THE SIGNIFICANCE OF METHOD IN INTELLECTUAL TASKS:
A STUDY BASED UPON AN ANALYSIS OF
INTELLIGENCE TEST ERRORS

A THESIS
PRESENTED TO
THE UNIVERSITY OF NEW ZEALAND

IN FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE
DOCTOR OF PHILOSOPHY

BY
PHILIP JOHN LAWRENCE
1955
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>The History of a Problem</td>
<td>1</td>
</tr>
</tbody>
</table>

**SECTION A: THEORETICAL FORMULATION.**

| I. THE NATURE OF INTELLIGENCE | 7 |
| Introduction | 7 |
| Spearman and General Intelligence | 10 |
| Modifications and Extensions of Spearman's Original Factorial Studies | 14 |
| Thurstone | 14 |
| Further Factorial Studies of Intelligence | 17 |
| Factorial Studies of Reasoning | 20 |
| Closure Factors | 27 |
| Summary | 33 |

Theoretical and Experimental Studies of Reasoning and Problem Solving | 36 |

Gestalt Studies | 37 |
Learning and Problem Solving | 44 |

Cognitive Theories | 53 |

Summary | 59 |

**II. STUDIES OF COGNITIVE ERROR.** | 62 |

Systematic Studies of Error | 63 |

Spearman | 63 |
Selz | 66 |

Relevant Experiments | 69 |

The Gestalt Account of Error | 74 |
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification of Errors found in the Literature on Problem Solving.</td>
<td>78</td>
</tr>
<tr>
<td>Inadequate Formulation of the Problem.</td>
<td>80</td>
</tr>
<tr>
<td>Distortion or Partial Operation of the Task.</td>
<td>81</td>
</tr>
<tr>
<td>Dislocation of the Elements in a Problem.</td>
<td>85</td>
</tr>
<tr>
<td>Influence of &quot;Set&quot;</td>
<td>90</td>
</tr>
<tr>
<td>Influence of Attitude, Emotional Involvement and Personal Preference.</td>
<td>92</td>
</tr>
<tr>
<td>Some Specific Errors in Intelligence Test Items.</td>
<td>95</td>
</tr>
<tr>
<td>Factorial Studies of Carefulness.</td>
<td>99</td>
</tr>
<tr>
<td>Conclusions.</td>
<td>100</td>
</tr>
</tbody>
</table>

**SECTION B: EXPERIMENTAL DESIGN AND ANALYSES BASED ON ERROR SCORES.**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>III. THE NATURE OF THE INVESTIGATION.</td>
<td>102</td>
</tr>
<tr>
<td>Selection of Test.</td>
<td>103</td>
</tr>
<tr>
<td>Selection of Subjects.</td>
<td>105</td>
</tr>
<tr>
<td>Preliminary Study.</td>
<td>110</td>
</tr>
<tr>
<td>Experimental Design.</td>
<td>124</td>
</tr>
<tr>
<td>Methods of Analysis.</td>
<td>128</td>
</tr>
<tr>
<td>IV. REPETITION OF ERRORS AND DIFFICULTY OF CATEGORIES</td>
<td>132</td>
</tr>
<tr>
<td>Repetition of Errors on Otis I and II.</td>
<td>132</td>
</tr>
<tr>
<td>Consistency of Errors in, and Difficulty of Categories</td>
<td>134</td>
</tr>
<tr>
<td>V. A COMPARISON OF 'HIGH ERROR' AND 'LOW ERROR' GROUPS OVER ALL CATEGORIES</td>
<td>142</td>
</tr>
<tr>
<td>Method of Selecting Groups</td>
<td>142</td>
</tr>
<tr>
<td>Results and Discussion for Each Category</td>
<td>145</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Category (1)</td>
<td>146</td>
</tr>
<tr>
<td>(2)</td>
<td>164</td>
</tr>
<tr>
<td>(3)</td>
<td>166</td>
</tr>
<tr>
<td>(4)</td>
<td>171</td>
</tr>
<tr>
<td>(5)</td>
<td>173</td>
</tr>
<tr>
<td>(6)</td>
<td>179</td>
</tr>
<tr>
<td>(7)</td>
<td>182</td>
</tr>
<tr>
<td>(8)</td>
<td>187</td>
</tr>
<tr>
<td>(9)</td>
<td>192</td>
</tr>
<tr>
<td>(10)</td>
<td>194</td>
</tr>
<tr>
<td>(11)</td>
<td>198</td>
</tr>
</tbody>
</table>

**Sex Differences Over All Categories.** 200

**Summary of the Analysis of H.E. and L.E. Groups.** 203

**VI. INCONSISTENCY OF ERROR RATIOS, AND 'CAREFULNESS-CARELESSNESS' RATINGS IN RELATION TO ERROR.** 206

- Inconsistency of Error Ratios. 206
- Ratings on Carefulness-Carelessness in Relation to Error. 214

**VII. DISTRIBUTIONS OF WRONG RESPONSES FOR ALL CASES.** 222

- Distractor Position. 222
- Concentration of Wrong Responses 225
- Sex Differences. 228
- Comparison of Otis I and Otis II 229

**SECTION C: ANALYSES BASED ON INTERVIEWS WITH CHILDREN AND SUMMARY AND CONCLUSIONS.**

**VIII. INDIVIDUAL ANALYSES OF ERRORS: SELECTION OF GROUP AND PROCEDURE.** 231

- Selection of Subjects 231
- Selection of Items 232
- Method of Re-working 233
Conditions and Rapport ................................................. 234
Method of Presenting the Results ................................. 236
IX. INDIVIDUAL ANALYSES OF ERRORS: RESULTS AND DISCUSSION .... 237
Classifications of Errors by Categories. ......................... 237

Category (1) ............................................................ 237
(2) .................................................................. 240
(3) .................................................................. 241
(4) .................................................................. 242
(5) .................................................................. 244
(6) .................................................................. 246
(7) .................................................................. 247
(8) .................................................................. 250
(9) .................................................................. 251
(10) ................................................................ 252
(11) ................................................................ 253

Discussion ..................................................................... 255

Spearman, Selz and Relevant Experiments .................... 257
The Gestalt Account of Error ........................................ 259
Errors found in the Literature on Problem Solving .......... 260
Specific Errors in Intelligence Test Items ....................... 264
Carefulness ................................................................. 264

X. SUMMARY AND CONCLUSIONS ........................................ 266

Summary ................................................................. 266
Section A .................................................................. 266
Section B .................................................................. 267
Section C .................................................................. 271

Conclusions ............................................................... 272

Intelligence Test Construction ....................................... 272
Classroom Teaching ..................................................... 275

BIBLIOGRAPHY ............................................................ 279

APPENDIX A. Mean Error Ratio Scores by Category, Sex, and I.Q. Group, for Otis I .... 290
| APPENDIX B | Example of Record Sheet for Marking Errors Etc. | 291 |
| APPENDIX C | I. Mean Error Ratio Scores by Category, Sex, and I.Q. Group for Otis II | 292 |
| APPENDIX D | II. No. of Cases in 7 I.Q. Groups for Otis II | 292 |
| APPENDIX D | Adjustment of Error Ratio Scores | 293 |
| APPENDIX E | M.Vs for Otis I and II Error Ratio Scores | 299 |
| APPENDIX F | Instructions and Record Sheet for Ratings on Carefulness-Carelessness | 300 |
| APPENDIX G | Comparison of Scores on Otis I and II, F.M.A. and Intermediate D | 303 |
| APPENDIX H | Classification and Illustration of Errors on all Items, by Categories | 305 |
| APPENDIX I | A Summary of Specific Weaknesses of Otis Items, with some Suggestions for Improvement | 399 |
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Distribution of Cases in 7 I.Q. Groups</td>
<td>114</td>
</tr>
<tr>
<td>II.</td>
<td>Median and Range of I.Q. Groups on Four Measures</td>
<td>119</td>
</tr>
<tr>
<td>III.</td>
<td>Percentage of Item Errors and Omissions Repeated (A), and Percentage of Identical Errors (B)</td>
<td>133</td>
</tr>
<tr>
<td>IV.</td>
<td>Comparison of Error Ratio Means over 11 Categories</td>
<td>135</td>
</tr>
<tr>
<td>V.</td>
<td>Omissions on all Categories (Otis I)</td>
<td>139</td>
</tr>
<tr>
<td>VI.</td>
<td>Unadjusted Error Ratio Score for Category (1) H.E. and L.E. Groups</td>
<td>147</td>
</tr>
<tr>
<td>VII.</td>
<td>I.Q. (Otis I and II Combined) of Category (1) H.E. and L.E. Groups</td>
<td>148</td>
</tr>
<tr>
<td>VIII.</td>
<td>W/A Scores (In Percentages) of Category (1) H.E. and L.E. Groups</td>
<td>148</td>
</tr>
<tr>
<td>IX.</td>
<td>Percentages of Boys and Girls in Category (1) H.E. and L.E. Groups</td>
<td>150</td>
</tr>
<tr>
<td>X.</td>
<td>H.E. and L.E. Groups on Category (1) compared on Other Categories</td>
<td>152</td>
</tr>
<tr>
<td>XI.</td>
<td>Category (1) H.E. and L.E. Groups Compared on P.M.A. Q and V</td>
<td>155</td>
</tr>
<tr>
<td>XII.</td>
<td>Category (1) H.E. and L.E. Groups Compared on Intermediate D Scores.</td>
<td>157</td>
</tr>
<tr>
<td>XIII.</td>
<td>Distribution of Specific Item Errors for H.E. L.E. and N (Neutral) Groups on Category (1) Items</td>
<td>160</td>
</tr>
<tr>
<td>XIV.</td>
<td>Comparison of Category (2) H.E. and L.E. Groups</td>
<td>164</td>
</tr>
<tr>
<td>XV.</td>
<td>Distribution of Specific Item Errors for Category (2)</td>
<td>165</td>
</tr>
<tr>
<td>XVI.</td>
<td>Comparison of Category (3) H.E. and L.E. Groups</td>
<td>166</td>
</tr>
</tbody>
</table>
TABLE

XVII. Distribution of Specific Item Errors for Category (3) .................. 169

XVIII. Comparison of Category (4) H.E. and L.E. Groups 171

XIX. Distribution of Specific Item Errors for Category (4) .................. 172

XX. Comparison of Category (5) H.E. and L.E. Groups 173

XXI. Comparison of Category (5) H.E. and L.E. Groups on Omissions in Intermediate D. 174

XXII. Comparison of Category (5) H.E. and L.E. Groups on P.M.A. Q and S 176

XXIII. Distribution of Specific Item Errors for Category (5) .................. 176

XXIV. Comparison of Category (6) H.E. and L.E. Groups 179

XXV. Distribution of Specific Item Errors for Category (6) .................. 181

XXVI. Comparison of Category (6) H.E. and L.E. Groups 182

XXVII. Comparison of Category (7) H.E. and L.E. Groups on Intermediate D Scores 184

XXVIII. Distribution of Specific Item Errors for Category (7) .................. 185

XXIX. Comparison of Category (8) H.E. and L.E. Groups 188

XXX. Comparison of Category (8) H.E. and L.E. Groups on P.M.A. and Intermediate D Scores 189

XXXI. Distribution of Specific Item Errors for Category (8) .................. 189

XXXII. Comparison of Category (9) H.E. and L.E. Groups 192

XXXIII. Comparison of Category (9) H.E. and L.E. Groups on Intermediate D Scores 193

XXXIV. Distribution of Specific Item Errors for Category (9) .................. 194

XXXV. Comparison of Category (10) H.E. and L.E. Groups 195
<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXXVI. Comparison of Category (10) H.E. and L.E. Groups on Intermediate D Scores.</td>
<td>196</td>
</tr>
<tr>
<td>XXXVII. Distribution of Specific Item Errors for Category (10)</td>
<td>197</td>
</tr>
<tr>
<td>XXXVIII. Comparison of Category (11) H.E. and L.E. Groups</td>
<td>198</td>
</tr>
<tr>
<td>XXXIX. Distribution of Specific Item Errors for Category (11)</td>
<td>199</td>
</tr>
<tr>
<td>XL. Percentage of Boys and Girls in each Category Group</td>
<td>200</td>
</tr>
<tr>
<td>XLI. Mean Differences between Mean Error Ratio For Boys and Girls</td>
<td>202</td>
</tr>
<tr>
<td>XLII. Teacher Ratings on a &quot;Carefulness-Carelessness&quot; Scale</td>
<td>216</td>
</tr>
<tr>
<td>XLIII. Comparison of &quot;Careful&quot; V &quot;Careless&quot; Group on Selected Measures</td>
<td>219</td>
</tr>
<tr>
<td>XLIV. Mean Number of Errors and Percentages over Five Distractor Positions</td>
<td>223</td>
</tr>
<tr>
<td>XLV. Percentage of Errors in Distractor Positions for Three I.Q. Groups</td>
<td>225</td>
</tr>
<tr>
<td>XLVI. Distributions of Percentages of Errors on 47 Multiple Choice Items</td>
<td>226</td>
</tr>
<tr>
<td>XLVII. Distributions of Percentages of Errors on 24 &quot;Open Answer&quot; Items and 4 Multiple Choice Items</td>
<td>227</td>
</tr>
<tr>
<td>XLVIII. Percentage of Total Errors on each Category for Boys and Girls</td>
<td>229</td>
</tr>
</tbody>
</table>

**LIST OF FIGURES**

Figure
1 Distribution of Percentages on Four Measures 120
2 Comparison of Error Ratio Means of Otis I and Otis II 135
3 Comparison of "Actual" and "Predicted" Difficulty of Mean Error Ratios 137
INTRODUCTION

THE HISTORY OF A PROBLEM

Some years ago, the writer completed a thesis, in the course of which an attempt was made to clarify some of the psychological problems involved in the concept of "work", and to relate these to educational techniques.¹ This led to a consideration of the relationship between "output" and "method", and an appraisal of some of the outstanding studies in industrial psychology² in so far as they were relevant to educational problems of the same general type. One of the problems centered around the study of work methods - e.g.- study habits and skills, and some of the implications of these methods of improving performance, for example the relation of performance to capacity in intellectual tasks. Since that time, the problem has undergone several stages of transformation and has finally materialised in the form of this thesis. A brief account of these transformations will put the present thesis topic in perspective, and account for the particular way in which it has been handled.


²e.g.- The Western Electric Studies - See Roethlisberger F.J. and Dickson W.J. Management and the Worker, 1939. (Harvard University Press).
STAGE 1.

Work methods must be learned, so that the question arose as to how and when children develop their particular methods, and to what extent they are conscious of them, or can be made more conscious of them by direct teaching. Clearly, technique varies greatly in different subjects from specialised methods in handling Latin 'unseens' to methods involved in solving logical problems. Thus, in order to ensure maximum transfer from any direct teaching of methods of tackling intellectual tasks of many varieties, it would be necessary to understand the role of method in cognitive functioning at a fundamental level.

STAGE 2.

This line of argument led on to a consideration of the methods involved in such processes as reasoning, or problem solving, or even more broadly — thinking. It was found that although there is a considerable literature on these topics, it is very difficult to find a consensus of theory and experiment which gives a reasonably clear picture of the whole field. Furthermore, the very terminology was found misleading, for example, many experiments in concept formation could be quite readily classified as experiments in problem solving, and vice versa. This in itself was not an important barrier, but the difficulty arose because
experiments studied were so often based on the assumption that the terminological differences marked off quite distinct fields of investigation.

**STAGE 3.**

While Stage 2 was in progress, Stage 3 began to take shape - namely the relationship of methods of work to the expression of intelligence. Could systematic training influence intelligence test performance? As a first move towards answering this question, two steps were taken, firstly an attempt to build up some sort of systematic training course, and secondly a survey of the literature on the effects of training and practice on tests of intelligence and reasoning.

The first step led to an attempt to construct exercises - e.g. increasingly complex examples of the logic of classes, using blocks and beads of different colours; the building up of complex sets of rules for a simple classroom game; the constructing of statements with varying degrees of evidence on which to judge probability. These exercises were not very original, but a more serious defect lay in the fact that they were not systematic, nor could they be, without a systematic theory of intelligence upon which to base them, thus a prior requirement was the formulation of some such theory.
The second step revealed that there was a considerable number of experiments on practice effects in intelligence tests, although fewer on direct training, and a number of training experiments of different types based on reasoning or problem solving situations. In none of these experiments was there an attempt to work out the implications of the results in terms of a systematic theory of intelligence, so that both this and the preceding step pointed to the necessity for studying the nature of intelligence in more detail, hence the next stage.

STAGE 4.

Although the two great N.S.S.E. Year Books on intelligence hinged around the nature-nurture "controversy", they did not give a systematic account of the types of theory assumed by the "protagonists". At this stage in the development of the problem, it was necessary to attempt to find a rationale for relating performance and capacity in intellectual tasks at a level which would be more fundamental, and in a way which would allow for a systematic treatment of method as an integral part of intelligence.

A survey of the literature available made it apparent that while there had been many developments, for example in

---

the factorial analysis of intelligence tests, there had been few attempts to clarify the implications of such studies and to relate intelligence to the broader cognitive domain. It was decided that the next step should be an attempt to provide a basis for interpreting the literature on intelligence and problem solving in such a way that (a) the study of cognitive processes in general should be made relevant to the study of intelligence, and (b) the relationship between performance and capacity in intellectual tasks should be clarified. This developed into a major problem in itself, and although naturally enough, not resolved, at least a tentative 'tidying up' of groups of studies was attempted and some of the significant outlines of a theory of intelligence blocked in. This stage resulted in the study on the nature of intelligence presented in chapter one of this thesis.

**STAGE 5.**

Although the previous stage brought into focus certain broad implications, it was still necessary to clarify the 'methodological' component of intelligence and to reduce it to something which could be tested against experimental findings. This step was taken by applying a widely used educational procedure in dealing with attempts to improve performance - i.e. first diagnose those areas where improve-
ment can take place, or in other words, find the weaknesses. It seemed reasonable then, to begin this clarification by concentrating on weaknesses, or more specifically - errors, i.e. to evaluate theories of error in intelligence and problem solving and the experimental evidence relating to these types of error, and to see whether the results of such an investigation fitted in with the broad theory of intelligence outlined. The results of this step have been presented in chapter two.

STAGE 6.

The last stage involved the reduction of this general method to a specific experimental design which could be used to study certain aspects of the problem more intensively, and this stage has been presented in the remaining chapters of the thesis.
SECTION A.

THEORETICAL FORMULATION.
CHAPTER I

THE NATURE OF INTELLIGENCE

I. - INTRODUCTION.

In a paper on "The Transmission of Knowledge", G. Patrick Meredith\(^1\) wrote -

"The kind of theory which bears fruit often originates in a tidying up process applied to existing data, whereby a clearer view of the inter-relationships of those data is obtained. The theory then suggests that hitherto unsuspected relations hold between the data, and it proceeds to predict further data satisfying these relations."

From time to time there has been a "tidying up" of studies dealing with certain aspects of intelligence, e.g. the recent symposium on the effects of coaching and practice in intelligence tests\(^2\) or the article by Burt\(^3\) on the differentiation of intellectual ability. The detailed


reviews of Wolfle⁴, Burt⁵, Vernon⁶ and French⁷ have brought some organization into the ever extending field of factorial studies of human ability, thus bringing intelligence into relationship with other aspects of cognitive functioning.

There is, however, a danger in misinterpreting factorial classifications due to the asking of the wrong question i.e. such classifications are an answer to the question - "How many factors are necessary to account for the given correlations, how can these factors be described, and how are they related?" This is not necessarily the same question as - "What is the nature of the psychological processes which operate in the answering of a given test item?", for the description of factors is usually given in terms drawn from the answers to the second question. Ryle⁸ has summed up the position by contrasting the "products" of

⁴Wolfle D., Factor Analysis to 1940, 1940, Psychomet. Mon. No. 3.
pondering with the "acts" of pondering, thus such terms as deduction or abstraction belong to the former rather than the latter.

On the other hand a **definition** of intelligence in terms of the psychological processes involved may be even more confusing as it is likely to depend on a **priori** judgments based upon introspective data or on analyses of logical forms of thought. Furthermore, definitions, as Spiker and McCandless⁹ have pointed out, may be **formally** correct, while "which of the several concepts of intelligence proves to be the most useful, in the sense of entering into laws which lead ultimately to more accurate predictions of human behaviour, remains to be seen." It is the purpose of this chapter to attempt a "tidying up" of several groups of studies in order to throw some light on inter-relationships among data which may point to ways in which traditional definitions of intelligence may enter into wider laws of cognitive functioning.

II.- SPEARMAN AND GENERAL INTELLIGENCE.

As Hearnshaw\textsuperscript{10} has noted, there has been a great advance in the elaboration of statistical methodology, but little advance towards a coherent theory of intellect. This is illustrated by the fact that while Spearman's methodological contributions have been elaborated, refined and changed, his analysis of the nature of intelligence has been accepted, often implicitly rather than explicitly, as a satisfactory definition of 'g'. In his original analysis he suggested that the three principles of noegenesis represented the qualitative aspect of intelligence, although the first principle (apprehension of experience) "has the least obvious claims to the name"\textsuperscript{11}. Although in later works\textsuperscript{12} he did not commit himself consistently on the exact relation of 'g' and intelligence, in his last paper\textsuperscript{13} he added 'abstraction' to noegenesis and claimed that these describe the psychological nature of 'g'.


\textsuperscript{12}e.g.- \textit{The Abilities of Man}, 1927.(London:Macmillan)

Although Vernon\textsuperscript{14} was careful to point out that we use 'g' to refer to the objectively established general factor, instead of the subjective and indefinable term intelligence", it has been usual to identify 'g' with 'general intelligence' and to describe general intelligence as "relational thinking". It is an inescapable fact that most of the situations in which intelligence is measured involve the manipulation of relations and things related (fundaments). To have conceded this, however, is to have made little more than a basic statement in formal logic, and in fact Spearman's analysis begins with the self evidence of the education of relations and the educating of correlates - e.g. "The mentally presenting of any two or more characters tends to evoke immediately a knowing of relation between them."\textsuperscript{15}

From a distinctly psychological viewpoint this analysis does not go far enough, for example, at least two of the crucial questions are:- (1) What conditions inhibit or facilitate the knowing of relations? (2) How does this knowing of relations develop from the very simple to the most complex situations? To answer these questions by re-


\textsuperscript{15}Spearman, Op. Cit. 1923. p63. italics not in original.
ferring to different amounts of 'g' i.e. in its quantitative aspect of "mental energy"\(^{16}\) is merely to make the problems more inaccessible. Of the other four quantitative principles (retentivity, fatigue, conative control and primordial potencies\(^ {17}\)) the last three refer to conditions which influence the free flow of mental energy, while the first is of a somewhat different nature. Retentivity is broken up into two manifestations - inertia, examples of this being perseveration, set and pathological complexes; and facilitation, - "the real key to at least two basal problems. The one, vital for theory is that of psychic or psycho-physiological "dispositions", "traces" or engrams. The other... is that of "formal discipline". Secondary to these... are the questions as to how facilitation accumulates ...how it decreases... and how it is modified...\(^ {18}\)

Although Spearman developed his concepts of inertia and facilitation as minor aspects of cognition, these are concepts which can be linked up with broader fields of psychological study, for example, the study of processes of reasoning and problem solving and the whole field of learning - these studies being concerned more with process

\(^{16}\)Ibid p131

\(^{17}\)Ibid p132 ff

\(^{18}\)Ibid p133
than with the classification of products. This suggests that the nature of intelligence might be more clearly delineated if the results of factorial analyses were brought into closer relationship with the results of experimental studies of problem solving and reasoning. Hearnshaw\(^{19}\) gave as his tentative definition of intelligence - "a cluster of high grade skills concerned with problem solving," while Vernon\(^{20}\) summed up a widely accepted viewpoint when he wrote - "reasoning ability is one of the commoner definitions of intelligence".

