large differences.

It does not necessarily follow that these differences, whether large or small, are valid indications of true differences in intellectual ability between the social groups. The argument has been put forward that test constructors are guilty of bias in their selection of test items, automatically favouring types of problem with which children of the upper social classes, by virtue of background and training, are more familiar. Although there may be some truth in this charge, it is difficult to substantiate. An examination of group tests in common use certainly leaves one with the impression that psychometrists tend to conservatism in their choice of items, preferring standard problems which have apparently proved their worth; and there is clearly room for experiment with new kinds of item. Even so it would be expected that in a test containing a large number of items, such as the one in the present investigation, some items would have slipped through the mesh the performance/

/performance of subjects on which would have shown a tendency to zero difference or difference in favour of the lower social groups.

There certainly are a large number of cases, mostly represented in comparisons of Classes A and B, B and C, and D and E where differences in performance are extremely small; there are a few cases of zero difference; and occasionally an item distinguishes, though nowhere significantly, in favour of a lower group. But as has already been pointed out, the general trend of item differences is such that it would not be legitimate to take an item showing a very small difference in favour of an upper group in a comparison as showing other than a relatively low tendency to distinguish; it can not be taken as showing a tendency not to distinguish. In fact only the differences in favour of lower groups , none of which are significant, and the zero differences, constitute relevant evidence against which to examine what may be called the "item bias" theory of intelligence testing. These will be looked at more closely in later chapters. But to the present investigator/

/investigator it appears that their numbers are too small to bear out the correctness of this theory.

When differences between means and standard deviations of index differences for the various comparisons are worked out little new information emerges. The mean index difference in favour of Class C when Classes C and D are compared is signif--icantly(at the .01 level) higher than the mean index differences in favour of Classes A, B, and D when Classes A and B, B and C, and D and E are compared. Similarly both the mean index difference in favour of Class B when Classes B and D are compared and the mean index difference in favour of Class C when Classes C and E are compared are significantly higher than the difference in favour of Class A when Classes A and C are compared. No significant difference exists between the mean index difference in favour of Class A when Classes A and D are compared and that in favour of Class B when classes B and E are compared. The gap between Classes C and D to which attention has been drawn earlier is obviously responsible for the significant differences in the primary and secondary comparisons of groups. Nevertheless it perhaps requires/