There are no clear conventional distinctions between the concepts intelligence, reasoning and problem solving, and it is likely that investigations in these latter fields may have considerable, sometimes complete, overlap with studies of intelligence. In order to explore this possibility, three groups of studies have been brought together i.e.- (1) Modifications or extensions of Spearman's original factorial studies. (2) Theoretical and experimental studies of reasoning and problem solving. (3) Cognitive theories which take account of the development of intelligence.

\(^{19}\)Hearnshaw, Op. Cit. p315.

The original two factor theory has been superseded by several alternative designs and it is not necessary to review these in detail here, nor to draw attention to the typical hierarchical structure of most British analyses in which 'g', in the objective sense in which Vernon described it (see above), accounts for most of the variance of cognitive tests. There are, however, a few developments which may throw some light on the nature of intelligence, the first of these arising from Thurstone's investigations of the primary mental abilities.

(a) Thurstone.-

In his first primary mental abilities study, Thurstone found inductive (I), reasoning (R), and deductive (D) factors, the last two being only tentatively identified. Wolfle in reviewing related studies pointed out the instability of these factors in other analyses, including

\[21\] Thurstone L.L., Primary Mental Abilities, 1938, Psychomet. Mon. No. 1.

that of Holzinger and Swineford\textsuperscript{23} in which tests of the deduction, reasoning and problem solving variety all disappeared into the general factor. He concluded - "It is an interesting and by no means implausible hypothesis that Spearman's 'g' and the reasoning factor are identical".\textsuperscript{24} The R and D factors did not appear in the Thurstones' later study\textsuperscript{25} although I was firmly identified, and a general factor appeared in the second order analysis, of which they wrote - "If further PMAs of children should reveal this general factor, it will sustain Spearman's contention that there exists a general intellective factor"\textsuperscript{26}. After repeating selected tests on 8th grade children, they found that "the inductive or reasoning factor has the highest correlation with the second order general factor"\textsuperscript{27}. Commenting on this general factor, they wrote: - "The psychological interpretation of the general factor must be only tentative at the present time"\textsuperscript{28}.

\textsuperscript{24}Wolfle, Op. Cit. p33
\textsuperscript{25}Thurstone L.L. and Thurstone G., Factorial Studies of Intelligence, 1941, Psychomet. Mon. No. 2.
\textsuperscript{26}Ibid. p26.
\textsuperscript{27}Ibid. p38.
\textsuperscript{28}Loc. Cit.
The later study of perception yielded two factors - A, speed and strength of closure (i.e. the ability to form a perceptual closure against some distraction out of an unorganised presentation), and E. flexibility of closure (i.e. flexibility in manipulating several conflicting gestalts). These two factors were closely related, and this is important because the reasoning tests of the primary abilities had significant loadings on one of them - E. Thurstone commented - "...one might wonder whether this factor represents one important aspect of intelligence". In the same study, the composite tests of the PMAs showed a common factor described as "...a second order general factor common to the primary abilities, ...it may be interpreted as Spearman's 'g'". The relation of the closure factors to intelligence was again specifically mentioned by Thurstone in a later paper:

"The question has been raised about the primary factors that have been identified, whether any of them represent a modern form of the central intellective factor that

29Thurstone L.L., A Factorial Study of Perception, 1944, Psychomet. Mon. No. 4.
30Ibid. p111.
31Ibid. p114
Spearman postulated 40 years ago. If we were to make a guess in answer to this question we should probably consider the inductive factor and the two closure factors. But the solution to the problem of the central intellective factor is probably more indirect!

These statements of Thurstone's have been given in some detail because they indicate firstly, that Thurstone was not definite about the relationship between intelligence, Spearman's 'g' and second order factors; secondly, the importance of distinguishing between Spearman's qualitative and quantitative principles, as it is not always clear whether 'g' is meant to refer to a central intellective factor or to intellective processes as well; thirdly, that the description of intelligence in terms of one reasoning factor may be inadequate. There is also an interesting parallel between the "closure" factors and Spearman's "inertia" and "facilitation".

(b) **Further Factorial Studies of Intelligence.**

Although Spearman rejected analysis and synthesis as genuine cognitive processes, the terms have been used to describe factors related to intelligence and reasoning,

---

e.g. - Rimoldi\textsuperscript{34} in a second order factor analysis of his 19 non verbal and performance tests of intelligence, found that of the two clearest factors one represented a sort of analytical activity - "it is related to what has been called abstraction and indicates a rather fundamental dimension of psychological dynamics such as capacity to abstract from complex presentations," while the other represented a certain synthetic or syncretic activity. This division is similar to the contrast between explicit and implicit apprehension of relations made by Burt\textsuperscript{35}. In a later study in which tests were selected which were claimed by their authors to be good measures of g, I, R and D, Rimoldi\textsuperscript{36} found six factors: - (1) reasoning - not only analysis or synthesis, but their dynamic interplay, (2) eduction of relations, (3) educing correlates, (4) bringing parts together into a meaningful solution, (5) likeness-seeing differences and similarities, (6) verbal.


\textsuperscript{35}Burt, Op. Cit., 1949, p191

This analysis is of interest because the factors were interpreted in terms of Spearman's noegenetic principles, but supplemented by processes (especially (1) and (4) above) which suggest an ability to manipulate noegenetic processes within the wider framework of the task. It is possible that this latter ability is an example of the "apprehension of principles of order" discussed by Hearnshaw\textsuperscript{37}. The results of the second order analysis strengthen this interpretation, the 3 factors identified being: (1) likeness, i.e. a particular type of relation (Cf. Spearman's emphasis on likeness as a fundamental relation\textsuperscript{38}, (2) educing relations and eduction of correlates, i.e. the mechanisms involved in dealing with relations, (3) synthetic process, i.e. "all the possible reshaping and redistribution of the elements and the use of the 'instruments' by means of which this process is accomplished"\textsuperscript{39}.

What is being suggested here is that the noegenetic principles do not take into account explicitly enough, the restrictions of the task which make some eductions relevant and others irrelevant. Support for this point of view comes

\textsuperscript{37}Hearnshaw, Op. Cit. p319.
\textsuperscript{38}Spearman, Op. Cit. 1923, p71.
from Yela\textsuperscript{40} in his interpretation of a second order general factor, common to all the primary cognitive tasks, as involving:

the capacity of the subject to understand his task, and his working, i.e. to see the meanings of words, percepts, spatial relations and logical relations, and to see the relation of these meanings to one another, and to the solution of the problem. This general feature in all intellectual tasks has been named by Spearman 'abstraction and noegenesis!'\textsuperscript{41}

The concept of control or direction over and above noegenesis is a re-emphasis of a concept developed by Binet\textsuperscript{41} in his analysis of thought.

(c) Factorial Studies of Reasoning.

If reasoning is regarded as either a major component of, or identical with intelligence, then factorial analyses of reasoning tests may add something to our understanding of the processes underlying intelligence. Guilford, in a series of studies beginning with the USAAF Research Programme, has attempted to break up reasoning into a number of sub factors - although Vernon\textsuperscript{42} has cast some doubt on the reliability and validity of the procedure. A detailed des-

\textsuperscript{40}Yela M., "Application of the Concept of Simple Structure to Alexander's Data," 1949, Psychometrika, 5, p131. (italics not in original).

\textsuperscript{41}Binet A. and Simon T., The Intelligence of the Feebleminded, 1916, (Baltimore:Williams & Wilkins) p135.

cription of all the factors extracted is not appropriate here, but a few points arising from various studies are relevant:- (1) The attempt to design tests of "judgement" and "integration," although within a specialized field\textsuperscript{43}, is an illustration of what may be done by departing from the conventional type of test (Cf. the Melbourne "Test 90" of Lafitte\textsuperscript{44}), and thus perhaps tapping aspects of intelligence involving a "time dimension" and "judgement" (viz. Hearnshaw\textsuperscript{45}). (2) A general reasoning factor has appeared repeatedly, the typical defining tests being arithmetical, and the major characteristic being something similar to Thurstone's restrictive reasoning. This factor has not been interpreted consistently e.g. Guilford\textsuperscript{46} wrote - "Our present conception is that it is an ability to structure a problem preparatory to its solution" - a process in which trial and error plays some part. Later\textsuperscript{47} the factor was


\textsuperscript{45}Hearnshaw, Op. Cit.

\textsuperscript{46}Guilford J.P., \textit{Some Recent Findings on Thinking Abilities and Their Implications}, 1952, TA. and D. Informational Bulletin, p52.

described as involving perhaps a trial and error manipulation of symbols (especially when problems become more difficult), or symbolic span, or speed of symbol manipulation. The earlier interpretation fits in well with the concept of control or direction suggested above, although the later interpretation does not. While Guilford did not believe that reasoning abilities could be reduced to one general factor, he did concede that "the two aptitude scores representing verbal comprehension and general reasoning cover fairly well the variance in traditional intelligence tests."\(^{48}\) (Cf. Vernon's g - v: ed.). (3) Of more direct interest is the naming of the other reasoning factors, deduction, which was called "sensitivity to logical necessity"\(^{49}\); eduction of perceptual relations, eduction of conceptual relations, eduction of conceptual patterns, and eduction of correlates\(^{50}\). If these latter factors be called "noegenesis" and the general reasoning factor "control," then there appears essentially the same scheme as developed above, plus a re-interpreted deductive factor (Cf. this last factor with Burt's\(^{51}\) small group factor, and Vernon's\(^{52}\) discussion on

\(^{48}\)Guilford and Lacey, Op. Cit. p. 55
\(^{49}\)Green, et. al. Op. Cit. p. 152
\(^{50}\)Ibid, pp. 153-156
logical group factors).

Adkins and Lyerly\textsuperscript{53} in a factor analysis of reasoning tests after a preliminary analysis of 38 A.A.F. tests loaded on factors associated with reasoning, selected 65 reasoning tests and applied Thurstone's centroid method. As well as the usual factors, the following reasoning factors were extracted - perception of abstract similarities (Cf. Rimoldi's "likeness" factor), - hypothesis verification (Cf. Green Guilford et al. "sensitivity to logical necessity"), - flexibility of perceptual closure (Cf. Thurstone), - deduction, - concept formation. No general reasoning factor appeared, nor was there any confirmation of separate factors for educing relations and eduction of correlates. The authors suggested that the first two factors plus concept formation are all related to Spearman's 'g'. A second order analysis\textsuperscript{54} of the same data resulted in the extraction of these factors:- precision in formation and use of verbal concepts - i.e. "the perception and manipulation of verbal relations" (Cf. "noegenesis"?); - general word fluency - i.e.


"a type of fluency in educing possible relations to be tested for similarity" (Cf. Adkins and Lyerly, perception of abstract similarities?); - visualization "involving maintenance of a framework" (Cf. strength of perceptual closure?); - speed in analysis, an analytical ability possibly at the perceptual level; - flexibility in analysis, with the highest loading on flexibility of perceptual closure. This analysis is closer to Thurstone's "guess" of a general reasoning factor (inductive) and closure factors as representing Spearman's central intellectual factor.

These analyses illustrate the difficulty involved in the description and naming of factors and the disparities in interpretation from study to study. Although one solution to this would be to extract a general factor first of all, this does not help to clarify qualitative differences in the processes underlying 'g', but may give a false sense of psychological stability in interpreting the general factor, and obscure the interplay of whatever processes are involved in manipulating relations. An example of the difficulties involved in interpretation is provided by the

analysis of reasoning tests by Corter\textsuperscript{56} and the subsequent reinterpretation of the same data (after re-rotation) by Bernreuter\textsuperscript{57}. Corter's eight factors included three of particular relevance - academic, "a learned activity factor that includes all those things that we ordinarily call 'academic' in nature"\textsuperscript{58}; judgement, "the weighing of hypotheses, evaluation of thought materials, criticism and discrimination. It finds its expression in a wide variety of tasks in which control of thinking is of primary importance"\textsuperscript{59}; flexibility, or the ability to change one's mental set. Bernreuter accounted for the same correlations with four factors - induction, deduction, synthesis and fractionization, "the breaking down of a gestalt in order to extract a sub-gestalt"\textsuperscript{60} - a very different pattern of

\begin{itemize}
\item \textsuperscript{56}Corter H.M., \textit{Factor Analysis of Some Reasoning Tests}, 1952, Psych. Mon. 6, No. 8.
\item \textsuperscript{57}Bernreuter R.G., "Implications of Recent Studies of Intelligence," \textit{Transac. N.Y. Acad. Sc.} 1953, pp301-305.
\item \textsuperscript{58}Corter, \textit{Op. Cit.} p17. Cf. this with French's interpretation of "general intellective" factors as "schooling" - "an experiential factor resulting from variance in amount of formal education or other intellectual activity" (French, \textit{Op. Cit.} p238); also with Halstead's central integrative field factor - "the organized experience of the individual" (Halstead W.C., \textit{Brain and Intelligence}, 1947 (Univ.Chicago Press) p96); also Vernon's verbal - numerical - educational factor (Vernon \textit{Op. Cit.} 1950).
\item \textsuperscript{59}\textit{Op. Cit.} p19.
\item \textsuperscript{60}Bernreuter, \textit{Op. Cit.} p305.
\end{itemize}
factors from that of Corter. The chief value of Corter's analysis, lies in the confirmation of "control" and "flexibility", however, in view of Bernreuter's interpretation, his results must be accepted with caution.

Two further studies warrant some attention, those of Howie\(^{61}\). In the first of these, using Burt's Group Factor method he extracted a general factor, and among the group factors a reasoning factor similar to Thurstone's restrictive reasoning which he called "conceptual concentration," a term which may be useful in describing more accurately what is implied by "control" - "a capacity to make and hold a pattern in complex conceptual material subject to all the disturbing effects of diversity of possible linkages and associations...the maintenance of a directive set in the face of distractions"\(^{62}\). On the other hand this factor may be psychologically closer to the closure factors, although in the 1953 analysis, which was designed to investigate this reasoning factor more fully, it was interpreted as "the ability to think or to order conceptual material in terms


of a constraining form"63 - i.e. a concept of control or restriction likened by Howie to Hearnshaw's "apprehension of principles of order" (Cf. Rimoldi "bringing parts together into a meaningful solution").

(d) **Closure factors.**

As has already been noted, Thurstone suggested that the two closure factors - speed and strength of closure, and flexibility of closure, may be related to the central intellective factor. In a later study64 he suggested that the first closure factor might be associated with inductive and the second with deductive thinking. The whole concept of closure is closely related to control or direction and thus merits closer study.

Reyburn and Taylor65 claimed to have shown that 'g' is complex and among the five factors identified was one described as the finding of a significant pattern in a mass of irrelevant detail. This appears to be similar to Thurstone's A - speed and strength of closure. Rimoldi

---


described a factor represented by conflict between two or more configurations - "the more plastic the subject the more he is likely to be 'gestalt free' and the better will be the solution"66 - and compared this with Thurstone's E, flexibility of closure and Meili's67 "plasticity"; Thurstone's A was specifically linked in this study with factor B - "perception of spatial relations"68. In his later study factor A or "reasoning" was described as the interplay of analysis and synthesis, with the stress on the "plasticity" required.69 Another factor, G, which he described as a bringing of parts into a meaningful solution was compared with Meili's "globalization". In the second order analysis this factor had a high loading on "synthetic process" - a general description of what is involved in "globalization".

A closure factor (Z) was found by Yela and interpreted as "speed of closure of a pattern to be formed following some formal instructions and against conflicting or irrelevant elements".70 He also referred to the ability to synthesise

70Yela, Op. Cit. p126
the elements given into a meaningful whole and suggested that his factor Z was probably a composite of Meili's two synthetic factors, globalization and plasticity. After pointing out that both the primaries Z and R (reasoning) had the highest saturations in the second order general factor, he wrote - "These results present some additional evidence to the findings of recent research in pointing out the importance of a closure or synthetic perceptual factor in the performance of tests of intelligence."\(^71\)

He also quoted supporting evidence from an unpublished thesis by Bechtoldt on the factorial study of perceptual speed. French in his summary of Bechtoldt's investigation described the two closure factors as - "G. Facility in reconstructing formal perceptual material possessing weak intrinsic structure...Y. Facility in organizing simultaneous visual configurations under distraction of continued act."\(^72\)

The nearest approach to closure factors in Guilford's A.A.F. study appeared to be two "integration" factors\(^73\), a factor representing the persistence of a complicated mental set operating in complex clerical type work (Thurstone's A?).

---

\(^71\) Op. Cit. p127


\(^73\) Guilford and Lacey, Op. Cit. p225.
and a factor representing adaptability of mental set (Thurstone's E?). No closure factors were found in the study of reasoning abilities\(^7\), possibly because tests were chosen to test four major hypotheses and a number of sub-hypotheses, none of which were concerned with closure. In the study of creative thinking abilities\(^5\), three relevant factors were described, the first of which, closure, represented a combination of Thurstone's two closure factors. The other two represented two types of flexibility - adaptive flexibility, the ability to change set to meet new requirements, and spontaneous flexibility or lability of ideas, which also requires a change of set but the direction of the change is not restricted by the requirements. (Cf. with the second "integration" factor, and note that Guilford distinguished these factors from fluency factors). A further factor of interest in this study is one named "redefinition" - a shifting of the function of an object or part of an object in order to use it in a new way. This may represent an aspect of closure, and is closely related to a process described by Duncker (see below).


Adkins and Lyerly\textsuperscript{76} described two closure factors - speed of perceptual closure, in which a chaotic perceptual field is organized and flexibility of perceptual closure; while in the second order analysis\textsuperscript{77} a factor was tentatively identified as "flexibility in analysis". A similar factor was described by Corter\textsuperscript{78} as "a sort of ability to change one's mental set, or plasticity, or mental flexibility, or ability to try a new task."

These studies point to the existence of processes related to the manipulating of relations under the constraint of the given task - processes important for reasoning and thus underlying intelligence. Two studies have been directed specifically towards the exploration of closure and reasoning, those of Botzum\textsuperscript{79} and Pemberton\textsuperscript{80}. Botzum identified speed of closure and flexibility of closure among the primaries, and in the second order domain a composite reasoning factor with high loadings on space,

\begin{flushleft}
\textsuperscript{76}Adkins and Lyerly, Op. Cit. \\
\textsuperscript{77}Matin and Adkins, Op. Cit. p.78 \\
\textsuperscript{78}Corter, Op. Cit. p.21. \\
\end{flushleft}
deduction, induction, and flexibility of closure. He explained the composite nature of this factor by pointing out that space and closure tests can be solved either analytically or synthetically, and furthermore both induction and deduction involve certain configurational and gestalt elements. The other second order factors are of interest - analytic "something akin to Spearman's noegenesis" configuration or closure, and a bipolar factor - at one pole speed of association with well practiced associations, and at the other pole speed of closure, i.e. completing unstructured configurations. This bipolar factor parallels the results from an experiment by Saugstad in which there was negative correlation (ranging from .10 to .60) between incidental memory and problem solving.

The aim of Pemberton's study was to test the hypothesis that the closure factors in perceptual tests would generalize to tasks requiring higher cognitive functions - an assumption often made implicitly in many of the foregoing investigations.

---


Her hypothesis was strongly supported with regard to the second closure factor - flexibility, although the evidence was not so clear cut for the first closure factor. One of the most interesting results was the confirmation of Thurstone's hypothesis that flexibility of closure (C2) is associated with a deductive or analytic thinking, while speed of closure (C1) is associated with inductive thinking. A "reasoning" factor, with its highest positive correlation with C1 was interpreted as involving a synthetic apperception of the solution - similar to Rimoldi's G, "Bringing parts together into a meaningful solution," and Meili's "globalization." On the other hand C2 was interpreted as corresponding closely with Meili's "plasticity" and Rimoldi's A - interplay of analysis and synthesis in which plasticity is required. In the second order analysis the four factors extracted were interpreted as involving analytic processes, synthetic processes, speed of perception and fluency - i.e. the closure factors were regarded as components of traditional analytic and synthetic activities.

(e) **Summary.**

This review of factorial studies has been presented in some detail because a mere listing of factors from many studies does not indicate the rather complex relationships among factors which may be superficially different. The
evidence presented has not been directed at the problem of the existence of 'g', for in fact most of the studies cited in detail have used multiple factor analysis. On the other hand the evidence has been used as suggesting hypotheses as to the nature of 'g', and this must be distinguished from the practical application of multiple factors as for example in test construction. If 'g' represents a general level of cognitive functioning it is just as arbitrary to describe this function as noegensis plus abstraction, as to describe it as involving e.g. - globalization, plasticity, fluency and complexity (Meili). Ultimately the factorial analyses must be supported by data drawn from other types of psychological investigations, i.e. must enter into wider laws of cognitive functioning.

In the preceding sections the main argument developed has been that intelligence consists of processes which, at least, involve the manipulating of relations, but that while Spearman drew attention to certain formal characteristics of relations and their fundaments, his analysis did not penetrate far enough. It appears from the factorial studies of intelligence and reasoning, that there is some confusion in the interpretation of factors. At times the terms used are based upon the formal logical requirements of the problem - e.g. -
deduction, induction\textsuperscript{84} and at other times upon somewhat vague introspective descriptions, especially of the Gestalt type - e.g. closure, globalization, restriction. This "struggle to interpret" as McNemar\textsuperscript{85} has called it, is a most crucial part of analyses in such a complex field as that of intelligence. Thurstone\textsuperscript{86} summed up the difficulty when he wrote - "It seems to be essential, and it seems to be psychologically reasonable, that we must judge introspectively the psychological processes that are involved in the principal task of a test."

Despite this difficulty the factors described above go at least part of the way towards enriching Spearman's analysis, particularly in stressing (a) the firmness of control or direction in the task situation - i.e. a

\textsuperscript{84}French, (Op. Cit. 1951) has deduction listed as a factor in 37 of the 69 analyses he covered - induction only 9 times. This identification of deduction may be due to this type of test item being relatively easy to construct, or to the fact that the investigators can recognise the formal requirements of deduction in a variety of situations more easily than for example induction.


selective or restrictive factor\(^{37}\), (b) the flexibility of manipulation where several patterns of relations conflict or distract i.e. a flexibility or plasticity factor, (c) the bringing together of parts into a pattern i.e. strength of closure, globalization, synthesis, (d) the interplay between, or bias towards, analysis or synthesis - possibly a characteristic of flexibility.

IV. - THEORETICAL AND EXPERIMENTAL STUDIES OF REASONING AND PROBLEM SOLVING.

This field is extensive, and the sections following are not intended as systematic reviews, but as illustrations of the ways in which the analysis of intelligence as developed above may be linked to non factorial studies of reasoning and problem solving. Both terms have been used to cover the same type of investigation, and deal with that aspect of behaviour usually designated as "thinking". One of the interesting features of such investigations, is the assumption, usually unformulated, that "intelligence" and "problem solving" are separate fields of research, thus apart from an occasional correlation of problem solving

\(^{37}\)This has been recognised by Knight in his definition based upon an elaboration of Spearman's noegenetic principles, see Knight R., Intelligence and Intelligence Tests, 1948, 4th ed., (London: Methuen) p.16.
scores with intelligence test scores, there have been few attempts to link up the processes underlying both. This would be understandable if intelligence was formally defined as 'g', in the sense of a statistical concept, but as a rule no such formal definition is rigidly adhered to. The explanation may lie in the fact that intelligence is usually thought of as a quantitative index, with little regard for the qualitative character of what is measured, apart from the assumption that it has to do with "relational thinking"—yet this could equally well be a general reference to problem solving and reasoning!

(a) Gestalt Studies.

Although the emphasis on insight in Kohler's original work did not throw much light upon the actual production of the solution, it did draw attention to the fact that one of the essential aspects of problem solving is the structuring of the situation around the focal points of data given and goal desired. Duncker examined this structuring more closely and built up a theory centered around the concept of "functional solution," based upon the analysis of protocols of subjects working through a number of problems. In a

88Duncker K., On Problem Solving, 1945 (Trans), Mon. No. 270.
problem the facts given are "explicated," i.e. elaborated, and relationships between the facts clarified, in terms of the goal. At the same time the goal is analysed by making explicit the conditions which must hold if the requirements of the goal are to be met. The gap between "explicated" data and "analysed" goal specifies the nature of the solution, and leads to the formulation of a "functional" solution or to a series of interlocking functional solutions. A further defining of the gap results from analysis of conflict (i.e. what is wrong?) and analysis of material (i.e. what can I use?), and these analyses lead to a restructuring of the problem - a process influenced by the functional fixedness (embeddedness) or looseness of the elements in the problem.

From this brief description of Duncker's theory, it can be seen that the processes involved go beyond the formal principles of noogenesis and link up with such factors as restrictive reasoning i.e. the analysing of the goal, or defining the task which will control the selection of relations; flexibility of closure and plasticity, i.e. the restructuring of the material, the loosening of embedded elements; speed and strength of closure, synthesis, globalization, i.e. the organizing of the unstructured material into a pattern, the combining of patterns, the forming of functional solutions;
hypothesis verification and sensitivity to logical necessity, i.e. the testing of a functional solution against the requirements of the "gap." Duncker related these processes to intelligence quite specifically - i.e. "It seems very probable that the greatest differences between individuals as to so called intelligence...are based on differences in the facility of restructurings"\textsuperscript{89}.

Wertheimer\textsuperscript{90} criticised those who explain productive thinking in terms of the seeing of relations and summarised his own approach thus:- "To realize any relations, even if they are correct is not decisive, what is decisive is that they must be the relations structurally required in view of the whole, arising, conceived, used as parts in their function in the structure". This is essentially the same view-point as Duncker's, although Wertheimer lays more stress on the process of restructuring - i.e. grouping, reorganizing, realizing structural hierarchy, structural transposibility, recentering etc. Both Rimoldi and Guilford make use of these concepts, e.g. Rimoldi's factor Gamma\textsuperscript{91} - reshaping and redis-

\textsuperscript{89}Ibid, p30.

\textsuperscript{90}Wertheimer M., Productive Thinking, 1945, (N.Y. Harper), pp43-44.

\textsuperscript{91}Rimoldi, Op. Cit. 1951.
tribution of the elements, and the Wilson Guilford et al. factor "redefinition" or recentering. Wertheimer also drew attention to the importance of the finding of the right question - "putting the productive question is often more important, often a greater achievement than solution of a set question". Wilson, Guilford et al. used a similar concept i.e. "sensitivity to problems" in interpreting their data, but as there were only two loadings greater than .30, this factor is of dubious value. This may, however, prove to be a fruitful line of investigation, especially in analysing the intellectual performance of high ability groups.

Two groups of experiments on problem solving - those of Maier and Luchins - clarify some aspects of closure and set. Maier's concept of "direction" emphasised the selective effect of the particular task as formulated by the subject, i.e. the direction "determines the way various objects will be experienced". Directions may become habitual and prevent the emergence of solutions, thus Maier contended

93 Wertheimer, Op. Cit. p123
that "reasoning involves the inhibition of persisting habits as well as the ability to form solution patterns." Although Maier did not relate his analysis to intelligence, the processes he described are directly related to the manipulation of relations and have a close bearing on the concepts of closure and of restrictive reasoning or control.

Luchins constructed a particular type of problem situation in which the structure of the material created an inhibiting set. This is apparently a more specialised case of Maier's habitual direction, and again indicates the importance of plasticity or flexibility in problem solving. There is, however, a more fundamental implication in Luchins' work - i.e. the set or rigidity may be generalized and represent an important dimension of cognitive functioning - e.g. the negative side of flexibility of closure. Although a general rigidity factor has been postulated by Rokeach, an analysis by Cattell and Tiner resulted in the

breaking up of this general factor into types of rigidity - e.g. rigidity characteristic of low intelligence, of low fluency, or of ideational inertia. Later work has thrown considerable doubt on the generality of rigidity, in fact Schroder and Rotter\textsuperscript{100} claimed that rigidity as a specific trait is no-existent but is "a failure to learn something rather than an inherent or original trait". They preferred to use the concept "flexibility" as referring to "a kind of higher level behaviour which consists of expecting change and looking for alternative pathways"\textsuperscript{101}, behaviour which results from specific learning experiences.

Goodstein\textsuperscript{102} found no evidence to support Rokeach's conclusions, while Forster et. al\textsuperscript{103} in an experiment based on a group of well known problems rejected the hypothesis that there exists a generalized factor of intellectual functioning such as rigidity - flexibility. Their dis-

\textsuperscript{100} Schroder H.M. and Rotter J.B., "Rigidity as Learned Behaviour," \textit{J. Exp. Psych.} 1952, 43, p142.

\textsuperscript{101} \textit{Ibid}, p143.


cussion on the use of such terms is worth quoting:–

We are inclined, however, to believe that terms like recentering refer to a kind of behaviour, rather than to a factor or ability - as one would say that fear is a kind of behaviour, so that anyone might display fear in some situations but not in others... if the basis for flexibility-rigidity is sought in the background of experience, then the notion that it is a general factor of intellectual functioning must be abandoned.

Although flexibility in problem solving may depend largely on experience, this would not automatically class it as being irrelevant to the functioning of intelligence, unless by intelligence is meant some purely "innate" capacity (see discussion later in this chapter). In a relevant experiment by Fattu et. al\textsuperscript{104} a complicated mechanical problem was used to test certain criteria which might distinguish between good and poor problem solvers. The results indicated a clear cut difference on only one of the criteria - namely stereotypy of behaviour, or problem solving rigidity.

These examples of investigations in the general area of problem solving, particularly from the Gestalt viewpoint, illustrate the sort of linkage between factorial analyses and non factorial experimental studies which may lead to a more fruitful interpretation of the nature of intelligence. There

is, however, another group of studies which presents an alternative explanation to the Gestalt theory of problem solving, and which poses some important questions in relation to intellectual functioning and development. Examples of this type of explanation have already been touched on in discussing the work of Schroder and Rotter, and Forster et al., namely the emphasis on learning in problem solving and reasoning.

(b) Learning and Problem Solving.

The word "learning" has been used in as great a variety of ways as "intelligence" and it is therefore not surprising that the relations between the two concepts have not been rigorously defined. In investigations of problem solving, however, there has emerged a sharpening of the distinction between what Harlow has called "uni-process theorists," i.e. those who explain problem solving in terms of learning, and "dual-process theorists," i.e. those who make a distinction between learning and productive thinking (e.g. Maier). Another way of illustrating the distinction is


in the difference in emphasis on the subject with his repertory of habits and methods, or the problem situation with its particular field structure.

Modifications of the original emphasis on insight as a sudden restructuring of the field, have been in the direction of stressing the active participation of the subject, e.g. in such terms as exploration (Alpert\textsuperscript{107}), orientation (Adams\textsuperscript{108}), search (Bulbrook\textsuperscript{109}), trial and error experimentation (Hartmann\textsuperscript{110}), preliminary adaption (Drever\textsuperscript{111}) or inspecting (Chrisof\textsuperscript{112}). In a repetition of some of Kohler's original experiments with chimpanzees, Birch\textsuperscript{113} has shown that one of the pre requisites for


\textsuperscript{110} Hartmann G.W. "Insight Vs Trial and Error in the Solution of Problems," \textit{Amer. J. Psych.} 1933, 45, pp663-677.


\textsuperscript{113} Birch H.G. "The Relation of Previous Experience to Insightful Problem Solving," \textit{J. Comp. Psych.}, 1945, 38, p381.
successful solution is experience with the materials used in the problem, and while he did not ignore the external field properties of the situation, he put the emphasis on learning. His discussion of the experiment serves as a good summary of the uni-process point of view.

This ability to reorganize previous experiences in accordance with the requirements of a new problem situation is the essential feature of problem solving. The ability to select from the available repertoire of recall and reorganize into new patterns of response previously learned, but not contiguously learned items of experience, makes possible an enormous expansion of the adjustive possibilities of an organism.

Hartmann distinguished between comprehension of the internal structure or organization of a situation and sagacity in selecting appropriate methodological approaches to the task. This distinction is of value in helping to

---


bridge the gap between uni and dual process theorists and at the same time gives a point of contact between these studies of problem solving, and intelligence. It would be generally conceded that there is some neurological limit to the complexity of structure which a given organism is capable of comprehending and if this could be measured accurately it would be an index of an innate component or dimension of intelligence, assuming that experience could not affect this limit (although such an assumption may not be entirely valid - see discussion on Hebb below). On the other hand there is in any problem, or in any situation which is regarded as a measure of intelligence, a "methodological" component or medium which cannot be completely separated from the act of comprehending, and which depends upon the experiences of the organism. This is such a reasonable assumption that its implications can easily be overlooked because of the comparative uniformity of the sorts of experiences available in a particular culture. Thus while it would not be surprising to find that certain characteristics of constructive thinking among Greenlanders\(^{117}\) may differ from those of groups of adults in our own culture, it would not be so easy to dis-

tistinguish differences in experience as these may affect the ways in which children approach and handle a problem, such as a complicated syllogistic type item in an intelligence test.\textsuperscript{118}

The active participation of the subject in searching, experimenting, inspecting etc. is the result of a long process of learning, a good deal of which is informal.\textsuperscript{119} Specific experimental evidence on this point is important, not only as supporting the general statement, which is after all commonsense, but because it may link up this discussion of learning and problem solving with the central topic of intelligence.

Harlow\textsuperscript{120} has reported experiments in which he was able to trace the development of learning sets in monkeys. His particular standpoint is illustrated in the following

\textsuperscript{118}Such differences in experience may be relatively unimportant from the point of view of prediction of scholastic success, but of considerable importance for a consistent theory of intelligence.

\textsuperscript{119}This leaves out of account the importance of motivation or ego-involvement in the task e.g. see Birch H.G. "The Role of Motivational Factors in Insightful Problem Solving," J. Comp. Psych., 1945, 38, pp295-317, and Adcock C.J. Intelligence and High Level Achievement, 1952. Public. in Psych. No. 1. (Wellington: Victoria Univ. College)

quotation - "We wish to emphasize that this learning to learn, this transfer from problem to problem which we call the formation of a learning set, is a highly predictable orderly process which can be demonstrated as long as controls are maintained over the subject's experience and the difficulty of the problems."\textsuperscript{121} In a summary of his position\textsuperscript{122} he claimed that the range of appropriate reaction tendencies\textsuperscript{123} (i.e. learning sets) is a primary determiner of the efficiency of thinking, and these reaction tendencies are the result of experience or instruction in a wide variety of situations. It is not difficult to find evidence on the effects of training or experience in problem solving e.g.

Chant\textsuperscript{124} found two types of approach to problems, the interpretative and the analytic, associated with different academic backgrounds; Maier\textsuperscript{125} presented evidence for a considerable

\textsuperscript{121}\textsuperscript{Op. Cit.} 1949, p54.


\textsuperscript{123}\textsuperscript{Cf. also Tolman who described the learning and use of "field cognition modes," (Tolman E.C. "There is More Than One Kind of Learning," \textit{Psych. Rev.} 1949, 56, pp144-155) and Berlyne who traced the development of "trains of thought" from simple responses to "ratiocination" (Berlyne D.E., Knowledge and Stimulus Response Psychology, \textit{Psych. Rev.} 1954, 61, pp245-254.}

\textsuperscript{124}\textsuperscript{Chant S.N.F. "An Objective Experiment in Reasoning," \textit{Amer. J. Psych.} 1933, 45, pp282-291.}

\textsuperscript{125}\textsuperscript{Maier, Op. Cit.} 1933
improvement following a short lecture containing general suggestions about methods of problem solving; Eagleson\textsuperscript{126} reported an improvement in puzzle solving; Sargent\textsuperscript{127} demonstrated the effects of different training procedures on anagram problems; and Faust\textsuperscript{128} found rapid improvement in a "20 question" problem series; Bloom and Broder\textsuperscript{129} have given a detailed description of the effects of remedial training in problem solving; Furst\textsuperscript{130} has shown that critical thinking becomes less dependent on scholastic ability as instruction increases; while Burt\textsuperscript{131} has stated that "ability to reason is to a large extent a trainable technique".

Such illustrations as these do not, of course, prove very much, because from such limited experiments it is not

\textsuperscript{126}Eagleson O.W. "A Study of Puzzle Solving," \textit{J. Psych.} 1940, 9, 259-268.


possible to secure evidence on the important problems of permanence and generalization of transfer. They do, however, pave the way for a much more basic approach to problem solving in terms of learning theory, e.g. Hull\textsuperscript{132} has shown the possibility of understanding thought and reasoning in terms of his "fractional antedating goal reaction," while his analysis of "direction" in terms of generalization of drive stimulus and goal stimulus is very similar to Duncker's analysis\textsuperscript{133}. An example of the type of learning experiment which can be set up to investigate complex processes is illustrated in a paper by Maltzman and Morrisett\textsuperscript{134} in which the building up of a set is shown to conform to Hull's principles (see also earlier reference to the work of Schroder and Rotter). These examples indicate ways in which experimental techniques and theories of learning may eventually link up with the study of the "methodological" component of intelligence.


\textsuperscript{133}Tbid, p308. Duncker has linked his theory with data based upon the discrimination learning of the rat as studied by Krechesky and Tolman—see Duncker K. and Krechevsky I. "On Solution Achievement," \textit{Psych. Rev.} 1939, 46, pp176-185.

It has been pointed out that it is necessary to distinguish between several uses of the word "intelligence"\(^{135}\). Hebb, for example distinguished between intelligence A - innate potential, and intelligence B - average level of performance or comprehension. While Hebb was mainly interested in the effects of early experience in relation to B, there is no reason why the influence of later learning should not also be considered. There have been many reviews of environmental influences on intelligence test scores\(^{136}\) although one of the great difficulties in interpreting these studies is that of terminology. Furthermore, while it is possible to specify and control certain influences - such as coaching - it is very difficult to specify the intellectual stimulation value of the "environment" in the broad sense of home or school - and even more difficult to control it, or to know at what stages of development certain stimuli are crucial. While the detailed technical problems associated with fluctuations


\(^{136}\)Interesting additions to the literature are the brief reports by Anastasi, in the Annual Review of Psychology (Stone C.P. and Taylor D.W. eds. (Stanford: Annual Reviews Inc). pp147-148) of studies in Sweden and Holland on the influence of length and type of schooling on intelligence test scores - Cf. similar results in the study by Lorge - "Schooling Makes a Difference," Teach. Coll. Rec., 1945, 46, pp483-492.
in intelligence test scores are gradually being clarified, it is necessary at the same time to study the implications of the experience or methodological factor in the more general descriptions of aspects of intelligence, such as Hebb's "B".

The various studies described in this section have been selected, not because they represent every aspect of problem solving and learning, but because such groups of studies help to specify the ways in which experience enters into the functioning of intelligence. In order to stimulate more research on problems related to this functioning it is necessary to broaden the concept of intelligence by linking it more closely with cognitive functioning in general, in terms not only of measurement but also of development.

V. COGNITIVE THEORIES.

Spearman developed his theory of intelligence within the framework of a more general cognitive theory, but since that time discussions of intelligence have in the main hinged around specific problems, while more general theoretical problems have been neglected. Although this has resulted in a necessary and valuable clearing of the ground, it seems that the time is ripe for a new approach to the concept of intelligence in the light of developments since Spearman's original theory was formulated. The "tidying up"
process developed in this chapter has been used in order to emphasise inter-relationships among certain groups of studies, and to suggest that a theory of intelligence must take into account qualitative differences in the processes which underly intelligence, the relation of these processes to perception and thinking, and the part played by learning in their development.

It has been pointed out that Hebb and Piaget have contributed towards a theory of intellect, yet one of the outstanding features of these two writers, which does not seem to have been sufficiently appreciated, is the complementary nature of their theories. Furthermore, both theories account for the influence of experience in intellectual development, and deal with intelligence in the context of a more general cognitive theory. In brief, their emphasis

---

137 These have not included studies on concept formation - Vinacke has given an excellent review of relevant studies, (Vinacke W.E. "The Investigation of Concept Formation," Psych. Bull., 1951, 48, pp1-51) and Hearnshaw has pointed out the implications of these studies for a theory of intelligence (Hearnshaw L.S. "The Psychological Study of Conceptual Thinking," Brit. J. Psych. 1954, XLV, P5). Studies on "fluency" have not been included, (e.g. see Rogers C.A. "The Structure of Verbal Fluency," Brit. J. Psych. 1953, XLIV, pp368-380) although it is likely that there exists some functional connection between fluency and flexibility.

is not on the abilities which make up the "structure" of the mind, but rather on the interplay between hereditary mechanisms and experience which results in the "construction" of the mind. What Thomson\textsuperscript{139} has described picturesquely as "an intricate network of possibilities of intercommunication...sub-pools...categories interlaced and interwoven..." Hebb\textsuperscript{140} and Piaget\textsuperscript{141} have described in terms which can be defined and used systematically, yet which express essentially the same point of view as Thomson's.

The two theories are not unique in showing the relationship between perception and intelligence, for example Heidbreder\textsuperscript{142}, Krech and Crutchfield\textsuperscript{143}, Combs\textsuperscript{144} and

\textsuperscript{140}Hebb, Op.Cit.
\textsuperscript{144}Combs A.W. "Intelligence from a Perceptual Point of View," \textit{J. Abnorm. Soc. Psych.}, 1952, 47, pp662-673.
Krech and Calvin have all attempted to clarify this relationship. They are, however, outstanding for the way in which they have dealt with experience as necessary for the development and completion of categories of thought. Although Hebb's theoretical "model" is neurological, and Piaget's biological, the correspondence between the two systems is very close, as can be seen from the following examples. Hebb's basic structure is the cell assembly, a diffuse structure of cells which develops through repeated stimulation - Cf. Piaget's schema built up from repetition of a reflex under varying conditions. Both can act as "closed systems", and in fact it is only by this action that they become consolidated, but while so acting they not only facilitate (Hebb) or assimilate (Piaget) new sensory data, but also accommodate (Piaget) or integrate through synaptic growth (Hebb). Repetition and association of cell assemblies leads to the development of phase sequences (Cf. Piaget's co-ordination of schemata) which when established have some independence from sensory control and are typically "short circuited" in a rapid discharge (Cf. the growing efficiency of Piaget's circular reactions). In both

145  
theories the selectivity of attention or set is explained in terms of the relationship of the structures built up, to the particular stimulus, i.e. central facilitiation (Hebb) intentional sensorimotor adaption (Piaget). The development of thinking is treated as the ever increasing complexity and freedom from sensory dominance of the phase sequences (Hebb) or the elaboration and operational grouping of mobile structures (Piaget).

These few illustrations do not really bring out the most significant feature of the theories - i.e. their demonstration of the operation of intelligence as "embracing the whole mental organization".\textsuperscript{146} Piaget speaks for both:

"Intelligence is thus only a generic term to indicate the superior forms of organization or equilibrium of cognitive structurings... this use of the term precludes our determining where intelligence starts"\textsuperscript{147}. From another point of view the theories spring from a common viewpoint represented again by Piaget\textsuperscript{148} - "The schema (Cf. Hebb's cell assembly) is therefore a Gestalt which has a history," and it is in the investigation of this "history" that the studies of learning and intelligence coalesce. Hebb shows why such a

\textsuperscript{146}Quoted from Meili by Myers, \textit{Op.Cit.} p22
coalescence is necessary and how it can operate, while a
more detailed investigation of Piaget's central principles
of assimilation and accommodation, would in effect be a
study of generalization and discrimination in learning.
Although the two models briefly described represent a pro-
gramme for research rather than completed sets of demon-
strated principles, they do point to ways in which those
aspects of behaviour called perceiving, learning, thinking,
problem solving, and intelligence may be related more
meaningfully.

In fairness to Spearman, it should be pointed out
that although he emphasised a logical concept of intelli-
gence, his conception of the development of intelligence,
which he did not attempt to elaborate, was not after all so
different in principle from the more systematically expressed
conceptions of Hebb and Piaget - as can be seen from the
following statement:—

149

The eductive growth is unlike the building of a wall...
It is more like the waxing of a tree, which does not
first complete its roots, then stem, then branches, then
leaves, all in succession, but develops all these over-
lappingly. So in cognitive cellulation also, the
lower levels are allowed only a limited degree of
priority. Whilst they are still extremely obscure,
the upper levels already begin growing also, and to
the extent that their as yet very imperfect under-
structure becomes from moment to moment capable of
supporting.

One difficulty should be briefly mentioned before concluding - namely the problem of terminology, for in a discussion of this nature it is an embarrassment to have no active participle for the word intelligence. This is more than a grammatical difficulty, but strikes at the very root of the problem of the nature of intelligence. The question arises as to whether there is something to be gained by abandoning the word intelligence, as has often been suggested, and using instead descriptions such as "general scholastic ability" which imply measurement and which are frankly dependent on the type of criterion against which test are validated, or the type of prediction for which they are constructed. By this means the way would be open for the analysis of these abilities (as defined above) in terms of qualitative differences in the cognitive processes involved, and would clear away many of the presuppositions which have grown up around the concept of intelligence.

VI.- SUMMARY OF CHAPTER I.