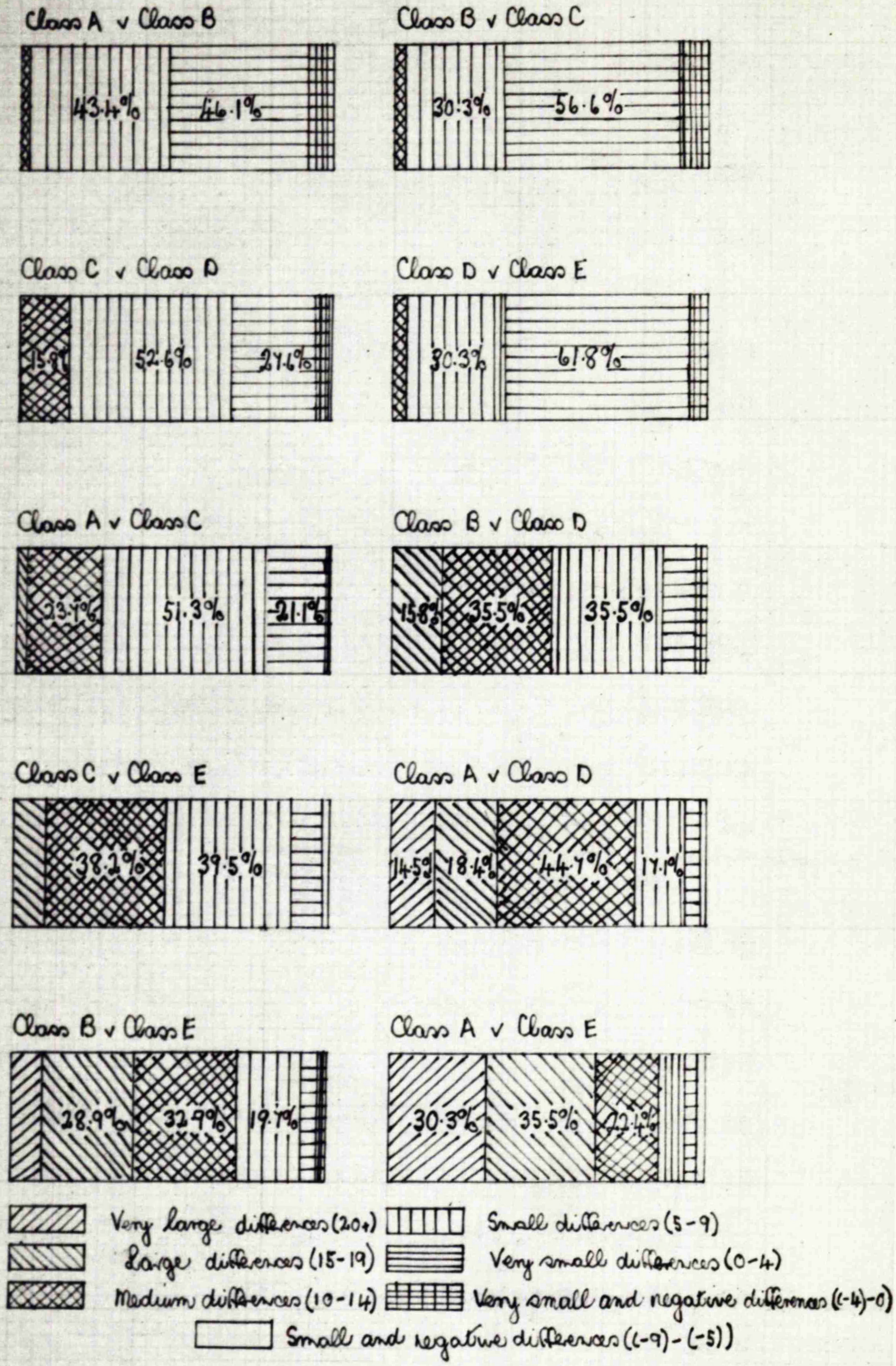

FIGURE 4. DISTRIBUTION OF TEST ITEMS BY SIZE OF INDEX DIFFERENCES BETWEEN SOCIAL CLASS GROUPS.

/requires to be pointed out that the mean index difference between Classes A and C, both middle-class groups, is greater, though not significantly so, than the mean index difference between Classes C and D.

The standard deviation of index differences between Classes C and D is significantly higher than that between Classes D and E. Also, the standard deviation of index differences between Classes B and D is significantly higher than that between Classes A and C. In neither case does significance reach the .01 level. This seems to be just a reflection of the general pattern whereby, as index differences between groups increase, so variation among items in the amount of difference shown also tends to increase.

Some indication of this range of variation can be gained from Figure 4 which shows graphically the distribution of items by size of index differences between groups for all ten comparisons. It will be noticed that these differences which have been designated as very large --- 20 index points or more -- are not found at all in any of the primary comparisons, but that they make up 30.3 per cent of all differences in/

/in the most extreme comparison, where Classes A and E are compared. Conversely differences which have been designated as very small or very small and negative make up only 2.6 per cent of all differences where Classes A and E are compared, while amounting to 67.1 per cent of all differences in comparisons both of Classes B and C and Classes D and E. This makes plain that any attempt to devise tests which would show no significant difference in performance between Classes A and E would involve a drastic remodelling of normal test material. The figures hold out little hope for the "culture-free" test. Differences in performance may be culturally determined; but even so the way to eliminate them would seem to be, not to devise a common test, but to develop a common culture.

However that may be, the considerable range of item variation is an interesting finding, and one which does not emerge from the usual kind of study of differences in test score between social groups. Clearly some items are much more responsible for these differences than others. Most of the remainder of this study will be concerned with the further analysis of this inter-item variability.

CHAPTER XXI

SPECIFIC AND GENERAL DISCRIMINATION OF ITEMS.

An interesting query which must occur to anyone who looks at the Tables in Chapter XX is, How far do items which tend to discriminate in one comparison of groups also tend to discriminate in another? This query may also be put in another way, How far is discrimination on the part of items specific and how far is it general?

The analysis of data in an attempt to answer these queries involves in the first instance the construction of a Table such as Table XLIII. There are four possible comparisons of/

/of contiguous groups, and an item may discriminate significantly in all four, in three, in two, in one or in none.

TABLE XLIII.

ITEMS SHOWING SIGNIFICANT DIFFERENCES IN RESPONSE IN COMPARISONS OF CONTIGUOUS SOCIAL CLASS GROUPS

No. of Item	Comparison A-B	Comparison B-C	Comparison C-D	Comparison D-E
3	-	-	-	.05
5	-	-	.05	-
7	-	-	.01	-
8	-	.05	-	-
9	.05	.01	-	-
10	-	-	-	.05
11	-	.05	.05	-
12	-	.01	.05	-
15	-	.01	-	.05
17	.01	-	.01	-
19	-	-	-	.01
20	-	.05	-	-
21	-	-	.05	-
23	-	-	.05	.05
24	-	-	.01	-

(contd..)