1. There has been a tendency to confuse the analysis of intelligence as product with the investigation of intelligence as process.

2. Spearman's noegenetic principles plus abstraction have been accepted far more uncritically than his statistical analyses.
3. Thurstone's work led to a partial rapprochement with Spearman although his conception of the central intellectual factor was broader than Spearman's - in particular his inclusion of "closure" factors.

4. Later factorial studies of intelligence may be interpreted as adding the concept of control or direction i.e. the ordering of relations in the light of the task.

5. Factorial studies of reasoning have produced a crop of factors - some being interpreted in terms of Spearman's principles, others confirming the concepts of closure and control, while still others have been interpreted in terms of analysis and synthesis, or deduction and induction.

6. In non-factorial studies of reasoning and problem solving, the Gestalt analyses have helped to clarify the factorial interpretations, and while the studies of problem solving as learning have created some difficulties for a theory of intelligence, these difficulties may lead to a clarification of the role of learning and experience in intelligence.

7. In general these groups of studies have been selected as evidence that advances in understanding the nature of intelligence may come from a more conscious attempt to link together the interpretation of factors from analyses of intelligence and reasoning, the results of
investigations into the nature of problem solving, and developments in learning theories and experiments. A type of cognitive theory which can deal with intelligence in relation to these other fields has been illustrated by linking the theories of Hebb and Piaget.
CHAPTER II

STUDIES OF COGNITIVE ERROR.

It has been shown in chapter I that it is reasonable to interpret intelligence, as a process, in terms of problem solving and reasoning, and that such an interpretation not only fits in with many factorial studies, but also opens up the possibility of examining more closely the "methodological" component of intelligence. Although it is difficult to make a direct attack on this problem, precedents in such fields as reading and arithmetic, for example, (or even in the study of medicine, to use a much more spectacular example) suggest that an indirect attack through a study of mal-functioning may yield much valuable information about the processes which underly normal functioning.

As applied to the arguments developed in the previous chapter, this means that a study of errors in tasks designated "intellectual" should link up with the study of errors in the whole cognitive domain. Before attempting to investigate a segment of this very broad problem, it is necessary to consider what has already been achieved, and to note its bearings on the interpretations of intelligence already discussed. Although there are not many major accounts of cognitive error, there have been three or four systematic studies, and a
number of incidental references. The plan adopted in this chapter has been to present (I) the systematic studies, (II) experiments bearing directly upon them, (III) the Gestalt account of error, (IV) a classification of errors found in the literature on problem solving and reasoning, (V) Some specific errors in intelligence test items, (VI) a brief account of factorial studies of "carefulness", (VII) conclusions.

I.- SYSTEMATIC STUDIES OF ERROR.

There are two particularly significant accounts of cognitive error - significant because they both spring from more general theories of thinking and because they present contrasting viewpoints - these are the studies of Selz and Spearman. Strictly speaking, Selz comes first chronologically, although from the point of view of exposition, it is more valuable to deal with them in the reverse order, particularly as Spearman made no reference to the work of Selz, and Selz's account covers those aspects of Spearman's theory which are most inadequate.

(a) Spearman

Spearman wrote an article on the origin of error\(^1\) in which he claimed that the only sound theory from which error

can be predicted is his theory of noegenesis, in particular as it is affected by one of the minor quantitative laws - i.e. retentivity. He summed up his position thus:- "...the liability to error must needs ensue whenever an item arises in any context by virtue of reproduction, but is then taken as being there by virtue of intuitive evidence such as invests the three principles of noegenesis..." After showing that the two conditions which most favour the displacement of the correct response by an incorrect one, are resemblance and contiguity, he wrote - "But these two conditions of displacement...are just the two conditions of reproduction itself."²

Although he quoted approvingly from the work of Laycock, who explained errors in terms of associative reproduction replacing eduction, including "reproduction of methods...without any attempt to determine their suitability in the...situation" - "pupils reacted to the situation as a whole" - "no attempt to analyse the situations into sub-problems"³, he did not comment on the possibility that the errors were equally - on the evidence of the very examples quoted - failures in analysis or in the adequate structuring

²Ibid, p51.
of the problem, as well as failures due to reproduction.
He did not concede the possibility of incorrect eduction
being a cause of error, in fact although he pointed out
that the ascribing of error to creative imagination (i.e.
the operation of educing correlates) would be a formidable
objection, he quickly dispensed with the objection by stat-
ing that it is the process of mistaking the mental pre-
sentation for real existence which is the error, not the
actual educing of the correlate\(^4\). This defence is not
convincing, in fact it weakens his whole theory, for accord-
ing to this line of thought noogenesis does not then refer
to a really basic process in thinking, but merely to the
logical verification of a correct answer.

One further example sums up Spearman's point of
view and illustrates its weakness - "For any item or relation
whatever may fit their original context perfectly well and
yet, on being retained when this context changes, may result
in presenting any degree of misfit."\(^5\) He emphasised the
condition of "being retained," yet it would seem that the
error springs from the failure to understand the implications
of the change of context, while the reproduction of the item
or relation arises as a result of this faulty approach to the

\(^4\)Ibid, p49.
\(^5\)Ibid, p33.
problem i.e. the crucial factor is "relevance" (a matter of noogenesis?) not retentivity.

(b) Selz.

A critical account of Selz's theory of thinking has been given by Humphrey⁶ who while dubious of the "rather rarified doctrine of the Relational Fact, with the other implications belonging to the Austrian school..." yet stressed the originality and importance of Selz's emphasis on "the idea of total integration in the process of thinking." This process of integration is illustrated by Selz's rejection of the Wurzburg analysis of thinking as reproduction under the stimulus of a determining tendency⁷ and his stress on the total Aufgabe or task complex which initiates and directs the finding of the means to the solution. The similarity of this conception of productive thinking to Duncker's analysis can be judged from the following quotation from Selz⁸:

All general operations of the Finding of Means have this in common, that they are introduced by the schematic anticipation of operations which lead to a determinate result... The special operations which are being sought are indirectly determined in this anticipation by their consequences, which constitutes the goal; that is to say, they are indirectly determined as means to this end.

⁶Humphrey G. Thinking, 1951, (London: Methuen), p149.
⁷Ibid, chaps. 2-4.
⁸Quoted by Humphrey Op. Cit. pp139-140.
Humphrey criticized Selz's theory on the grounds that while it may be useful to refer to productive thinking as dependent upon a schematic complex, it is false to assume that thinking proceeds by complex - completion, for it is really a case of complex - formation\(^9\). Bearing in mind this limitation, it can be seen that Selz has given a useful clue to the understanding of error, namely in pointing to the possibility of a false or inadequate schema for a given problem, which then determines what means to the end will be chosen from the mental operations and concepts already available. There is not really very much difference between Selz's "schematically anticipated complex" and Spearman's noogenesis, but the main advantage of Selz's theory lies in its emphasis on the total structure of the task - just that aspect which is most neglected in Spearman's theory.

The discussion above would not have been warranted if the problem of error, in this theory, had been merely described in so general a way; however, its purpose has been to introduce the necessary background in order to give body to the specific theory of error developed by Selz. This theory has been described in detail by Wilcocks\(^10\)

\(^9\)Ibid, p143.

and the following details have been drawn from this description.

The theory is based on the assumption that errors arise, not because of strong reproductive tendencies, but as a result of the particular task directed intellectual operations which serve the solution. "The most common ground for these errors was, in his experiments, that a related problem served instead of the one actually set"\(^{11}\) - this may be called "task displacement". Conditions which may produce such displacement appear to be dependent upon a law of part efficiency - i.e. not all parts of the task actually set are sufficiently active. This law operates through the dropping of some characteristics of the task, the prepotency of particular elements, the low dispositional effects of some elements and readiness to attend to the more abstract or general characteristics of the task. What does remain of the original task complex, i.e. that part which still operates, is called the "residual complex"\(^{12}\).

\(^{11}\)Ibid, p27.

\(^{12}\)An interesting set of principles, in many respects very similar to those developed by Selz, has been worked out in systematic form by Hollingworth - namely his general laws of redintegration, (Hollingworth H.L. "General Laws of Redintegration," J. Gen. Psych. 1928, 1, 79-90,) which cover such characteristics as effectiveness of partial stimuli, completeness of response as a function of completeness of stimuli, greater instigative potency of prepotent details, reinforcement (synergy) or neutralization (inhibition) of stimuli, organization of stimuli as an expression of total past history, incomplete stimuli functioning as complete stimuli as a result of previous contexts. Many of these principles can be trans-
A number of conditions may affect or cause the operation of part efficiency - e.g. in general, misunderstanding of the task (a related but not identical task is attempted), or more specifically, lack of clarity which reduces consciousness for those parts of the task with low dispositional efficiency; distractions of attention; restless or hurried approach to the task; underestimation of the difficulty of the task; displacement of meaning of the task in the direction of a more familiar problem with an easier solution.

These conditions, as related to the general theory, give a much more satisfactory account of the genesis of error than that given by Spearman, although a combination of this account and Spearman's analysis of reproduction is probably even more satisfactory.

II.- RELEVANT EXPERIMENTS.

In the paper quoted previously, Wilcocks set out to examine a number of introspective protocols of simple reaction tasks in order to produce evidence which would show that Selz's description does not exclude the process of substitution - a source of error in thinking formulated by Muller and Pilzecker in terms drawn from the traditional

[Text continues...]

lated quite readily into a form compatible with Selz's error theory.
laws of association,\textsuperscript{13} and very similar to Spearman's description of reproduction. The evidence supported the conception of substitution or association as an error producing factor and pointed to the conclusion that part efficiency and substitution are not mutually exclusive processes.

One of the great limitations of Selz's work was that he based his experiments on a very simple situation e.g.- a type of reaction such as "Give the coordinate of parson". In view of this limitation an experiment by Fortes,\textsuperscript{14} in which the errors on perceptual intelligence tests were analysed in the light of Selz's theory, is significant. His subjects were children of an average age of 13 years, and his analysis was based on objective records checked by questioning "a large number of children" although he does not give any details about this questioning.

Types of errors found\textsuperscript{15} may be summarized in the following way:

\begin{itemize}
\item \textsuperscript{13}See Wilcocks, Op. Cit. p29.
\item \textsuperscript{14}Fortes M. "A Study of Cognitive Error," Brit. J. Educ. Psych., 1932, 2, 297-318. From the description of the tests it appears that these were similar to the Raven Matrices.
\item \textsuperscript{15}Ibid, p311ff
\end{itemize}
(1) Given figures cognized with insufficient determinateness, e.g.- educing an incidental relation.

(2) Reproductive effect of the actual test material, or perseveration, accounted for very few errors.

(3) Errors of subsumption - i.e. the given figure subsumed under a familiar concept, often leading to distortion of the figure by supplementation.

(4) Displacement of cognitive energy from a total situation to a part only - i.e. certain elements prepotent.

(5) Correct general inference made, but not tied to the specific context, or beginning and end points of the sequence of relations.

(6) Constitutive relations displaced by conjunctive or additive relations i.e. an easier or more familiar type of relation.

(7) Displacement of the critical relation in an eductive process by a contingent experience.

Fortes pointed out that his results support Selz in placing the greater weight on "displacement" as a source of error, rather than reproduction in Spearman's sense, and although he suggested that indeterminate cognition of the total perceptual situation was the antecedent condition for their occurrence, yet he realized that there are still more basic questions e.g.-
..is this a purely cognitive phenomenon, a problem of mental span or of the distribution of attention; or is it a conative phenomenon, a problem of the maintenance of a mental set or of volitional control of the cognitive exploration of a situation?...It may be conjectured that reproduction is resorted to in perceptual situations which are capable of evoking associative reproduction when, for conative or cognitive reasons, the subject fails to apprehend the relations constituting the situation. It would thus be a sort of regression to a more primitive level of coping with cognitive situations.16

In an extensive study based on items used in the Scholastic Aptitude tests of the College Entrance Examination Board, Brigham17 explored various types of test material and methods for handling results, in order to throw some light on the possibility of a systematic study of solutions as a didactic device. One of his major findings was that errors are merely "characteristic solutions of the problems set,"18 and he rejected Spearman's theory of error as merely a labelling of the one process noogenesis if the solution happens to be right, and reproduction if it happens to be wrong.- "In any given case of error the assignment of a certain portion to education and the rest to reproduction will merely generate quibbling"19.

16Ibid, p317
17Brigham C. A Study of Error, 1932, (N.Y. College Entrance Examination Board).
18Ibid, p84.
19Ibid, p80.
As a result of a study of solutions to synonym tests he pointed out that although the problems may be presented, and the answers recorded simply, yet the psychological structure of the situations may be exceedingly complex. The displacement of a problem from a narrow to a much broader (and mainly irrelevant) context is illustrated in his observation that the group refused to accept premises as printed, as true, but judged them "in the wider universe of experience"\(^{20}\). On the whole his attitude is summed up in the statement "...our studies have indicated error to be not capricious, but orderly"\(^{21}\).

From these few studies in which direct reference has been made to the two systematic studies, it is apparent that Spearman's theory of error, although accounting for a part of the error producing process, is inadequate, while Selz's theory in adding the important principles of task displacement and part efficiency, has proved useful in drawing attention to the perception of the problem by the subject, and in providing a useful set of concepts for analysing answers which to the experimenter may appear capricious but which from the point of view of the subject, are orderly.

\(^{20}\) Ibid, p194.

\(^{21}\) Ibid, p195.
III. - THE GESTALT ACCOUNT OF ERROR.

Although the assumption is made throughout Gestalt writings, that failure to structure, reorganize, etc. may lead to error, yet there are a few specific references to error worth noting briefly.

Kohler[^22] described three types of errors observed in his experiments with chimpanzees - (1) "Good" errors - i.e. the observation of the animal's behaviour, rather than pre-occupation with human achievements, will sometimes show an error to be quite reasonable in the circumstances. (2) Errors due to a complete lack of comprehension of the conditions of the task. (3) "Crude stupidities arising from habit in situations which the animal ought to be able to survey" - habits which are not accidental, but the "after effects" of previous solutions. He mentions drowsiness, exhaustion, colds or excitement, as conditions favouring this type of error[^23]. These three types of error fit in well with some of the types listed by Selz and Fortes, and point again to the significance of understanding the nature of the task and judging the relevance of habitual methods of solution.


[^23]: Ibid, p195, Cf, with Selz i.e. influence of distraction, restlessness, lack of clarity.
Duncker\textsuperscript{24} summed up this general point of view - "A solution conceived without functional understanding often betrays itself through non-sensical errors" - and in the analyses of some of his subjects protocols, reported errors due to attempts to escape from the original setting of the problem, "single track" proposals which leave out of consideration all other factors, and confusion of the original problem with another for which the correct solution is known.\textsuperscript{25} Similar sources of error were noted by Wertheimer,\textsuperscript{26} i.e. inadequate view of the whole situation; lack of breadth of view - "Even when he has it at the beginning he may lose it in the process because he is busy with details"; overextended view; premature or short cut closure - "Often the impatient desire to find the solution focuses the eye too exclusively, too strongly." (Cf. Selz - hurried approach).

Errors arising from the "blind" transfer of habits, pointed out by Kohler and demonstrated by Wertheimer\textsuperscript{27} in

\begin{itemize}
\item \textsuperscript{25}Ibid, Footnotes 17, 18 and 20, P15.
\item \textsuperscript{26}Wertheimer M. \textit{Productive Thinking}, 1945, (N.Y. Harper) p195, Cf. Fortes type 5 errors.
\item \textsuperscript{27}Ibid, p141ff.
\end{itemize}
his experiments on the area of a parallelogram, have been
described in more detail by Maier and Luchins as "habitual
direction" and "mechanization", the one resulting from the
typical experiences of the individual with the materials
of the problem, the other being created by the particular
mode of presentation of the materials.28

Some objections to the Gestalt interpretations are,
that they are altogether too subjective and even speculative,
that they lack the rigour of the more systematic theories of
learning, for example, and that they are descriptive, not ex-
planatory. These criticisms may be justified, yet Gestalt
concepts do appear to give more immediate meaning, however

28See discussion in previous chapter on Maier and
Luchins. That Maier's conception of "direction" may be an
oversimplification has been suggested by Weaver and Madden,
who repeated one of Maier's crucial experiments with somewhat
different results (Weaver H.E. and Madden E.H. "'Direction'
Furthermore, the same applies to Luchins' interpretation of
mechanization, for as Goodstein has pointed out -(Goodstein
Soc. Psych, 1953, 48, p351)"...it is questionable
that a shift from the old, set solution, to an alternate
solution represents any real gain. Why should the sub-
ject want to change his problem solving technique? The
previous technique was adequate; the directions in this
study and all similar studies were to find the solution,
not find the most direct solution, or to find more than one
solution, the rigid and non rigid solutions were still
correct."
subjectively (or perhaps because of this subjectivity), than do concepts drawn from learning theory. This may be inevitable at the present stage of psychological knowledge, for learning theorists, such as Hull or Skinner, have been interested in observing behaviour in strictly controlled experimental situations. Because of experimental isolation, they do not so readily yield to pressure from those, such as teachers, who want concepts that they can transfer immediately to everyday learning situations. On the other hand, investigators like Selz, Duncker and Wertheimer have observed, in a much less rigorous way, typical samples of everyday behaviour, in the form of human problem situations. This characteristic explains why the various analyses of error presented above (and those presented in the next section) are often expressed, and interpreted in concepts drawn from Gestalt literature.

As examples of the way in which more rigorous analyses may develop, two approaches may be briefly described. Firstly that of Ammons\(^{29}\) who gave an example of part of a postulate system which could define error in terms of properties of stimuli and responses; the particular sample

presented, being mainly to demonstrate increase in errors as a result of a high drive. His short exposition included a description of a type of experiment in which drive was measured in terms of inclusion of hard questions among easy ones, and errors in terms of erasures, crossouts, write overs and mis-spellings. It is difficult to assess the significance of Ammons error theory until it is developed more fully. The other approach is illustrated briefly in Miller's work on language and communication\textsuperscript{30}, in which he discussed errors in terms of communication theory (i.e. "noise") and linked them to linguistic change, e.g. "analogic" change - "changes in the direction of greater regularity or systematization". This type of change of material has already been illustrated\textsuperscript{31} and is referred to many times in the errors described below.

IV. - CLASSIFICATION OF ERRORS FOUND IN THE LITERATURE ON PROBLEM SOLVING.

The following classification is by no means rigid, and in fact incorporates overlapping groups; although the errors reported come from many different sources, with different age


\textsuperscript{31}See Selz and Fortes above. The same type of change has been noted in the experiments described by Bartlett in his work on remembering. (Bartlett F.C. Remembering, 1932, Cambridge University Press).
groups and different materials, in the interests of clarity they have been reduced to a few descriptive, although not logically distinct, categories. Frequency with which certain types of errors are mentioned cannot be taken as an indication of the significance of the error in problem solving; it is probable that certain types of error are more easily noted or described than other no less significant errors, furthermore it should be borne in mind that in most of the studies mentioned, there was little direct exploration of error (an exception is the study by Bloom and Broder). Despite the arbitrary nature of this classification, it does at least help to fill the gap between theoretical accounts and direct evidence, and has the advantage of arising from many diverse, yet first hand observations of the course of problem solving. 32

32 There is a rather obvious condition which may underly incorrect responses - namely lack of factual information, and this has not been included in any of the groups, for logically it should lead to failure to give any response at all. Where a response is given, it is dangerous to assume that it is entirely random, because the field of possible random responses is wide, so that where lack of factual information is the "predisposing" condition of failure, there is still the problem of describing the "precipitating" condition which leads to a particular wrong response being given at all.
(a) **Inadequate formulation of the problem.**

This may take several forms and has been described in somewhat similar terms by different writers. Piaget\(^{33}\) described one type of inadequate formulation when he pointed out that a child's mistakes may not be actual fallacies - mistakes in reasoning - but the result of "a faulty point of view, due to the fact that he has not yet asked himself the question as we ask it," i.e. there is some formulation of the problem, but not an adequate formulation. Other reports of the occurrence of this type of error are:

Chrisof\(^{34}\), failure to formulate the problem, or an "inability to penetrate the occasion"; Eagleson\(^{35}\), insufficient knowledge of directions; Smith and Tyler\(^{36}\), failure to study the description of the problem fully enough to appreciate all the limiting factors; Bloom and Broder\(^{37}\), confusion about the requirements of the problem; Burak\(^{38}\), incomplete formu-

---


lation of the task; Jones\(^39\), failure to see the problem clearly or to consider all its aspects; Doyle\(^40\), failure fully to understand the requirements of the problem.

Errors of this type exemplify what Selz emphasised, i.e. the response is incorrect because the task is not clearly structured. They also support the distinction between eduction of relations (or correlates) in general, and relevant eduction of relations in a particular context.

(b) **Distortion or partial operation of the task.**

Although this is really a refinement of the first category, rather than a completely different type of process, yet it is worth presenting separately as it includes some interesting phenomena. Piaget may again be quoted as illustrating the essential characteristics\(^41\) -

"Syncretistic understanding consists precisely in this, that the whole is understood before the parts are analysed, and that understanding of the details takes place rightly or


\(^40\)Doyle R.M. *An Investigation Into the Problem Solving Habits of a Selected Group of Boys, Together with an Attempt at Remedial Training*, 1955, Unpublished M.A. Thesis. (Canterbury University College Library) p72.

wrongly, only as a function of the general schema". This schema, he wrote elsewhere\(^{42}\) is "unconsciously applied to all the objects that will more or less fit into it".

One of the most penetrating accounts of the way in which such a schema operates has been given by Bartlett in his studies of perceiving and remembering\(^{43}\), and his observations on levelling, sharpening, conventionalizing, emphasising etc. have been confirmed in studies of problem solving.

Maslow\(^{44}\), has pointed out the tendency to place a present problem in a certain category of problems and then give a solution typical for that category. He used the word "rubricizing" to refer to a "partial, token or nominal response"\(^{45}\) and suggested that events (including problem situations) are not attended to in their own right, but in an abstracted, categorized or rubricized form. This same

\(^{42}\)Op. Cit. 1924, p133.

\(^{43}\)Bartlett, Op. Cit. It is surprising to find how little attention has been paid to Bartlett's work, especially by American writers, and interesting to find many developments which substantiate his results, not only in detail, but in his broader interpretations.


\(^{45}\)Ibid, p23.
point has been emphasised by Krech and Crutchfield\textsuperscript{46} who refer to the levelling and sharpening of perceptions in order to create a more stable organization. Efforts to produce stability in a schema may lead to the attending to only a part of the situation, a characteristic included in the first category of errors already described, and noted specifically by Bloom and Broder\textsuperscript{47} who found a tendency among their students to select an answer to a problem on the basis of only a very few clues. Although this may be due to "rubricizing" - i.e. to reading the question only partially and then presuming that it is a particular type of question, without checking the detail further, it may also be due to the fact that the first glance at the question could give the impression of difficulty, and lead, probably unconsciously, to a repression of the difficult parts.