TABLE XLIII (contd..)

No. of Item	Comparison A-B	Comparison B-C	Comparison C-D	Comparison D-E
25	-	-	.01	-
26	.05	-	.05	-
28	-	-	.05	-
29	.05	-	.01	-
30	.05	-	-	.05
32	-	-	.01	.05
33	-	.05	.01	.05
34	-	-	.05	-
35	.01	-	.01	-
36	-	-	.01	-
37	-	-	.05	-
38	-	-	.05	-
39	.01	-	.05	-
40	-	-	.05	-
41	-	-	.01	-
42	-	-	.01	.05
43	-	.05	.05	-
44(a)	-	.05	-	-
(b)	-	-	.01	-
46	-	-	.01	.01

(contd..)

TABLE XLIII (contd..)

No. of Item	Comparison A-B	Comparison B-C	Comparison C-D	Comparison D-E
47	.05	-	.05	-
48	-	-	.01	-
49	-	-	-	.05
50	-	-	.01	.01
51	.01	-	.05	-
52	-	-	.01	.05
53	-	-	.01	-
54	-	-	.01	.05
55	-	.01	-	-
56(a)	-	-	.01	-
(b)	-	-	.05	.05
(c)	.01	-	-	.05
57	.05	-	.01	-
58	-	-	.05	-
59	.01	-	.01	.05
61	-	.01	-	-
62	.05	-	-	.05
63	.05	-	-	-
66(a)	-	-	.05	.05
(b)	-	-	-	.05
67	-	.05	.01	-

(contd..)

TABLE XLIII (contd..)

No. of Item	Comparison A-B	Comparison B-C	Comparison C-D	Comparison D-E
68	-	-	.01	-
69	.05	-	.01	-
70	.01	-	.05	-
71	-	.01	-	-
Totals	16 7 .01 9 .05	13 6.01 7.05	43 24.01 19.05	20 3.01 17.05

60 items are included in the table; therefore 16 items do not discriminate significantly between contiguous groups. 30 items discriminate significantly only once, 28 items twice, and two items three times. No item discriminates significantly in all four comparisons.

It is evident from this table that although the test as a whole always discriminates in favour of the upper group when an upper and lower group are being compared, in each comparison some items discriminate more than others and are responsible for a disproportionate part of the general/

/general difference between the two groups. A line cannot be arbitrarily drawn, of course, between items which differentiate at the .01 or .05 level of significance and items which do not differentiate significantly. Any broad dichotomizing of items which is undertaken must be based on the relative contrast between items which tend to exhibit large differences in performance in favour of upper groups and items which tend to exhibit small differences in performance in favour of upper groups. As stated earlier, there are a few cases where items show either zero difference between groups or small differences in favour of lower groups. But as will be deduced from Table XLIV, these almost invariably disappear in the more extreme comparisons, although there are one or two minor exceptions. In a comparison of Class A with Class E, for instance, only 2 of the 76 items do not discriminate significantly in favour of Class A.

TABLE XLIV (contd..)

TABLE XLIV

DISTRIBUTION OF TEST ITEMS BY THE LEVEL OF SIGNIFICANCE OF DIFFERENCES IN RESPONSE FOR COMPARISONS OF NON-CONTIGUOUS SOCIAL CLASS GROUPS.

Comparison	Significant at .01 level	Significant at .05 level	Not Significant
A-C	33 items	14 items	29 items
B-D	53 "	8 "	15 "
C-E	51 "	14 "	11 "
A-D	65 "	3 "	8 "
B-E	64 "	4 "	8 "
A-E	71 "	3 "	2 "

As the social distance between groups increases, not only do more items show significant differences, but the level of significance of these differences tends to become very high. Whatever factors are responsible for the appearance of zero or negative differences for some items on some comparisons are obviously scarcely strong enough to compensate for extreme social differences in groups being compared.

One question which must occur to the reader is,/

/is, How far is the tendency of items to discriminate or not to discriminate in favour of the upper group in one comparison associated with that tendency in another comparison? If the only differences between groups relevant to test performance are differences in intelligence, a close association should be expected, although, in default of detailed information as to the way in which this test was constructed, results must be interpreted with caution.

Table XLV gives an indication of how far significant discrimination in one comparison is associated with significant discrimination in another comparison.

TABLE XLV (contd..)