Piaget\textsuperscript{48} noted a similar characteristic in children's language - e.g.- "He lets all the difficult words in a phrase slip by, then he connects the familiar words into a general schema". When this results in the logical relations between


\textsuperscript{47}Bloom and Broder, Op. Cit. p26ff.

the words being dropped, and a relation such as "this goes with" being substituted, then a type of "juxtaposition" in the face of the unfamiliar, occurs. Such a reaction to difficulty is very interesting and may result in the complete reshaping of the problem. Maslow\textsuperscript{49} noted this reshaping of novel problems, while Bloom and Broder\textsuperscript{50} reported that in complex situations the students substituted one or more questions or sub-questions which they could manage, instead of the given question. A very good description of this tendency has been given by Hildreth\textsuperscript{51} - i.e. "the difficulty reduction tendency," and her illustrations drawn from observations of children's perception, reading and thinking are very relevant. She noted that problems were simplified or allied to familiar one, the guiding principle being the effort after meaning - "The subject remakes the problem so that it will be a meaningful whole for him."

A further influence of difficulty has been reported by Luchins and Luchins\textsuperscript{53} in their experiments in which subjects traced mazes by mirror. They pointed out that the

\textsuperscript{49}Maslow, Op. Cit. p32.
\textsuperscript{50}Bloom and Broder, Op. Cit. p50.
\textsuperscript{52}\textit{Ibid}, p311.
introduction of the difficulty factor (i.e. use of mirror) resulted in "the development of narrow strip like cognitive maps as opposed to broader, more comprehensive maps". It is apparent that the combined effects of "rubricized" attention and difficulty, may result in any one of a number of operations on the problem situation, so that the actual problem the subject sets about answering may be more general, more specialized, simplified, distorted or only partially interpreted.\textsuperscript{54}

(c) \textit{Dislocation of the elements in a problem.}

Thorndike\textsuperscript{55}, in a brief analysis of errors in reading relevant to the psychology of thinking, used the terms "dislocation or disrelation of elements," and this serves as a convenient label for certain types of errors reported, in which the failure seems to be due to an inadequate analysis and weighting of the elements in a problem. A contrast between those using an analytic approach and those using an

\textsuperscript{54}That this is not just a characteristic of low ability is borne out by observations of the way university students at all levels sometimes misinterpret difficult or unfamiliar questions. Leaving aside those who are consciously suiting themselves, there is still a number who seem to be unconscious of what they are doing and are surprised when it is later pointed out to them. An interpretation in terms of Freudian repression would seem to be appropriate in these cases.

"intuitional" approach characterized by much use of previously established associations, was made by Chant, errors resulting mainly from the intuitional approach. Symonds also mentioned errors due to failure in analysis, while Burak found that failure to analyse out major variables was one of the important criteria of unsuccessful solution. In situations which were arranged to contrast personal and impersonal involvement in problem solving, Marks, and Marks and Ramond reported that the awareness of the given elements in the problem, or the capacity for analysing out the variables involved was one of the most important factors in successful solution.

A condition which influences the efficiency of analysis is speed, and this was specifically listed as a cause of error by Eagleson i.e. hasty activity without careful

deliberation; by Bloom and Broder\textsuperscript{61} - little time given to the consideration of the problem; and by Doyle\textsuperscript{62} who noted that some of the boys studied tended to be impulsive or hasty in their approach to the problem.

One of the causes of dislocation in analysis is the failure to give certain elements their proper weighting - a characteristic described by Thorndike\textsuperscript{63} as under potency or over potency of elements. This may, of course, vary from a hasty emphasis on one element because of familiarity, to a considered choice on reasonable grounds. The tendency to pick on a familiar or interesting element in the problem situation was noted by Welch and Long\textsuperscript{64} who reported errors due to the neglecting of the relationships involved because of attraction to the particular objects of the problem. On the other hand, Maslow\textsuperscript{65} has pointed to the potency of the unusual and unfamiliar in distorting analysis, this pre-potency of certain elements leading to the overlooking of


details which are necessary for solution - a characteristic observed by Doyle.66

These examples of prepotency of elements in the presentation of a problem can be supplemented by the well known experiments of Woodworth and Sells67 on atmosphere effect in syllogistic type problems, e.g. the global effect of "all positive" or "all negative" premises displaced the logical analysis of relationships. Morgan68 found that where subjects had to choose, on the basis of induction, certain letters and geometrical designs as significant factors in a problem situation, the irrelevant features of position and size of the elements kept on intruding and producing errors.69

69A theory has been outlined by Heidbreder in which such errors as those described by Morgan can be accounted for systematically. She wrote (Heidbreder E. "Toward a Dynamic Psychology of Cognition," Psych. Rev. 1945, 52, pl) "...the typical and dominant cognitive response in human beings is the perception of concrete objects...all other cognitive responses may be regarded as in some sense approximations to or modifications of this form". If this theory could be expanded through developmental studies of children's perception and thinking-(and in a series of experiments Welch and Long have made a beginning, while Templin has reported an experiment on the effects of deafness on children's reasoning- see Templin M.C. The Development of Reasoning in Children with Normal and Defective Hearing, 1950, Inst. Child Welfare, Mon.No.24), it could lead to some important educational implications, especially in the first years of schooling. Stern, for example, has built up a method of teaching arithmetic which although inspired by Wertheimer's point of view, is in fact a demonstration of the practical applications of Heidbreder's theory. (Stern C. Children Discover Arithmetic, 1949 (N.Y. Harper).
In a study of puzzle solving Hildreth\(^70\) found that some errors occurred with children working at assembling puzzles even though they had been given the chance to see the completed puzzle beforehand, and explained the failures as due to the prepotency of a bright foreground, and the neglect of the important details in the background. An unusual problem solving experiment by Simmel\(^71\) based on the detecting of an underweight coin among different numbers of sound coins by means of a minimum number of weighings, resulted in further evidence on the prepotency of elements as causing error. It was found that certain aspects intrinsic to the problem "snared" the subjects — these were symmetry, totality and divisibility. The two latter aspects were the most important — totality referring to "the overall undifferentiated impression of the task which the subject gains on first encountering it,"\(^72\) an impression which may inhibit analysis and lead to crude attempts to solve without getting down to the root of the problem. Divisibility referred to the more obvious properties of numbers, e.g. there is something "fourish" about 8, which may lead the subject into error


\(^{72}\)Op. Cit. p231.
until he can break down this obvious property, and think of 8 as made up of \(3 + 3 + 2\). This study of Simmel's stresses not only the prepotency of elements in attracting attention, but also, the effects of overlearning of certain ways of perceiving data, as an error producing factor.

(d) Influence of "set"\(^73\)

Overlearning may produce a "set," and apart from the major studies of Maier and Luchins, set or stereotyped response has been reported as a cause of error in a number of studies. Symonds\(^74\) mentioned set or fixation as a cause of error; Chrisof\(^75\) reported inhibition of the clear meaning of the task because of a particular set; Eagleson\(^76\) included "erroneous set" as one of the six major types of error noted; Maslow\(^77\) referred to errors caused by stereotyped or rote habits; Staudt\(^78\) found perseveration as a significant error.

\(^73\) Strictly speaking, the word "set" should no longer be used, because of a looseness of definition, clearly brought out by Gibson (Gibson J.J. "A Critical Review of the Concept of Set in Contemporary Experimental Psychology," \textit{Psych. Bull.}, 1941, 38, 781-817.)


\(^75\) Chrisof, Op. Cit. p182.


\(^78\) Staudt V.M. "The Relationship of Certain Personality Traits to Errors and the Correct Responses in Several Types of Tasks Among College Women Under Varying Test Conditions," \textit{J. Psych.} 1949, 27, p477.
producing factor; Bloom and Broder\textsuperscript{79} noted failures due to an inability to see a particular idea in a different function and inability to change a relationship once it had been established (Cf. Duncker's "embeddedness"); Fattu et al\textsuperscript{80} found that the best criterion for distinguishing the poor from the good problem solvers was "stereotypy of behaviour".

As noted earlier, the word "set" may cover a wide variety of conditions, and may arise in a variety of ways, yet the essential characteristic in the above studies is the persistence of responses or methods of handling the data, when these are no longer appropriate. In a way this is another example of "difficulty reduction," although it must not be overlooked that the use of sets in problem solving is analogus to the use of concepts in thinking (in fact a "set" may be classed as a type of "concept"). What is to be avoided is not "set" as such, for economy of method depends upon some sort of "set", but indiscriminate transfer of methods and assumptions from one situation to another. This interpretation covers such errors as those described by Jones\textsuperscript{81} as


the transfer of the wrong principle for the type of problem being worked, while the converse of this condition was noted by Smith and Tyler\textsuperscript{32} as failure to see the applicability of a known principle, and by Bloom and Broder\textsuperscript{33} as failure to find an effective technique for starting the problem.

Another condition which is related to this general type of error is the weakening or loss of the "task set", i.e. losing track of the problem after having started out in the right direction. Bloom and Broder\textsuperscript{34} described this as losing sight of the plan of the problem; Doyle\textsuperscript{35} noted a tendency for the subjects to become sidetracked, and for the problem to be worked correctly except for the expressing of the answer in the actual manner required by the instructions. This latter type of error was also reported by Eagleson\textsuperscript{36}.

(e) Influence of Attitude, emotional involvement and personal preference.

In a study of the relation of personality traits to errors on tests such as verbal analogies, arithmetic and

\textsuperscript{32}Smith and Tyler, Op. Cit. p84.
\textsuperscript{34}Op. Cit. p30.
\textsuperscript{35}Doyle, Op. Cit. p72.
\textsuperscript{36}Eagleson, Op. Cit. p268, Cf. also with Fortes Categ. 5 errors.
cancellation, Staudt\(^{87}\) concluded that "the more maladjusted the individual...the more errors she makes in material of the habitual and more complicated perceptual tasks". Related findings came from the experiments of Marks and Marks and Ramond\(^{88}\) which indicated that when problems are presented in an impersonal situation, there are likely to be more solutions than when the same problems are presented in a more personal or simulated "real life" situation. The greater emotional involvement of the real life situation resulted in an inhibiting of the subject's awareness of all the elements in the problem. Although this may not be the only interpretation of the results obtained, it is certainly a very plausible one.

Combs\(^{89}\) developed the thesis that the effectiveness of an individual's behaviour - i.e. intelligence, will depend upon the adequacy of his perceptions "in his phenomenal field at the moment of action". Among the factors which limit perception, he mentioned three relevant to this particular section\(^{90}\) - (1) the effects of goals - i.e. "It is conceivable that low intelligence may be at least in part, no more


\(^{88}\)Marks, and Marks and Ramond, Op. Cit. These experiments are of interest for the method used, as well as for the results.


than a function of the goals an individual is striving to reach in achieving his need satisfaction"; (2) the self concept, i.e. a child's conception of his abilities may severely limit his achievements - "...the conception one holds of himself is a vital factor in determining the richness and the variety of perception selected"; (3) threat - this may result in the narrowing of the field, and this "causes the individual to protect and cling to the perceptions he already holds". These conditions are of relevance in problem situations, because a child's reaction to a problem which appears difficult to him is likely to involve all these factors.

Supporting evidence comes from Bloom and Broder⁹¹ who found that the format of the problem was sufficient to discourage students from attempting any attack, while some adopted an all-or-none attitude, i.e. "either I can do this type of problem or I cannot do it". Doyle also found errors which were due to lack of confidence rather than failure to understand the problem,⁹² while Eagleson⁹³ reported a situation which is commonly recognised in school work but which is per-

haps too often unrecognised in testing, i.e. lack of interest.

The operation of a related condition - strong personal attitudes or prejudices - has been noted by Bloom and Broder, while Vinacke has summarized studies by Morgan, and Morton and Lefford in which errors on syllogistic problems were due to personal convictions and involvement in emotionally toned items. Brigham reported this type of error and summed up the situation concisely - "The conclusion is judged in relation to that larger universe of premises which constitutes the individual's experience...we seem to be dealing not only with the situation of the printed page but also with a larger universe supplied by the individual tested."

V. SOME SPECIFIC ERRORS IN INTELLIGENCE TEST ITEMS.

The five categories do not cover all sources of error, but at least they indicate some of the processes which operate when a subject is confronted with a problem or test situation and produces an incorrect answer. Unfortunately the few studies which report on errors in intelligence test items, apart from those of Brigham and Fortes, have not yielded much


information relevant to the arguments developed above. For example Wile and Davis\textsuperscript{97} studied failures on the Stanford Binet, of children referred to a psychological clinic. They found in their analysis of failures that "inadequacies in visual percepts and memory, with pattern reversals, supplement poor visual span and at least hint that the visual inadequacies for form and design, recognition, copying and reproduction, play a definite part in causing confusion in the learning process". Feifel\textsuperscript{98} analysed children's errors on the Stanford-Binet vocabulary test, although he was mainly concerned with changes in types of errors with age. He reported that there are no sharply demarcated patterns of error characterizing stages in development, the difference being that "more errors are made by the younger children seemingly because they 'know' less than the older ones". 75\% of all the errors made were omissions, and 20-24\% were described as "wrong definition".

An extensive study directed towards the specific hypothesis that differences in socio-economic status would influence the type of answers given to intelligence test


items, has been made by Eells et al.\textsuperscript{99}. In brief the main finding was that practically all the items showing unusually large status differences were verbal in nature and reflected the richer cultural background of the higher socio-economic group, as contrasted with the less sophisticated background of the lower socio-economic group. There were, however, other points worth noting: (1) that there was "some tendency for pupils in both status groups to check one of the first two distractors of a five choice question more frequently than any of the last three"; (2) where there were differences in the patterns of wrong choice, these were interpreted in terms of differential "opportunity for familiarity with certain objects, words or processes"\textsuperscript{100}; (3) on a number of items the high status group tended to concentrate errors on distractors which were classified as next best or near correct responses, distractors with associational value, distinctively unfamiliar distractors or those which look or sound like the given word, while the low status group tended to spread wrong responses over all the distractors more uniformly\textsuperscript{101}.


\textsuperscript{100}Op. Cit. p55

\textsuperscript{101}Op. Cit. p302.
Hamid\textsuperscript{102}, in an analysis of 'good' and 'bad' test items (in terms of prediction), reported one characteristic which has some relation to Eells' observations on distractor choice, namely that while the bright subjects discovered the correct answer wherever it occurred, the dull subjects tended to choose the first of the alternatives "if it appears to have the slightest relation to the subject of the test, as if they desire to terminate the emotional tension produced by the attitude of doubt".

As mentioned earlier, Brigham pointed out that errors are characteristic responses, and that "even with problems having large numbers of probable answers the distribution of erroneous answers concentrates at certain definite regions of the predicted range"\textsuperscript{103}. Fortes\textsuperscript{104} also noted that inspection of the distribution of error frequencies over three broad I.Q. groups showed that the relative error frequencies of the various distractors were approximately the same for the different groups. He also found, like Hamid and Eells, that the first distractor position was the one most often chosen.

\begin{itemize}
\item \textsuperscript{102}Hamid S.A. "Some Factors of Effectiveness in Mental (Intelligence) Tests," \textit{Brit. J. Psych.} 1925, XVI, p103.
\item \textsuperscript{103}Brigham, \textit{Op. Cit.} p306.
\item \textsuperscript{104}Fortes, \textit{Op. Cit.} pp309-310.
\end{itemize}
VI. - FACTORIAL STUDIES OF CAREFULNESS.

A further source of error which may be significant is suggested by the identification of a "carefulness" factor in a few factorial studies. Thurstone\textsuperscript{105} pointed out the possibility of a factor concerned with accuracy or caution as a result of loadings on four ratio scores - i.e. ratio of correct responses to total attempts. His suggestion "that the relative frequency of errors may represent a unique trait" was strengthened by Guilford and Lacey's analysis of wrong responses,\textsuperscript{106} in which they also found what they called "carefulness" - "which as presently conceived, is a trait of temperament"\textsuperscript{107}. Fruchter\textsuperscript{108} analysed separately correct responses and error scores and again isolated a factor interpreted as a carefulness factor. A further study, by Howie,\textsuperscript{109} resulted in the identification of the same factor, of which Howie wrote - "It is to be noted that for both sex groups it tends to be higher in the multiple choice tests. Without further research it is impossible to say more precisely what is implied by the description of carefulness or caution. Does it involve carefulness in the sense of checking over, or does it involve caution in the sense of 'canniness', not hazarding an answer when doubtful?"

\textsuperscript{105}Thurstone L.L. "The Perceptual Factor," \textit{Psychometrika} 1938, 3, p10.


\textsuperscript{107}Op. Cit. p695.


As Howie suggested, there is considerable difficulty in describing the psychological nature of this factor\textsuperscript{110}, for although it may be related to some such personality trait as persistence, or merely defined as a temperamental trait, yet it could also be the result of a "set" induced by the particular interpretation of the instructions by the child, or the work methods of children consciously or unconsciously developed by teachers. It is likely that the factor is a good deal more complex than its name suggests.

\textbf{VII.- CONCLUSIONS.}

This chapter has covered a variety of studies of error, has shown that there is a substantial amount of evidence favouring the type of description introduced by Selz, and has listed a number of other sources of error ranging from ego-involvement to "atmosphere" effect. All these conditions or sources of error throw more light on the operation of the methodological components of intelligence, many of them, e.g.

\textsuperscript{110}Vernon deals briefly with this factor in his chapter on 'expressive movements' in personality assessment - Vernon P.E. Personality Tests and Assessments, 1953, (London: Methuen) pp66-67; while Himmelweit and Summerfield found that and "index of accuracy," i.e. ratio of the number of wrong responses to the total number of responses made, correlated significantly with a criterion of examination performance, although they included this "index" as a personality characteristic, (See Himmelweit H.T. and Summerfield A, "Student Selection - An Experimental Investigation: II," Br.J.Sociol. 1951, pp64 and 66.)
task displacement, partial efficiency, insufficient determinateness, subsumption, set, inadequate formulation, dislocation of elements, difficulty reduction - having been reported a number of times, although not always in the same terminology.

It does not seem unreasonable to suggest that these are the types of errors which can be directly related to such concepts as firmness of control, ordering of relations in the light of the task, selection, restriction, flexibility, closure and analysis i.e. those characteristics which give to the concept of intelligence some of the qualitative distinctions found in factorial interpretations and widely used in the fields of problem solving and reasoning.

To this extent, the problems involved in interpreting the operation of the methodological component of intelligence, as developed in the previous chapter, have been brought into focus, and the next step involves the investigation of a small segment of this very complex problem, through an analysis of errors made by children on intelligence test items.
SECTION B.

EXPERIMENTAL DESIGN AND ANALYSES
BASED ON ERROR SCORES.
CHAPTER III

THE NATURE OF THE INVESTIGATION.

Up to this point a broad theoretical position has been developed, a general method of investigation proposed and evidence presented to show that this method can be a fruitful procedure in clarifying some of the problems arising from the theory. In the light of these considerations it was decided that one of the most direct ways of investigating the processes which produce error, would be to take a representative intelligence test, analyse the errors from a number of angles, and note the extent to which these errors could be interpreted in terms of the conditions described in the two previous chapters or any other conditions which might be relevant.

The first task, was to choose a suitable test, then decide on the selection of subjects to take the test. Following this it was necessary to work out ways of handling the data economically and to experiment with methods which could be used - thus a preliminary study was called for. Having cleared this ground it was possible to set out an experimental design and finalise the methods of analysis. All these steps have been described in this chapter.
I.- SELECTION OF TEST.

The most widely used intelligence test in New Zealand is the Otis Self Administering Test, Intermediate examination, this test being given to a large number of Primary School children, with the resulting I.Q. being entered, as a rule, on each child's school record card. In many cases the Otis I.Q. is one of the determining factors in the placing of children in ability and attainment streams in Intermediate schools, and is often referred to by teachers when discussing the capacities of various pupils. It is thus a test which has played, and is in fact still playing, an important part in educational guidance, and this reason alone was sufficient to justify an investigation into the type and origin of errors which children make on this test.

This reason was strengthened by the fact that the Otis is made up of a number of different types of items which are traditionally (although not necessarily factorially) associated with processes such as induction, deduction, concept formation, elucidating relations and correlates, classifying, and verbal comprehension. With such a wide variety of types of items, there should be scope for the interplay of many facets of problem solving and reasoning processes.

An added reason, was the fact that the Otis test is typical of the many group verbal tests which have been used
in both educational and experimental situations. Furthermore, the reasons for the particular score, or the particular items correct or incorrect, have been formulated from experimental analyses using the total score, rather than from studies of the interaction of subject and particular test item.

For these reasons the Otis test was selected as an appropriate source of material for investigation, the particular edition used being the one which was adapted and standardized by the New Zealand Council for Educational Research.1 The Otis Intermediate examination is suitable for children in classes ranging from Standard 3 to Form 2, (although there may be a ceiling effect with older and brighter children e.g. a 13 year old2) and is available in parallel forms; the form used in this investigation was Form B. Four sample questions are included on the front page and the instructions read:—

The test contains 75 questions. You are not expected to be able to answer all of them, but do the best you can. You will be allowed half an hour after the examiner tells you to begin. Try to get as many right as possible. Be careful not to go so fast that you


2Ibid, p91, p94.
make mistakes. Do not spend too much time on any one question. No questions about the test will be answered by the examiner after the test begins...

II. - SELECTION OF SUBJECTS.

Various criteria were set up to govern the choice of a sample from the school population:-

(a) The children should be old enough to cover most of the items, thus giving scope for the production of as many errors as possible. On the other hand they should not be at the top of the age and class range as this would increase the likelihood of the "ceiling" effect being too widespread.

(b) They should be unfamiliar with the Otis or similar tests (although it is difficult to establish this fact in every case).

(c) They should be reasonably representative in regard to age, I.Q., socio-economic status of parents, and sex, for the particular class level chosen.

(d) The children selected should be accessible for testing and interviewing without too much disruption of class organization.

In the light of these criteria it appeared that the most satisfactory sample would be Form I (or Standard five) of an Intermediate School with a median chronological age of
about 12 years, a wide range of age, I.Q. and socio-economic status, and reasonable accessibility in terms both of school organization and facilities for testing and interviewing.

At the time when the investigation was first launched there were two fully functioning Intermediate schools in the city, one of which had a satisfactory testing programme whereby the first assistant teacher administered the Otis Intermediate test to the Standard four classes of the contributing schools in the November prior to their entry to the Intermediate school. As far as could be ascertained the children in these classes had not taken the Otis test previously, although one or two children may have had some experience with it in other schools or for some special reason.

This Intermediate School was chosen as likely to meet the sampling criteria as fully as possible, and as having available completed Otis test forms for almost all Form I children. (A few came from schools outside the district, and some children had missed the test through absence).

These tests had been administered (during a period of two days) by the first assistant teacher, who went to each of

---

3Ibid, p65, Table 14, giving the median chronological age of Standard five classes as 12 yr 1 month at 1st April. Intermediate schools in New Zealand normally take Form I and II children from several "contributing" schools, thus the number of children in Form I classes may be between two hundred and three hundred.

4Christchurch South Intermediate.
the four contributing schools - thus the time and the conditions of administration were similar for all classes.\(^5\) Marking had been done by a group of teachers, so that it was necessary to check both marking and calculation of I.Q.s before the results could be used for research purposes.

In order to check this sample of Form I classes against the criteria of (a) and (c) above, the following measures were calculated from all the Standard four test forms (\(N = 292\)).

(a) **Age** as at 18th-19th November, 1953.

\[
\begin{align*}
\text{Median} & = 11 \text{ years } 5 \text{ months} \\
\text{Q1} & = 11 \text{ years } 0 \text{ months} \\
\text{Q3} & = 12 \text{ years } 0 \text{ months} \\
\text{Range} & = 10 \text{ years } 3 \text{ months to } 13 \text{ years } 11 \text{ months}
\end{align*}
\]

It is necessary to add on 4 months to this median age in order to make it comparable to the Redmond and Davies median of 12 years 1 month for all Standard five classes (i.e. difference between 19th November and 1st April) - and with this correction the median becomes 11 years 9 months, 4 months below the median (in 1940) for all classes, rural as well as urban.

---

\(^5\)The particular teacher concerned (Mr. A. Coates), had had experience in administering the tests in previous years and was familiar with the necessity for adhering to instructions, careful timing, and standard conditions, thus there was no reason to believe that the test results were in any way influenced by varying, or inferior conditions of administration.
(b) I.Q. Mean = 104.3 S.D. = 12.6 Range = 71 to 143

Redmond and Davies did give a mean I.Q. for Intermediate Schools based on all age groups, although the mean I.Q. for the 12 year olds was 105.5 with a S.D. of 11.6 score points\(^6\) which is equivalent at this level to between 12 and 13 points of I.Q.

(c) Socio-economic status of parents. According to Havighurst\(^7\), the best single estimate of socio-economic status is occupational status, with a correlation coefficient of over .9 between the latter and a composite score of socio-economic status. As a composite score (taking account of occupational, house residential area, and source of income ratings), was considered too cumbersome for the purposes of this study, and in view of the fact that the occupational rating index gives such a good prediction of status, this single index was used. The information was gathered from details filled in by the child on the first page of the Otis, checked against the school records, and in doubtful cases checked with the children in question and their teachers.

\(^6\)Ibid, P73 Table 17, and P118 Table 40.

Ratings for the various occupations were given on the basis of the Congalton-Havighurst 7 point scale\textsuperscript{8}, the median rating for the whole sample being 5.19, i.e. "Office and Sales Workers" group of occupations (e.g. Government Office Clerk = 5.20; Book-keeper = 5.13; Stenographer = 5.26). Median ratings for the four contributing schools were as follows:- School A = 4.91, B = 5.22, C = 5.26, D = 5.40 - the order of these ratings reflecting the judgements often made about the status of the areas served by these schools.\textsuperscript{9}

(d) Sex. There were more boys than girls in the sample:-

\begin{center}
Boys, N = 168, \% of sample = 57.5  
Girls, N = 124, \% of sample = 42.5
\end{center}

From the above measures it can be seen that the sample selected was in reasonable accord with the criteria set up,


\textsuperscript{9}The percentage of the sample in each of the 7 ranks was (1) 0.0, (2) 0.8, (3) 2.7, (4) 14.8, (5) 22.0, (6) 49.0, (7) 10.7. Congalton did not provide any figures indicating the percentage in each of these ranks in the census returns for major cities, although elsewhere he gave percentages based on a small town electoral roll (N = 1,484):- (1) 4.1, (2) 5.4, (3) 6.5, (4) 13.3, (5) 24.9, (6) 21.7, (7) 24.1. (See Congalton A.A. "Social Grading in New Zealand," \textit{Brit. J. Sociol.} 1953, 4, p50). Even allowing for the difference between a small rural town and a major city, it is apparent that this sample draws more heavily on the lower status groups and has a poor representation of the first three groups.
except that it was rather too heavily weighted with boys and had a disproportionate number in the lower status ranks.

III.- PRELIMINARY STUDY.

In order to try out various ways of handling the item errors, a preliminary analysis was made of the 292 test papers with a view to clarifying the following points:-

(a) Classification of Otis items into groups of similar items.
(b) Method of expressing error for a given item within a group.
(c) Type of distribution of such errors within homogenous I.Q. groups.
(d) Overall measures of error, or omission, or both, for each test paper.
(e) Correlations of such measures with I.Q.
(f) Method of recording (a), (b) and (d) for easy reference.

(a) Classification of Otis Items.*

The 75 items are arranged in order of difficulty, not in sub tests of items all of one type, thus the child changes from one type of item to another, with certain types intermittently recurrent. These "recurrent" items were grouped into eleven categories, some categories being more homogeneous in item structure than others - (e.g. (3), (8) and (11) contain two or more types).

(1) Word Meaning - 5, 16, 23, 40, 44, 54, 61, 75.

- An event which is sure to happen is said to be?
  (1) probable, (2) certain, (3) doubtful, (4) possible, (5) delayed.

* A Specimen Otis Test is included after the last Appendix.
(2) **Opposites - 3,8,24,63.**

   e.g.- Which one of the five words below means the opposite of difficult?
   (1) hard, (2) quick, (3) soft, (4) easy, (5) common.

(3) **Word or Sentence Attribute or Definition - 2,14,18,27,31,35,52,66.**

   e.g.- What tells best just what a horse is?
   (1) It has a tail, (2) a live thing, (3) a thing that works and eats, (4) a large, four legged animal, (5) something to pull a wagon.

   e.g.- What is the most important reason that motor cars have displaced horses and carriages?
   (1) Horses were getting scarce, (2) Horses often run away, (3) Motors save time and "time is money," (4) Motors are cheaper than carriages, (5) Motors cost less to repair than carriages.

(4) **Forming a Concept from given Examples - 1,9,25,41,58.**

   e.g.- Which one of the five things below is most like these three: apple, peach, pear?
   (1) seed, (2) tree, (3) plum, (4) juice, (5) peel.

(5) **Mixed Sentences - 12,28,32,49,53,60.**

   e.g. Do what this mixed up sentence tells you to do. letter Write the the in A brackets.

(6) **Classifying and Arranging Elements of a Concept in a Given Order. - 6,10,20,29,43,55,64,69,73.**

   e.g.- Which one of the five things below is the largest?
   (1) bud, (2) branch, (3) tree, (4) twig, (5) limb.

(7) **Analogies - 4,7,11,15,19,22,26,30,34,38,42,46,50.**

   e.g. Hat is to head as thimble is to what?
   (1) finger, (2) needle, (3) thread, (4) hand, (5) sewing.
(8) **Verbal Reasoning** - 36,37,45,48,51,57,62,68.

*Example:* If George is taller than Frank, and Frank is taller than James, then George is (?) James.
(1) taller than, (2) shorter than, (3) just as tall as, (4) (cannot say which).

*Example:* In a foreign language, boy = Kolo
good boy = Kolo Daak
The word that means  **good** begins with what letter?

(9) **Proverbs** - 33,59,65,70

*Example:* There is a saying "A stitch in time saves nine." This means?
(1) A little sewing may save nine shillings, (2) It pays to attend to troubles before they get worse.
(3) Work hard and save as much as you can. (4) You can save time by sewing.

(10) **Number Series** - 17,21,47,67,74.

*Example:* One number is wrong in the following series. What should that number be?
1 2 4 8 12 32 64.

(11) **Arithmetical Reasoning** - 13,39,56,71,72.

*Example:* If a man has walked east from his home 7 miles and then walked west 4 miles, how many miles is he from home?

*Example:* If I have a large box with 2 small boxes in it and 5 very small boxes in each small box, how many boxes are there in all?

Although this classification is superficial, and may not bear much relation to different types of abilities, or to the apparent processes involved in solving the items, yet it is probable that the particular item structure may have some bearing on the types of errors made, particularly in view of the errors described under (a) and (b) of IV in the previous chapter.
(b) **Method of expressing error for a given item within a group.**

All the errors made in a given test paper were listed in categories (as above), along with the omissions, and the combined total of errors and omissions thus provided a measure of failure in the various categories. By itself, this measure was misleading, because a total of 3 may indicate something different, if in one case it is 3 errors out of 3 attempted, and in another, 3 out of 6 attempted. In order to take account of this factor, all error totals were put in the form of a ratio, i.e. total error over number of items covered (in the particular category). A typical category error analysis illustrates a resulting pattern of error ratios:

**Girl, A.M. I.Q., 117.**

Categories:— (1) 4/7, (2) 1/4, (3) 3/7, (4) 2/5,

(5) 3/6, (6) 0/6, (7) 1/13, (8) 1/7,

(9) 1/2, (10) 2/3, (11) 2/3.

(It should be noted that "error ratio" includes errors and omissions)

The whole point of listing such error ratios was that if two children of approximately the same I.Q. have very different ratio patterns, and if an analysis of these patterns throws any light on the causes of these errors, then it may be possible to trace qualitative differences in the ways in which the two children approach or handle test
items.

(c) **Type of distribution of such errors within homogeneous I.Q. Groups.**

Having listed the errors in all the test papers, it was necessary to break up the total sample into groups with a reasonably small I.Q. range, and examine the possibility that within any one I.Q. group the error ratios on each category might show considerable variation.

Seven I.Q. groups were selected by taking 6 I.Q. points (i.e. approximately half the S.D. of 12.6) as a class interval and arranging the classes in such a way that the mid-point of the middle class interval fell as near as possible to the mean for the whole group (104.3) - the resulting distribution being as follows:

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISTRIBUTION OF CASES IN 7 I.Q. GROUPS.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I.Q. Group</th>
<th>Range</th>
<th>No. of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>120+</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>114-119</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>108-113</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>102-107</td>
<td>47</td>
</tr>
<tr>
<td>5</td>
<td>96-101</td>
<td>54</td>
</tr>
<tr>
<td>6</td>
<td>90-95</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>89-</td>
<td>34</td>
</tr>
</tbody>
</table>

**10**Because of the small numbers in the half S.D. intervals below 89 and above 120, these intervals were combined into one of 89 and below and one of 120 and above.
In order to compare the error ratios and put them into frequency distributions, each ratio was converted into a proportion of ten - e.g. 4/7 became 5.7/10; 1/13 became 0.8/10 - thus the distributions all had the same possible range of 0 to 10 and comparisons between categories were more straightforward. While this procedure reduced all the ratios to the same base, its greatest defect was in obscuring the number of items upon which the ratio was based, e.g. 2/4 became 5.0, 4/8 became 5.0 - an identical ratio arithmetically, but clearly different in terms of reliability.

Distributions in each of the seven I.Q. groups on each of the eleven categories for boys and girls separately, and together, were tabulated.¹¹ At this stage of preliminary analysis an inspection of the distributions confirmed the observations made while listing ratios on the test sheets - namely that even with a narrow I.Q. range, there were contrasting cases in regard to error ratio or degree of failure on items attempted or covered. Several examples illustrate these observations:

¹¹The means of the distributions are given in Appendix A.
Group 2 Category (1)
Error Ratio: 0 1.3 1.7 2.5 2.9 3.3 4.0 4.3 5.0 5.7 6.7
No.of Cases: 5 5 5 3 5 1 1 3 4 4 2

Group 2 Category (2)
Error Ratio: 0 1.3 1.7 2.0 2.9 3.3 3.8 4.0 4.3 5.0 5.7
No.of Cases: 11 4 3 2 6 1 5 1 1 3 1

Group 4 Category (4)
Error Ratio: 0 2.0 2.5 4.0 5.0 6.0 7.5 8.0
No.of Cases: 15 3 9 10 2 2 2 4

Group 6 Category (6)
Error Ratio: 0 1.7 2.0 2.5 3.3 4.0 5.0 6.0 6.3 6.7 8.0 8.6
No.of Cases: 3 1 2 4 3 4 1 6 2 2 1 1

Group 7 Category (7)
Error Ratio: 1.1 1.8 2.3 2.7 3.3 3.3 5.8 4.4 5.4 6.2 6.7 7.7
No.of Cases: 1 1 4 1 6 3 8 5 2 2 1

From one point of view there is nothing unusual about these distributions, for it would be most unlikely that all children, within a half S.D. range in I.Q. would have the same or even nearly the same ratio of failures to number attempted. Thus in Group 2 Category (1) for example, although 5 children made no errors and 4 children had error ratios of 5.7, i.e. just over half the number attempted or covered, these deviations from the mean, which in this case was 3.0, merely indicate what is already known, viz., that given 38 children and this range of error ratios, there are bound to be many factors operating which will cause a considerable spread (e.g. number of items attempted in relation
to age; difficulty of items in the category; unreliability).

There is, however, another point of view which may be equally legitimate, namely, that if two children of approximately the same I.Q. and overall error tendency, score very different error ratios on a given category, then while this may be due to the factors mentioned above, it may equally well be due to some difference in the way the two children approach this type of item, or react to its particular wording or structure. These differences may cancel out over a number of items of different types, and in terms of a quantitative estimate of performance this may be quite satisfactory, but in terms of a qualitative analysis of the particular process which results in the erroneous answer, the differences on any one item or group of items may be significant.

This preliminary analysis indicated that:— (a) The whole sample could be broken up into seven I.Q. groups, the middle five of which showed a variation of 6 I.Q. points within each group, while the number of cases ranged from 27 in Group 1, to 60 in Group 3. (b) Within each category and group there were variations in error ratio large enough to suggest further analyses in which I.Q., overall error tendency, and unreliability would need to be controlled.
(d) **Overall measures of errors or omissions or both, for each test paper.**

This analysis was undertaken in order to explore the possibility of finding a measure of failure, which would not be highly correlated with I.Q. or in other words a measure which would show considerable variation within each I.Q. group. Underlying this attempt was the assumption that where children of approximately the same I.Q. have very different measures of number of errors made or number of omissions, then these differences may throw some light on the methods of approach to test situations. One child may race through the test covering 70 items and making 30 errors, while another may work more slowly covering 45 items and making only 5 errors. This difference may be specific to the particular test, due to some temperamental factor, or to a method of working developed in classroom tests (e.g. mental arithmetic).

Four measures were calculated for each test paper:—

(a) Percentage of the total number of items (75) in the test actually covered - i.e. as measured by the last item attempted. This measure included correct, incorrect and omitted items, e.g. Girl A.M. last attempted item 63; score (correct items) 43, wrong 9, omitted 11; % covered = 84.

(b) Percentage of total items which were attempted - i.e. as measured by score + errors (but not omissions), e.g.-
A.M. 43 + 9 = 52/75 = 69%.  

(c) Percentage of the number covered (score + errors + omissions) which were omitted e.g. A.M. 11/63 = 17%.  

(d) Percentage of number attempted which were wrong, e.g. A.M. 9/52 = 17%.  

These four percentages were then tabulated for each of the seven I.Q. groups with the following results:

**TABLE II**  
**MEDIAN AND RANGE OF I.Q. GROUPS ON FOUR MEASURES.**

<table>
<thead>
<tr>
<th>I.Q. Groups</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Md</td>
<td>Range</td>
<td>N</td>
<td>Md</td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>94</td>
<td>77-100</td>
<td>91</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>84</td>
<td>68-100</td>
<td>82</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>78</td>
<td>62-100</td>
<td>76</td>
</tr>
<tr>
<td>4</td>
<td>47</td>
<td>77</td>
<td>59-100</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>54</td>
<td>75</td>
<td>56-100</td>
<td>72</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>70</td>
<td>38-100</td>
<td>65</td>
</tr>
<tr>
<td>7</td>
<td>34</td>
<td>66</td>
<td>41-100</td>
<td>59</td>
</tr>
<tr>
<td>All</td>
<td>292</td>
<td>77</td>
<td>38-100</td>
<td>74</td>
</tr>
</tbody>
</table>

The distributions for the whole sample are given in linear form below - (each I.Q. group distribution being of a similar general shape).
FIGURE (1)

DISTRIBUTION OF PERCENTAGES ON FOUR MEASURES.

From (a) and (b) above it can be seen that a considerable proportion of the children covered all or nearly all the items (13.7% for the interval 98-100), some of them attempting all or nearly all (6.2%). In both cases the upward trend is apparently due to the fact that the test has too few items for a group of fast workers, although it cannot be assumed that fast working in this case necessarily means accurate working, in fact the overlaps in the ranges given in Table II (a) and (b), show that in all I.Q. groups except one, some individual reached 100%.
The distribution for (c) shows clearly the great concentration of percentages around 0%-4%, with very few percentages greater than 20%, although from the ranges it is apparent that within any one I.Q. group there were great contrasts, e.g. - between one child with 0% omissions and another child in the same I.Q. group (e.g. 4) with 25%.

In view of this piling up of percentages around 0-4% this particular measure is not likely to be useful in discriminating between methods of approach to the test, except in a few exceptional cases.

A piling up of percentages at one end of the distribution does not occur in (d), this being the only distribution to approach normality. This may have been due to a close relationship between this measure and I.Q., a possibility indicated by the ranges for each I.Q. group, although the great overlap from group to group indicates some striking contrasts in individual cases, e.g. - a group 2 child had 37% of attempted items wrong, while a group 6 child had only 16%.

Contrasts mentioned here, and elsewhere, may depend on age differences,¹² so that if two children were of the

¹² Possibly on sex also, hence separate analyses for sex were made on (a), (c) and (d), with very little evidence for any differences, the medians for the seven I.Q. groups combined being - (a) boys 79% girls 76%, (c) boys 4% girls 4%, (d) boys 30% girls 33%.
same I.Q. but different ages, the younger child would have a lower score and could thus have more errors, assuming that both covered the same number of items. A rough check on this, was made by taking the 108-113 I.Q. group and splitting it into three smaller groups of 2 I.Q. points range, i.e. 108-109, 110-111, 112-113, while each of these three groups was split into nine age groups of three months each, i.e. the total age range being from 10 years 3 months to 12 years 5 months. Scores on the four measures were then tabulated for each of the groups. This procedure resulted in a number of small groups, homogeneous in regard to age and I.Q., and in the twelve groups where there were five or more cases the scatter was examined for evidence of individual contrasts. In (a) examples of the amount of the scatter were (in percentages):- 72-99, 63-96, 66-99; in (b):- 69-93, 63-96, 66-90; in (c):- 0-14, 0-6, 0-10; in (d):- 20-40, 10-40, 12-38. These figures are of little significance when considered in terms of the total sample, but they are relevant when a question is put such as "Why did this child make errors in 40% of the items he attempted, when this other child of the same age and I.Q. made only 10% errors?"
Correlations of measures with I.Q.

Of the four measures, (a), (b) and (d) had a wide enough distribution to merit further examination. Although as shown in the illustrations above, there were contrasting cases within each I.Q. group, yet the general trend in the percentages was in the same direction as the I.Q.s. This is indicated by Pearson Product - Moment correlation coefficients of:

- (a) and I.Q. = .46 ± .03
- (b) and I.Q. = .61 ± .02
- (d) and I.Q. = -.66 ± .02

It is evident that all these measures are influenced considerably by I.Q. and cannot therefore be used as independent measures except in groups of similar I.Q. The lower figure for (a) was probably due to the greater piling up of cases on 100%, and with a longer test it is likely that the distribution would even out, with the lower I.Q. groups dropping away from 100% and thus raising the correlation coefficient slightly.

Method of recording.

Various ways of recording the errors were tried and finally an error record was drawn up for each pupil in order to provide information which would be available for cross checking and for use in separate analyses using the same data. A completed record is illustrated in Appendix B.
IV.- EXPERIMENTAL DESIGN.

The preliminary study was useful in clearing the ground for further analyses of errors, yet its greatest defect lay in the fact that there was no guarantee that the types of errors made would be repeated, especially as some of the measures used - e.g. error ratio, were based on only a few items. In order to overcome this difficulty, the test was given again (i.e. the same Otis B) approximately seven months later. It was argued that if similar patterns of error appeared after an interval of this length, then the chances of purely accidental or irrelevant conditions being the cause of error, were very much reduced.

A further possibility was that the patterns of errors might be related to differences in abilities more specific than a general measure of intelligence could indicate, and in order to check this, it was necessary to administer a test with sub-scores for different abilities or group factors. A suitable test meeting these requirements is the S.R.A. edition of Thurstone's Primary Mental Abilities, Intermediate Form, for ages 11 to 17, and this test was

13 At the beginning of July, 1954; by this time the children had been in the Intermediate School five months. The test was given to the various groups in Form 1 over a period of three days.
given to all the children approximately a week after the second Otis.\textsuperscript{14}

Even though patterns of errors might be consistent from the first to the second Otis, they may or may not be of the same type on another test of similar design. To throw some light on this, the A.C.E.R. Intermediate D was given approximately five months after the second Otis. This test is a group verbal test very similar to the Otis Intermediate in form, and designed for children between the ages of 10 and 14 years.\textsuperscript{15}

One further measure related to "carefulness" and "carelessness"; this was a rating scale on which each teacher was asked to rate each of his children on a trait described as "carefullness" at one extreme and "carelessness" at the other. This rating was done in December, as was designed to throw some light on conditions which might influence test performance, especially in view of the fact that some such factor has appeared in factorial analyses of wrong responses.

\textsuperscript{14}Separate scores, (percentiles) are given for Verbal, Spatial, Reasoning, Number and Fluency, and a weighted average of these scores (Q) which gives an overall measure similar to I.Q. Tests were administered over a period of one week.

\textsuperscript{15}The test was given over a period of four days early in December. This and the other tests mentioned were administered by the writer.
From the records of errors it is possible to test certain hypotheses and to make some assumptions about the causes of error, yet there is always the possibility that the investigator may read into the figures an interpretation based on a theory of intelligence or the influence of some condition on intelligence. There is no simple way of knowing directly why a child chooses this answer instead of that, thereby making an error. Even if the child could introspect fully, there would still remain the probability that some of the essential processes leading up to the production of error, would not be recognised.16

As it is clearly impossible to obtain detailed introspections from children, and as it is also clear that an objective analysis of errors needs to be strengthened by some more direct evidence on the reasons for a particular choice, it was decided to take a representative group of children and to ask each one to work over aloud the items on which he had made errors, and to answer questions about his methods. This method of individual analysis was attempted with 84 children in November and December just

16Maier found that students did not remember and even denied receiving, hints which had initiated the processes leading to the solution - Maier N.R.F. "Reasoning in Humans II The Solution of a Problem and its Appearance in Consciousness," J. Comp. Psych. 1931, 12, pp181-194.
prior to the giving of the Intermediate D.

The complete programme of testing and analysis was spread over the period November, 1953 to December, 1954\(^\text{17}\) the particular order and timing of tests being determined partly by the requirements of the analysis, partly by the school organization - (e.g. examinations, availability of classes), and partly by the time available to the examiner. It was therefore not a rigidly determined experimental design, otherwise the Intermediate 'D' would have been administered soon after the P.M.A. tests. A summarized representation may clarify the procedure.