TABLE XLV

DISTRIBUTION OF SIGNIFICANT ITEM DISCRIMINATIONS AMONG COMPARISONS OF SOCIAL CLASS GROUPS.

	n
Items Discriminating in Comparison A-B only	1
Items discriminating in Comparison B-C only	6
Items Discriminating in Comparison C-D only	18
Items Discriminating in Comparison D-E only	5
Items Discriminating in Comparisons A-B/B-C	1
Items Discriminating in Comparisons A-B/C-D	10
Items Discriminating in Comparisons A-B/D-E	3
Items Discriminating in Comparisons B-C/C-D	4
Items Discriminating in Comparisons B-C/D-E	1
Items Discriminating in Comparisons C-D/D-E	9
Items Discriminating in Comparisons A-B/C-D/D-E	1
Items Discriminating in Comparisons B-C/C-D/D-E	1
Total	60

From/

/From this table it will be seen, as mentioned earlier, that in thirty cases an item discriminates significantly only for one comparison, in twenty--eight cases it discriminates for two comparisons, and in two cases it discriminates for three comparisons. The distribution has to be interpreted with caution: since the numbers of items discriminating in each comparison are not equal, certain associations of discriminations are more likely to occur than others. Nevertheless certain points stand out. In the first case an element of specificity of discrimination appears to be present, as warranted by the large number of items which discriminate only once. Secondly, of these specifically discriminating items, too few seem to discriminate in the comparison of Classes A and B. Thirdly, association between discrimination in the comparison of Classes A and B and discrimination in the comparison of Classes C and D is more marked, and association between discrimination in the comparison of Classes C and D and discrimination in the comparison of Classes D and E is less marked, than would be expected on the/

/the basis of a chance association of discriminations.

The objection may be made that this approach to the problem is falsified by the fact that for some comparisons the mean index differences for items between groups is small. It would be possible, for instance, for the difference in performance on an individual item between Classes B and C to be relatively large without being significant; the difference between Classes C and D might bear an exactly similar relationship to the differences for other items, but this time be significant.

To meet this objection items were divided for each comparison into two groups, according to whether their index differences were more or less than the mean index difference for that comparison. Tetrachoric correlations were then worked out between each pair of comparisons, in order to determine how closely the way an item discriminated for one comparison of groups was associated with the way it discriminated for other comparisons. The results were tabulated in Table XLVI.

TABLE XLVI (contd..)

TABLE XLVI

CORRELATIONS BETWEEN ITEM DISCRIMINATIONS FOR FOUR COMPARISONS OF SOCIAL CLASS GROUPS.

Associations of Comparisons	rt
A-B/B-C	-.49
A-B/C-D	.47
A-B/D-E	.04
B-C/C-D	-.15
B-C/D-E	.19
C-D/D-E	-.14

This table confirms, but more strikingly, general impressions gathered from Table XLV. None of the smaller correlation coefficients are significant ; but correlation coefficients both for A-B/B-C and A-B/C-D are significant at the .01 level.

It is evident that items which tend to discriminate well between Classes A and B tend to discriminate poorly between Classes B and C, and vice versa. The same items tend to discriminate well again between Classes C and D. For the other assoc--iations of comparisons there is no marked tendency either/

/either way, in fact no sound evidence for believing in any correlation at all, although it is worth noticing that two of the coefficients are negative.

In general terms the interpretation of this evidence is clear enough, although perhaps a little surprising. On the whole, it appears that whether an item discriminates between two social groups to a large extent or a small extent depends, at least when the groups are not socially too far apart, on the nature of the groups. There is no general tendency for items to discriminate consistently well or poorly; if anything, inconsistency is the rule. If findings are classified in social terms, they run as follows. Items which discriminate well or poorly between an upper-middle-class group and a middle-middle-class group similarly tend to discriminate well or poorly between a lower-middle--class group and an upper-working-class group, but the tendency is only moderate. On the other hand, items which distinguish well between a middle-middle--class group and a lower-middle-class group tend to distinguish poorly between an upper-middle-class group/

/group and a middle-middle class group, and items which distinguish poorly in one case tend to distinguish well in the other. There may be a slight tendency for items which distinguish well or poorly between a lower-middle-class group and an upper-working-class group to behave in the opposite way when an upper-working-class group and a lower-working-class group or a middle-middle--class group and a lower-middle-class group are compared. A slight, but perhaps not meaningful, positive association is found between the way in which items distinguish when a middle-middle-class group and a lower-middle-class group and an upper-working-class group and a lower-working-class group are compared. It is unlikely that there is any association between good or poor discrimination of items in a comparison of an upper-middle-class group and a middle-middle-class group and similar good or poor discrimination in a comparison of an upper-working-class group and a lower-working-class group.

As/