<table>
<thead>
<tr>
<th>Time</th>
<th>Test</th>
<th>Comments</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov. 1953</td>
<td>Otis I</td>
<td>Taken in the four contributing schools.</td>
<td></td>
</tr>
<tr>
<td>Nov.-June, 1954</td>
<td>Otis II</td>
<td>Preliminary Analysis etc.</td>
<td></td>
</tr>
<tr>
<td>Nov.-Dec., 1954</td>
<td>Int. D</td>
<td>Questioning of children on errors in Otis I and II.</td>
<td></td>
</tr>
<tr>
<td>Dec., 1954</td>
<td>Rating</td>
<td>Taken immediately following questioning.</td>
<td></td>
</tr>
<tr>
<td>Dec., 1954</td>
<td>Scale</td>
<td>Completed during questioning and Int. D.</td>
<td></td>
</tr>
</tbody>
</table>

\(^{17}\) Although the original sample was made up of 292 children, the number of children taking the various tests varied because of absences, change of school, cafeteria duties etc. Comparisons of the results of the tests do not bear directly on the work reported below but they may be of some value to other investigators in New Zealand and are accordingly reported in Appendix G.
V. METHODS OF ANALYSIS.

Although details of the methods of analysis are discussed along with the results, an overall description of the procedure may be of some value in specifying the types of questions which these methods and results should elucidate.

(a) Repetition of errors on Otis I and II.

The first question arises from the fact that the analyses of errors take into account performance on two Otis tests, the assumption being that where errors are repeated then the error producing process is due to something more than passing "accidental" factors. The question may be formulated thus: "To what extent did children make errors on the same items despite the fact that the tests were seven months apart? Furthermore, what percentage of these item errors were repetitions of exactly the same choice of distractor?"

(b) Consistency of errors in, and difficulty of categories.

As some distractors have a higher mean error ratio than others, it could be assumed that these categories contain questions which are more difficult than others. This concept of "difficulty" is, however, complex and may result in a circular argument depending on the criterion of difficulty. The question arises: "Is the pattern of mean
error ratios (i.e. a measure of difficulty) similar on Otis I and II? Is it similar to the predicted pattern, i.e. predicted in terms of the way the test was constructed?"

(c) **Comparison of "High Error" and "Low Error" groups over all categories.**

If some children score consistently high error ratios on a particular category, while other children of comparable I.Q. score consistently low error ratios, then the question arises as to the cause of the difference. It cannot be explained as a superiority in "general intelligence" as this is ruled out by the comparability of the two groups in I.Q. It may be due to any number of factors, and the questions which an analysis of high error vs low error groups should throw some light upon, may be such as: "Is the difference due to age, sex, socio-economic status? Is the difference on one category also apparent when the two groups are compared on a similar category? Is the difference still apparent on a similar category in another test altogether? Is the difference related to a special ability? Are the errors made by the two groups different in type as well as quantity?"

A further question, of a somewhat different order, arises from the criterion of "consistency" as used in this
analysis, namely - "Are there cases of marked inconsistency in error ratios from Otis I to Otis II. If so, how is this inconsistency related to I.Q.?'"

(d) Ratings on carefulness-carelessness in relation to error.

Where children have been rated as "very careful" or "very careless" in school work, assuming the rating to have some value, this trait may be reflected in the way in which they approach a test situation such as is provided by the Otis. Relevant questions are:- "How do children rated at the extremes of the scale, compare on I.Q., errors, omissions etc.? If any relation exists between the trait and Otis test performance, is this carried over to the P.M.A.s or the Intermediate D? Is there any relation between position on the scale and consistency from Otis I and II?"

(e) Wrong distractor choices in I.Q. groups over all Categories.

I.Q. groups may be differentiated not only in terms of total score, but also in terms of type of error made, and the type of error made on any given item or category, i.e. the particular wrong distractor choice, may give a clue to the process operating to produce error in that item.

Questions which may be asked are:- "Are there differences between the I.Q. groups in wrong distractor choices? If so, are they more apparent in certain categories rather
than in others? Do the wrong answers cluster around certain choices or are they spread over all the wrong distractors? How are wrong responses distributed on items where the child must produce the answer rather than choose from alternative answers?"

(f) **Questioning of children on Otis errors.**

Whatever the pertinency of the various questions noted in the preceding five analyses, and however interesting the attempted answers, there is no question as important as this one: "How did this child arrive at this particular error on this item, as suggested by his own explanation and by relevant questions? What is his attitude towards his own error?" Such a question may be partially answered by working over items with individual children, and the results from this analysis should check on, and clarify, the results of the more formal analyses.

This section, although the last, is the most important, and follows on logically from the analysis of errors in Chapter II. The other sections lead up to, and provide the background for this individual analysis.
CHAPTER IV

REPETITION OF ERRORS AND DIFFICULTY OF CATEGORIES.

I.- REPETITION OF ERRORS ON OTIS I AND II.

As the item errors and the distractor choices or wrong answers were recorded for each individual on Otis I and II, it was possible to make comparisons of repetitions. The first comparison was based on the percentage of errors and omissions on Otis II which were repetitions of errors or omissions on Otis I. No items in Otis II beyond the last item reached in Otis I were included, thus the percentage was based on those items common to both tests. An example may clarify this:

Girl I.Q. 108/110.

(1) Last item attempted, in Otis I = 62.
    in Otis II = 67.

(2) Ignore all items in Otis II after 62.
    Count wrongs and omissions = 15.

(3) Mark items wrong or omitted in Otis I
    which were also wrong or omitted in
    Otis II. e.g. -

    6 items wrong repeated in II.
    (5 with identical answers)

    5 items omitted in I
    (4 of these omitted in II)
    (1 attempted an wrong in II)

(4) Percentage repeat errors and omissions
    = 11/15 = 73% (A)
The second comparison was based on the percentage of identical errors in the items wrong in both Otis I and II. In the above example 6 items were wrong in both tests, and of these, 5 were due to identical errors \( = \frac{5}{6} = 83\% \) (B). Distributions of these two measures (A and B) are given below.

**TABLE III**

PERCENTAGE OF ITEM ERRORS AND OMISSIONS REPEATED (A), AND PERCENTAGE OF IDENTICAL ERRORS (B)

<table>
<thead>
<tr>
<th>Percentage</th>
<th>(A)</th>
<th>(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95-100</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>89-94</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>83-88</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>77-82</td>
<td>30</td>
<td>11</td>
</tr>
<tr>
<td>71-76</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>65-70</td>
<td>49</td>
<td>24</td>
</tr>
<tr>
<td>59-64</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>53-58</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td>47-52</td>
<td>10</td>
<td>39</td>
</tr>
<tr>
<td>41-46</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>35-40</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>29-34</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>23-28</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>17-22</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11-16</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5-10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0-4</td>
<td>2</td>
<td>9</td>
</tr>
</tbody>
</table>

\[ N^1 = 248 \quad 248 \]

(A)...\( M = 63.4\% \). S.D. = 17.7\%

(B)...\( M = 58.0\% \). S.D. = 23.2\%

Of the original 292 children, 44 were either not available or had gone to another school at the time of Otis II testing.
The concentration of cases at the two extremes of (B) is due mainly to the percentages based on only a few repeated error items - e.g. 0/2 or 3/3 etc. If these two extremes are ignored the M and S.D. are both reduced. Although the percentage of repeat errors and omissions is high (63.4%), it would have been even higher if allowance had been made for the fact that during the 7 months period between the tests the children had matured and some of the easier items originally wrong were bound to be right on the second testing.

The distributions give evidence that in a considerable number of cases the types of items wrong, and the specific errors made, were repeated from Otis I to Otis II, thus the assumption that there is some consistency in the processes underlying errors (and omissions) seems warranted.

II. - CONSISTENCY OF ERRORS IN, AND DIFFICULTY OF CATEGORIES.

From Appendix A which lists the mean error ratio scores for each I.Q. group over each category, it is possible to calculate a mean of these means for each category, thus providing a basis for comparison between the eleven categories. The same can be done with the table of error ratio means on Otis II (Appendix C).  

2This Table was based on 274 cases, i.e. all those who took Otis II. Of this number 243 took both tests, the remainder Otis II only. This means that the two Tables were not based on exactly the same sample, as Otis I included 44
TABLE IV
COMPARISON OF ERROR RATIO MEANS OVER II CATEGORIES.

<table>
<thead>
<tr>
<th>Category</th>
<th>Otis I</th>
<th>Otis II</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.3</td>
<td>4.1</td>
</tr>
<tr>
<td>2</td>
<td>2.6</td>
<td>2.7</td>
</tr>
<tr>
<td>3</td>
<td>3.3</td>
<td>2.9</td>
</tr>
<tr>
<td>4</td>
<td>3.1</td>
<td>2.7</td>
</tr>
<tr>
<td>5</td>
<td>5.3</td>
<td>4.6</td>
</tr>
<tr>
<td>6</td>
<td>3.4</td>
<td>3.1</td>
</tr>
<tr>
<td>7</td>
<td>2.4</td>
<td>1.9</td>
</tr>
<tr>
<td>8</td>
<td>4.6</td>
<td>4.1</td>
</tr>
<tr>
<td>9</td>
<td>5.8</td>
<td>5.0</td>
</tr>
<tr>
<td>10</td>
<td>3.7</td>
<td>3.6</td>
</tr>
<tr>
<td>11</td>
<td>4.0</td>
<td>4.4</td>
</tr>
</tbody>
</table>

The general trend and the comparison between the two sets of means can be seen from the following graphical presentation.

FIGURE 2.
COMPARISON OF ERROR RATIO MEANS OF OTIS I AND OTIS II.

2 children not included in Otis II (M=103.2 SD=10.4), and Otis II included 26 children not included in Otis I (M=106.2 SD=14.1). The higher mean of the Otis II group is in keeping with the higher mean of the total Otis II sample (i.e., 107.7 SD 13.4). The distribution of Otis II scores in 7 I.Q. groups based on approximately half SD classes is also given in appendix C.
The two patterns of mean error ratios are very similar, indicating a consistency in performance over the two tests, with the mean for Otis II being slightly lower overall. It is clear that on some categories there were many more errors and omissions than on others, and this was presumably due to the position in the test of the various items. Thus from the classification of items (see above 110-112) it can be seen that category (7), for example, has more items appearing earlier in the test than category (8).

This assumption, (that the mean error ratio is dependent on position of item in the test), can be tested by giving each category a value calculated from the mean ordinal number of the items in the category; e.g. category (8) contains items 36, 37, 45, 48, 51, 57, 62, 68; mean = 50.5. These values then represent the "predicted" difficulty of the categories and can be compared with the "actual" difficulty i.e. the mean error ratios. If the mean ordinal numbers are divided by 10 to make them comparable in size to the mean error ratios they can be entered on the same graph as the

---

3 According to the original Otis manual the items were placed in order of difficulty, based on "the number of passes of each item by the students taking the preliminary editions". Extract from 1928 Manual of Directions.
mean error ratios, thus making possible a rough comparison. In the following graph the actual difficulty is based on a combination of the two sets of mean error ratios given above (Table IV).

**FIGURE 3**

COMPARISON OF "ACTUAL" AND "PREDICTED" DIFFICULTY OF MEAN ERROR RATIOS.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although it is legitimate to compare the general pattern of the two graphs, it is not legitimate to draw any conclusions based on the absolute difference on any given category, because the values entered for the predicted difficulty are arbitrary reductions (i.e. divided by 10) of the mean ordinal values. An inspection of the two patterns shows that on categories (1) to (4) the general trend is the same. Category (5) shows a discrepancy which is difficult to eliminate, for if the graph of predicted difficulty is superimposed on the graph of actual difficulty
so that there is agreement at category (5), then all other categories are distorted. If instead, it is superimposed in such a way that the (6)-(11) categories are in point to point agreement, then the discrepancy on (5) is magnified.

This suggests, although it cannot prove, that the particular sample used differed from the original sample on which the test was standardised, in making more errors and omissions on (5), or in other words in finding category (5) items (i.e., mixed sentences) more "difficult". The question arises as to the nature of this difficulty and this requires a further examination of failures on category (5). On analysing individual failures it was apparent that there were many omissions, hence these were listed for each I.Q. group on each category. In the table given below (based on Otis I) the I.Q. groups have been combined as the same gross pattern was evident in each - i.e., the great preponderance of omissions on (5).

---

4 It has been pointed out (Thyne J.M. Patterns of Error in the Addition Number Facts, 1954, (Univ. of London Press) that "difficulty" as an objective observation, based on length of time taken to complete the test, or frequency/error, cannot necessarily be equated with difficulty as a quality of experience of the subject (p3,4). Heim has also discussed the complexities involved in trying to clarify the concept of 'difficulty' in test items (See Heim A.W. The Appraisal of Intelligence, 1954, (London: Methuen) p117ff.)
TABLE V

OMISSIONS ON ALL CATEGORIES (OTIS I).

<table>
<thead>
<tr>
<th>Categ</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total Cases</th>
<th>Total Omissions</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>11</td>
<td>3</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>43</td>
<td>60</td>
<td>5.9</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>10</td>
<td>10</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>5</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>25</td>
<td>30</td>
<td>2.9</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>3</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>20</td>
<td>25</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>44</td>
<td>24</td>
<td>17</td>
<td>21</td>
<td>6</td>
<td>172</td>
<td>429</td>
<td>42.2</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>28</td>
<td>10</td>
<td>5</td>
<td>.</td>
<td>.</td>
<td>78</td>
<td>141</td>
<td>13.9</td>
</tr>
<tr>
<td>7</td>
<td>44</td>
<td>6</td>
<td>2</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>53</td>
<td>67</td>
<td>6.6</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>.</td>
<td>44</td>
<td>75</td>
<td>7.4</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>5</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>26</td>
<td>33</td>
<td>3.2</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>77</td>
<td>98</td>
<td>9.6</td>
</tr>
<tr>
<td>11</td>
<td>36</td>
<td>5</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>42</td>
<td>49</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Altogether 42.2% of the total omissions were on Category (5), the next greatest percentage being 13.9 on (6). Of the 292 children, 172 or 58.9% omitted one or more items on (5), thus the high percentage of omissions was not due to a large number of omissions by a few children, but was well spread over the whole group. A separate analysis of category (5) showed that 65% of all girls omitted one or more items, and 55% of boys. One of the interesting features of the table is the spread on (5) from 1 to 6 omissions - i.e. the total number of items in the category; thus, for example 14% omitted 3 items, 9.9% omitted 4, 12.2% omitted 5 and 3.5% omitted 6.
These figures show quite clearly that this particular type of item was avoided by many children - either after having tried one or two unsuccessfully, or without any attempt at all; hence the high mean error ratio does not merely reflect failures due to an inability to solve the problem correctly, but also failures due to an attitude of avoidance. Doubtless this attitude springs from the fact that the problems are of an unfamiliar type when compared with many of the other items, furthermore the answer cannot be produced by a superficial examination of the words, and there is no clue in the form of distractors. That this attitude was not confined to a particular I.Q. group is evident from the following figures. - i.e. omissions on (5) as a percentage of total omissions for each I.Q. group.-

1=32.4  2=45  3=44.8  4=42.6  5=41.9  6=46.3  7=38.4

Although the results for category (5) support an interpretation such as the influence of an attitude of avoidance, this may be just a magnified example of attitudes, not necessarily dependent on level of intelligence, which may be aroused by many different types of item. When a child works through an intelligence test he is placed in a situation which is bound to give rise to a number of feelings, attitudes or specific sets. To cover all these conditions by a term such as "rapport" or "motivation" is to obscure conditions
which may exert a subtle influence on test performance. The child may react to an item in a way which might be verbalized as "I've done problems like this before - I know what is required" or, in the case of (5), "These look difficult, I've never come across them before, I can't see just how to start off, so I'll skip this type or I'll be wasting too much time." Once having skipped, there is a tendency at least among some children to skip a second, or third etc. The net result may be only a small loss in score and thus an unimportant variation in I.Q., yet the investigator interested in the processes involved in the solving of intelligence test items cannot afford to ignore any conditions which may affect success or failure.5

5Cf. the types of error described above (p94), i.e. the format of a problem discouraging any attempt to attack it; all-or-none attitude to problems; lack of confidence.
CHAPTER V.

A COMPARISON OF 'HIGH ERROR' AND 'LOW ERROR' GROUPS OVER ALL CATEGORIES.

If two children have approximately the same I.Q., and approximately the same mean error ratio score over the eleven categories, and if on any given category one child has a consistently high error ratio score over the two Otis tests, while the other child has a consistently low error ratio score, then this is a difference worth investigating. If instead of two children, two groups could be selected, matched for I.Q. and mean error ratio score, and showing consistency over the two tests, then some hypotheses could be tested, for example - that the difference is related to sex, socio-economic status, special ability or "carelessness".

I.- METHOD OF SELECTING GROUPS.

One major difficulty lay in securing groups of children whose error ratio scores were clearly 'high' or 'low' and at the same time ruling out the influence of I.Q. and overall error tendency as measured by mean error ratio score. A further limitation was the requirement of consistency.¹

¹It has been pointed out (Snodgrass F.T. "Unreliability of Group Test Profiles," J. Educ. Psych. 1954, p133) that too much faith in the precision of subtest profiles is not warranted. As a result of her investigation into the reliability of the Terman - McNemar "Test of Mental Ability" subtest profiles, Snodgrass concluded that "most of the irregularities of a profile, as measured by the test used, are due to errors in measurement. Profile differentiations are reliable only about 25% of the time in the most reliably tested area."(p134-135).
The difficulty could be overcome if the error ratio scores could be adjusted to eliminate the influence of I.Q., to allow for overall error tendency and to reflect the degree of consistency. With these adjustments the error ratio scores would be directly comparable from individual to individual for any given category, although not from category to category, within any individual error pattern. If, however, the error ratio scores could also be adjusted to allow for the different degrees of difficulty (as indicated by the I.Q. group, mean error ratio score, for any given category), then the final error ratio scores would be comparable between categories as well as individuals. This would make it possible to select two contrasting groups for each category directly from the adjusted error ratio scores, and the two groups should then be matched for I.Q. mean error ratio score and consistency.

This method of adjustment was attempted by using the following procedure:

1. Error ratio scores for each category listed.

2. Deviation of each score from the I.Q. group mean error ratio score for that category, (i.e. as given in Appendix A) listed.

3. Deviation adjusted up or down by an amount determined by the difference between the individual mean error ratio score and the I.Q. group mean ratio score (i.e. overall error ratio mean given as the total M of Ms in Appendix A).

4. Same steps repeated with Otis II error ratio means.
5. Two sets of adjusted deviations (i.e. steps 3 and 4), added.

This procedure resulted in a single list of eleven deviations (plus or minus) for each individual, which could be used for the selection of high error and low error groups for each category.²

The next problem involved the setting up of a criterion in order to separate out the high deviation (or H.E.) group from the low deviation (or L.E.) group. An arbitrary deviation - e.g. 2.0 (± or -) could not be used because of the different means and distributions for the various categories and groups. On the other hand the use of standard deviation as a criterion would have resulted in the elimination of too many cases; furthermore, it was not desired that extreme deviations should influence the measure of deviation unduly. In the light of these considerations the mean variation was chosen as a satisfactory measure and the M.V.³ for all the means calculated. (see Appendix E.)

For the purposes of selecting the H.E. and L.E. groups for each category, a deviation was judged as falling into the 'high' group if it was greater than two M.V.³ above the mean,

²A detailed description of the method of adjustment etc. and the rationale underlying it is given in Appendix D.
(i.e. M.V. on Otis I + M.V. on Otis II), or into the 'low' group if two M.V.'s below the mean. Two restrictions were imposed:

1. The deviation was rejected if it was made up of one very high deviation on one test, plus a low or opposite deviation on the other. (This restriction ensured consistency in the cases selected).

2. The deviation was rejected if the number of items upon which the error ratios were calculated was less than five (i.e. Otis I + Otis II). In most cases this minimum qualification was far exceeded, although in category (9) many cases had to be dropped. (This restriction reduced the influence of chance factors in producing high deviations).

The final step was to take each category in turn and to select the H.E. and L.E. groups in terms of the criteria discussed above. In some cases where the two groups were very uneven, a few cases were included in which the Otis II alone had been worked, although in view of the fact that there was no guarantee of consistency, the criterion was raised to $1\frac{1}{2}$ times the M.V. (equal to 3 times M.V. in the summed deviations).3

II. RESULTS AND DISCUSSION FOR EACH CATEGORY.

The results for each category are given separately, with category (1) in some detail in order to illustrate the

---

3 In the example given at the beginning of Appendix D, the summed deviations (in final column) which fall into H.E. or L.E. groups are underlined, i.e. Category (7) and (10) H.E., (5) and (9) L.E.
types of hypotheses set up and the methods of interpreting the data, while categories (2) to (11) are treated in the same general manner but in less detail.

(a) Category (1) Word Meaning. (Items 5, 16, 23, 40, 44, 54, 61, 75).

Example: No. 23.
A boy who often tells big stories about what he can do is said to ( )
1 lie, 2 fake, 3 cheat, 4 joke, 5 brag............( ).

It was first necessary to check the two groups in order to see whether the procedure outlined in Appendix D had been successful. If so, then the two groups should differ significantly in their mean error ratio scores, but should not differ significantly in I.Q. or in overall error tendency. As a measure of the first of these it was necessary to use the unadjusted error ratio scores, while as a measure of the last, the W/A score could be used (i.e. total errors over all items attempted). The W/A score does not include omissions and this is an advantage as one of the major hypotheses to be tested deals with the errors only and ignores omissions.4

4As a check, the total number of omissions over the whole test was compared for the H.E. and L.E. groups on each category, with no differences being observed except in the case of category (5) (to be dealt with later). Thus it made no difference which measure of overall error tendency was used - W/A, which does not include omissions, or the mean of the error ratio scores over the eleven categories, which does include omissions.
The following figures provide the necessary data for checking.

### TABLE VI

**UNADJUSTED ERROR RATIO SCORE FOR CATEGORY (I)**

**H.E. AND L.E. GROUPS.**

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th></th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>29</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>M</td>
<td>6.25</td>
<td>6.18</td>
<td>1.88</td>
</tr>
<tr>
<td>S,D</td>
<td>2.06</td>
<td>1.53</td>
<td>2.38</td>
</tr>
<tr>
<td>S,E,M</td>
<td>0.39</td>
<td>0.29</td>
<td>0.30</td>
</tr>
</tbody>
</table>

S,E. diff.\(^5\)  
Otis I = 0.63  
Otis II = 0.42  
Critical Ratio  
Otis I = 6.9  
Otis II = 9.8  
(t for 51 df at 0.01 level 2.68) - Otis I  
(t for 56 df at 0.01 level 2.66) - Otis II

Thus the difference between the L.E. and H.E. groups on the unadjusted error ratio score means is significant at well beyond the 0.01 level of confidence.

---

\(^5\)Throughout this whole section the small sample formula for S,E,M has been used i.e. 
\[
\sqrt{\frac{N - 1}{SD}}
\]
TABLE VII.

I.Q. (OTIS I AND II COMBINED) OF CATEGORY (1) H.E. AND L.E. GROUPS.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>106.1</td>
<td>107.0</td>
</tr>
<tr>
<td>SD</td>
<td>10.6</td>
<td>14.2</td>
</tr>
</tbody>
</table>

These results indicate that the procedure has in fact resulted in the selecting of groups of similar I.Q., although in this particular category there is an unaccounted difference in variability of 3.6 points of S.D.

TABLE VIII.

W/A SCORES\(^6\) (IN PERCENTAGES) OF CATEGORY (1) H.E. AND L.E. GROUPS.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otis I</td>
<td>32.0</td>
<td>29.7</td>
</tr>
<tr>
<td>Otis II</td>
<td>27.7</td>
<td>28.5</td>
</tr>
<tr>
<td>M</td>
<td>10.4</td>
<td>10.2</td>
</tr>
<tr>
<td>SD</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>SE(_M)</td>
<td>1.9</td>
<td>2.6</td>
</tr>
</tbody>
</table>

SE diff. Otis I 3.2, C.R. 1.34, (t for 51 df 1.68 at 0.10 level)
SE diff. Otis II 3.2, C.R. 0.38, (t for 56 df 1.67 at 0.10 level)

The method of eliminating overall error has not been entirely successful as there is a difference of 4.3% between

---

\(^6\)Using the measure - percentage of the 75 items which were covered - instead of W/A, the resulting figures were Otis I H.E. 78.7, L.E. 77.3; Otis II H.E. 90.8, L.E. 90.3; i.e. there is no difference between the two groups in amount of test covered.
the two groups on Otis I and 1.2% on Otis II. These differences are not significant, especially for Otis II, although even in Otis I the difference is well below the 0.10 level.

Taking tables VI to VIII together, there is justification for the claim that the two groups show considerable difference in tendency to make errors on Category (1), despite the fact that there is little difference in mean I.Q. and only a small (and insignificant) difference in overall error tendency.

The question now arises - what causes the one group to make so many errors while the other group makes so few errors? In order to throw light on the question some hypotheses can be stated and tested.

**Hypothesis 1.**

Word knowledge depends on experience, and experience in this particular category gained in the home and in the school, is partly dependent on age. Thus if the two groups are equated for mean I.Q., the low error scorers may be found among the older children who make up on word knowledge what they may lack in such areas as verbal or arithmetical reasoning. The hypothesis, then, is that the mean age of the L.E. group should be greater than that of the H.E. group. This must be rejected because the mean age of the two groups
(as at mid-November 1953) shows no difference at all - i.e. mean age of both the H.E. and the L.E. group = 11 years 7 months.

Hypothesis 2.

Although in the construction of a test of this nature items which show sex differences are usually excluded, there may still be some small sex differences on word meaning in view of the greater verbal maturity of girls on the average. The hypothesis, then, is that more boys should appear in the H.E. group than in the L.E. group, and vice versa with girls.

TABLE IX.

PERCENTAGES OF BOYS AND GIRLS IN CATEGORY (1) H.E. AND L.E. GROUPS.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>19 (66%)</td>
<td>19 (66%)</td>
</tr>
<tr>
<td>Girls</td>
<td>10 (34%)</td>
<td>10 (34%)</td>
</tr>
</tbody>
</table>

The hypothesis must be rejected as the proportions remain constant.

Hypothesis 3.

The difference between the groups may be one of socio-economic status, for as with age, children from a socially sophisticated environment may make fewer errors on word meaning but more errors on 'reasoning' items, while children from a less sophisticated environment may gain the same I.Q. but by
just the opposite process. This is all the more likely in view of the findings of the Eells' study\(^7\), that the children from higher socio-economic status homes had better scores on verbal items than the children from the lower socio-economic status homes. The hypothesis is that the H.E. group should have a lower socio-economic status ranking than the L.E. group.

Mean status rankings on the 7 point socio-economic scale (as described above pp108-109) were H.E. 5.2, L.E. 5.5. This is a very small difference, and in fact for the purposes of this hypothesis it is in the wrong direction! Thus socio-economic status is apparently not a contributory factor.

Hypothesis 4.

Children who make many errors on word meaning may also make many errors on categories which are of a similar "verbal" type, or in other words, these children are perhaps weak in "verbal ability". This does not, of course, fully explain the errors, it merely shifts the explanation from one plane - i.e. weakness in general intelligence, to another more specific plane, i.e. weakness in verbal ability.

\(^7\)Eells K. et.al Intelligence and Cultural Differences, 1951, (Chicago, Univ. Chicago Press) - See discussion above on page 97.
Despite the limitations of such comparisons, a brief consideration of the performance of the category (1) H.E. and L.E. groups on other categories may be worth while, for whatever magnifies or reduces the errors in the two groups may be due to something more than variations in verbal ability.8

Although it is dangerous to try and predict what process or ability underlies a given item, an a priori inspection of the Otis items suggests the hypothesis that the H.E. group on category (1) should also show a higher mean error ratio on categories (2) (opposites), (3) (word attribute or definition), and (4) (forming a concept from given examples).

<table>
<thead>
<tr>
<th>Categ.</th>
<th>H.E. Otis I</th>
<th>Otis II</th>
<th>L.E. Otis I</th>
<th>Otis II</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.25</td>
<td>6.18</td>
<td>1.88</td>
<td>1.90</td>
</tr>
<tr>
<td>2</td>
<td>2.24</td>
<td>1.95</td>
<td>2.46</td>
<td>2.61</td>
</tr>
<tr>
<td>3</td>
<td>4.35</td>
<td>3.03</td>
<td>2.92</td>
<td>2.74</td>
</tr>
<tr>
<td>4</td>
<td>2.50</td>
<td>1.84</td>
<td>3.05</td>
<td>3.24</td>
</tr>
<tr>
<td>5</td>
<td>4.77</td>
<td>3.35</td>
<td>5.27</td>
<td>5.72</td>
</tr>
<tr>
<td>6</td>
<td>2.62</td>
<td>2.39</td>
<td>3.42</td>
<td>3.34</td>
</tr>
<tr>
<td>7</td>
<td>2.09</td>
<td>2.29</td>
<td>2.48</td>
<td>2.13</td>
</tr>
<tr>
<td>8</td>
<td>3.80</td>
<td>3.56</td>
<td>4.14</td>
<td>3.55</td>
</tr>
<tr>
<td>9</td>
<td>6.55</td>
<td>5.17</td>
<td>5.56</td>
<td>4.63</td>
</tr>
<tr>
<td>10</td>
<td>2.66</td>
<td>2.33</td>
<td>3.52</td>
<td>4.53</td>
</tr>
<tr>
<td>11</td>
<td>3.94</td>
<td>3.80</td>
<td>3.54</td>
<td>4.18</td>
</tr>
</tbody>
</table>

One such 'extra' would be the 'uniqueness' or particular structure of the items; again not an explanation of the error, but at least a restricting of the field of search for the error producing processes.
Of the differences between the 22 means given in the table, only 4 are significant, i.e. categories (3), (4), (5) and (10). In terms of the hypothesis: category (3) shows a difference in the expected direction (consistent and significant on Otis I); category (2) shows a difference in the opposite direction (consistent but not significant), while category (4) also shows a difference in the opposite direction (consistent and significant on Otis II).

As category (2) is based on only 4 items any interpretation is likely to be unreliable, although the difficulty involved in the items may be due not so much to the uncommon nature of the words, as to the particular structure of the questions. This possibility will be examined in the analysis of wrong responses for each item.

Category (4) also, has few uncommon words, the necessary condition being one of concept formation or inductive

9 Cat. (3) Otis I Obs. diff. 1.43, SE diff. 0.61, CR 2.36, t for 51 df 2.01 at 0.05 and 2.40 at 0.02 level, i.e. significant between 2 and 5% levels.
Cat. (4) Otis II Obs. diff. 1.40, SE diff. 0.54, CR 2.59, t for 56 df 2.39 at 0.02 and 2.67 at 0.01 level, i.e. significant between 1 and 2% levels.
Cat. (5) Otis II Obs. diff. 2.37, SE diff. 0.70, CR 3.39, t for 56 df 2.67 at 0.01 level, i.e. significant beyond 1% level.
Cat. (10) Otis II Obs. diff. 2.20, SE diff. 0.68, CR 3.22, t for 56 df 2.67 at 0.01 level, i.e. significant beyond 1% level.
reasoning rather than of familiarity with the meanings of the words.

The other two categories with significant differences are (5) - mixed sentences and (10) - number series, in both cases being differences in the opposite direction to the differences on category (1). There is no point in speculating on the reasons for these differences at this stage, although if the differences appear again in the analyses of categories (5) and (10), then some interpretation may be attempted. On the whole, the data presented in Table X, far from supporting the a priori predictions, confirm the futility of trying to predict performance on a given item from a somewhat similar type of item. This suggests the necessity for supplementing any classification of items with an inquiry into the influence of changes in form or wording on the structure of the items as perceived by the child.

Hypothesis 5.

It is possible that the difference between the two groups is dependent on the particular sampling of words used in the Otis test, one group having had the necessary experiences and the other group for some reason having failed to meet or make use of these particular words. If this is so, then with a different sampling of words the difference
should disappear. This hypothesis can be tested by comparing the performance of the two groups on a different type of test altogether, for example the Primary Mental Abilities test for "V". As it is 'error' ratio upon which the two groups were selected, it is necessary to compare the groups on the ratio of errors to correct responses in V\(^\text{10}\) rather than on the positive scores, although the latter are also shown in the table along with the Q scores, as a check on the comparability of the groups.

**TABLE XI**

CATEGORY (1) H.E. AND L.E. GROUPS COMPARED ON P.M.A. Q AND V.

<table>
<thead>
<tr>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>27</td>
</tr>
<tr>
<td>Q, Mean Score</td>
<td>105.2</td>
</tr>
<tr>
<td>V(Error ratio or W/A)</td>
<td></td>
</tr>
<tr>
<td>Mean %</td>
<td>30.5</td>
</tr>
<tr>
<td>V%ile</td>
<td>54.1</td>
</tr>
</tbody>
</table>

Q - SE diff. 3.62; CR 0.92; t for 53 df at 0.10=1.68
V(W/A) - SE diff. 4.66; CR 2.00; t for 53 df at 0.05=2.01
V%ile - SE diff. 6.27; CR 1.46; t for 53 df at 0.10=1.46

\(^{10}\)The ratio 'errors over number of items attempted' does not include omissions as, by the arrangement of the questions on the P.M.A. test sheet, some children are able to skip quickly down part of the list and pick out an item which is known while omitting up to 6 or 7 items between 'skips'.

---
From these results it appears that the hypothesis cannot be sustained, for although the difference on W/A for V is only just significant at the 5% level of confidence, the H.E. group has a higher mean O score which would act to reduce the gap between the two groups on V (W/A). There is also a difference (in favour of the L.E. group) on the positive scores for V, although due to the very large S.Ds for V, the difference is not significant.\footnote{A further comparison, i.e. on the Word Fluency scores of the two groups resulted in these mean percentiles - H.E. 53.6 and L.E. 54.4.}

The results indicate that there may be a difference in verbal ability, as measured by the P.M.A. verbal test, which accounts, at least in part, for the difference between the two groups. Although it is clear that the extreme difference between the groups, based on consistent Otis Category (1) errors, is not reflected to such an extent in the P.M.A. verbal errors, it is not possible, with the data given here, to know how much of the Otis error uniqueness depends upon the particular sampling of words and how much on the structure of the items. This point should be clarified in working through errors with individual children.
Hypothesis 6.

A further check on the generality of word meaning errors can be provided by comparing the performance of the two groups on a type of test very similar to the Otis - namely the A.C.E.R. Intermediate D (given approximately 6 months after the second Otis). In the light of the comparison made with the P.M.A. (V) errors, it could be predicted that at least part of the difference would carry over to the Intermediate D, and this hypothesis can be tested by comparing the performance of the two groups on category (3) of Intermediate D (i.e. Synonyms - items 3, 19, 22, 25, 28, 51, 54, 67). As with the Otis a check on the comparability of the two groups is provided by giving the Intermediate D I.Q. and total W/A as well as the error ratio means for category (3).

TABLE XII
CATEGORY (1) H.E. AND L.E. GROUPS COMPARED ON INTERMEDIATE D SCORES.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>Mean I.Q.</td>
<td>103.8</td>
<td>103.9</td>
</tr>
<tr>
<td>Mean W/A%</td>
<td>38.7</td>
<td>36.2</td>
</tr>
<tr>
<td>Mean (3)</td>
<td>4.9, S.D 2.27</td>
<td>3.6, S.D 2.00</td>
</tr>
</tbody>
</table>

S.E. diff. on (3) 0.59, CR 2.2, t for 51 df 2.01 at 0.05
2.40 at 0.02 level
The difference is significant at the 2-5% level of confidence, and this is of interest, because the 8 synonym items are cast in a different form from the Otis word meaning items e.g.:-

**Item 16 Otis.**

When a new kind of machine is thought of, it is usually called?
1 a discovery, 2 adoption, 3 a creation, 4 a novelty, 5 an invention.

**Item 19 Intermediate D.**

**SOLEMN** means
1 sermon, 2 preach, 3 serious, 4 wisdom, 5 happy.

The Otis item structure is more complex, and this may give some clue to the cause of error, over and above any verbal ability which may underly both sets of items.\(^\text{12}\)

So far the evidence has been mainly negative - i.e. it has ruled out certain hypotheses, rather than given direct indications of areas in which to look for error causing processes. Results arising from the last two hypotheses have been interpreted as indicating that given two groups with the same I.Q., qualitative differences can appear (such as a

\(^\text{12}\)In the discussion of Table X (hypothesis 4), it was pointed out that there was a significant difference (beyond 1% level of confidence) on category (10) - number series. As this category also occurs in Intermediate D, a comparison could be made between the H.E. and L.E. groups, thus the error ratio means for Category (4) - number series, were calculated, the H.E. mean being 2.5 and the L.E. mean 2.9. The difference is not significant, although it is in the same direction as the difference found on Otis category (10).
tendency to make few or many errors in word meaning items) which to a lesser extent persist in similar items of different tests. This can be accounted for readily enough in terms of a similar factorial content, in this case the influence of a verbal factor, yet this does not completely dispense with the problem involved in deciding how errors were made - for example, to what extent a particular error was the result of a guess because of ignorance, or to what extent a result of the way in which the item happened to be framed. Hypothesis 7.

Some light may be thrown on this problem by noting the specific errors made on each item in the category and comparing the frequency of the particular errors made by the two groups. This would provide data for answering the question: "Do the children in the H.E. group make a different type of error from those in the L.E. group?" It seems likely that if the particular item structure influences the response made, then the H.E. group may show a tendency to make a different type of error, and this hypothesis can be tested by examining the numbers choosing each wrong distractor for each item in the two groups.\textsuperscript{13}

\textsuperscript{13} Figures for Otis I only have been given, as the patterns of errors on Otis II were consistently similar to those on Otis I.
TABLE XIII

DISTRIBUTION OF SPECIFIC ITEM ERRORS FOR H.E., L.E. AND N(NEUTRAL) GROUPS ON CATEGORY (1) ITEMS.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>5</td>
<td>HE</td>
<td>LE</td>
<td>N</td>
<td>HE</td>
<td>LE</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>64</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>82</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>75</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>59</td>
<td>26</td>
<td>15</td>
<td>34</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>20</td>
<td>61</td>
</tr>
<tr>
<td>80</td>
<td>4</td>
<td>1</td>
<td>20</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>50</td>
<td>8</td>
<td>1</td>
<td>20</td>
<td>7</td>
<td>44</td>
</tr>
<tr>
<td>34</td>
<td>12</td>
<td>10</td>
<td>29</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>57</td>
<td>4</td>
<td>2</td>
<td>30</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>39</td>
<td>9</td>
<td>7</td>
<td>30</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>21</td>
<td>7</td>
<td>2</td>
<td>30</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>2</td>
<td>25</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>2</td>
<td>30</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>61</td>
<td>10</td>
<td>1</td>
<td>31</td>
<td>19</td>
<td>59</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>56</td>
</tr>
<tr>
<td>29</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>9</td>
<td>50</td>
</tr>
<tr>
<td>61</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>29</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>17</td>
<td>3</td>
</tr>
</tbody>
</table>
One of the most obvious features of this table is the uneven distribution of choices over the four distractors for each item — e.g. (in the H.E. group) item 16, 64% chose distractor (1); item 23, 82% chose (1); item 40, 61% chose (4); item 54, 49% chose (5); item 61, 59% chose (4); item 75, 61% chose (1). Because, by its very nature, the L.E. group made few errors over all the items, it is not possible to make an exact comparison between the two groups, although an inspection of the distribution of the few errors listed in the table may indicate whether similar patterns of errors occurred.

On item 23 the 2 groups showed a similar pattern — i.e. H.E. 82% on distractor (1); L.E. 75% on (1).

On item 40 both groups chose distractors (1) and (4) although of the total 5 errors made by L.E. group, 4 were made on (1), compared with 30% of H.E. total, while on (4) the H.E. group made 61% compared with 1 out of 5 for L.E. group. The actual item reads: "An event which is sure to happen is said to be(?)" 1 probable, 2 certain, 3 doubtful, 4 possible, 5 delayed.

Of the four distractors, (1) is closer in meaning to the correct answer than (4), while (4) has a close physical proximity to the question mark (as set out in the printed test), which might lead the hurried or careless reader to choose it in preference to (1). A hurried or 'careless' approach could be one of the characteristics which distinguishes the two groups, however, this question cannot be answered on the basis of the evidence given here, but may prove to be a useful clue in the individual analyses.

On item 44 the 2 groups spread their errors over (1) and (5), although the H.E. group also spread them over (3) and to a lesser extent (4). The distribution for this item is inconclusive.

On item 54 both groups tended to choose (5) (H.E. 49%, L.E. 50%) more often than (1), (3) or (4), while on item 61, both chose
(4) (H.E. 59%, L.E. 56%) more often than (2), (3), or (5).

On items 5, 16, 75, there were too few errors made by the L.E. group for any comparison.

From the data analysed it appears that on the whole the H.E. and L.E. groups tend to make the same type of error, the difference being one of quantity of error rather than quality.14 This could be interpreted as indicating that the difference is due to the amount of verbal experience of the two groups; but on the other hand a certain amount of the difference may be due to different reactions to "distraction" (in the sense of the particular structure of these items with their four distractors). Although the similarity of the error patterns for the two groups gives a certain amount of support to the latter interpretation, it is not clear why such "distraction" should be confined to word meaning items.

---

14 As a check, a "neutral" group was selected, i.e. a group of children whose summed deviations were zero or near zero, thus falling on the mean of category (1) error ratio scores. The range of deviations included was -0.4 through zero to +0.4; mean I.Q. 107.5. W/A mean 29.0% (Otis I and II averaged); hence the group was comparable to H.E. and L.E. groups. An analysis of errors for each item revealed a pattern similar to that of the H.E. group - i.e. item 23, 59% on (1) H.E. = 82%; item 40, 50% on (1) and 44% on (4) H.E. = 30% and 60%; item 44, 39% on (1) 30% on (3) 22% on (5) H.E. = 34%, 29% and 26%; item 54, 69% on 5, H.E. = 49%; item 61, 39% on (3), 50% on (4) H.E. = 31% and 59%; item 75, 29% on (1), 35% on (3) H.E. 61% and 11%. In the case of 23, 40 and 75 the neutral group varied in the direction of the L.E. group.
Summary for category (1).

Although the substantial difference (of mean error ratio scores on category (1)) between the H.E. and L.E. groups cannot be explained in terms of I.Q. level, or overall error tendency, the various hypotheses suggested are also inadequate explanations.

(a) The difference is not dependent on age, sex or socio-economic status.

(b) A priori grouping of similar categories does not throw any light on the difference.

(c) Although the difference is also apparent in performance on other tests, such as P.M.A. Verbal test and Intermediate D, Synonyms, there still remains the problem of explaining the actual form of particular errors.

(d) The two groups tend to have similar error patterns, although the small number of errors made by the L.E. group makes an accurate comparison difficult.

(e) In most items, responses are not spread evenly over the four distractors but are concentrated on one or two particular distractors. This points to the necessity of attempting to clarify the effects of item structure on the production of error.
(b) Category (2) Opposites. (Items 3, 8, 24, 63).

Example: No. 24.
Which one of the five words below means the opposite of difficult?
1 hard, 2 quick, 3 soft, 4 easy, 5 common.............( )

The following table gives the data necessary for checking on the comparability of the two groups and determining the influence of sex and socio-economic status.

**TABLE XIV**

**COMPARISON OF CATEGORY (2) H.E. AND L.E. GROUPS.**

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>31</td>
<td>34</td>
</tr>
<tr>
<td>Otis II I.Q.</td>
<td>102.4</td>
<td>104.2</td>
</tr>
<tr>
<td>W/A</td>
<td>33.9%</td>
<td>34.0%</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>19 (61%)</td>
<td>17 (50%)</td>
</tr>
<tr>
<td>(Girls)</td>
<td>12 (39%)</td>
<td>17 (50%)</td>
</tr>
<tr>
<td>S.E.Status</td>
<td>5.6</td>
<td>5.6</td>
</tr>
<tr>
<td>Error Ratio (2)</td>
<td>6.16</td>
<td>0.95</td>
</tr>
</tbody>
</table>

From these figures it may be concluded that:-

(a) The two groups show little difference in I.Q. and overall error tendency. (i.e. check on selection procedure).

(b) There is a substantial difference in the mean error ratio scores which is significant at well beyond the 1% level of confidence.

15 In this and the following categories, the I.Q., W/A, and category mean error ratio scores are based on Otis II only, as experience with Category (1) indicated that it was unnecessary to duplicate these measures.
(c) There is no difference in socio-economic status ranking.

(d) There are sex differences between the two groups, more girls (proportionately) than boys falling into the L.E. group.16

As there are only four items in this category, it does not warrant detailed treatment, although one further set of figures may prove relevant - i.e. the patterns of errors over the four items for the two groups:--

TABLE XV.

DISTRIBUTION OF SPECIFIC ITEM ERRORS FOR CATEGORY (2).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>4</td>
<td>21</td>
<td>3</td>
<td>16</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>LE</td>
<td>12</td>
<td>32</td>
<td>3</td>
<td>8</td>
<td>20</td>
<td>53</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td></td>
<td>20</td>
<td>1</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The whole question of sex differences will be discussed when all the categories have been covered.
The only outstanding figure is the 82% on distractor (1) for item (24) (H.E. group), the two errors in the L.E. group also occurring on this distractor. On item 63, the highest percentage of errors occurred on distractor (1) with slightly lower percentages on (3), (4) and (5) in the H.E. group, while the L.E. group errors were distributed in a similar manner. Although these figures throw no light on the differences between the two groups, they support the suggestion made for category (1) that the L.E. group errors are of the same general type as those made by the H.E. group, the difference being one of quantity not quality.

(c) Category (3) Word Attribute or Definition. (Items 2, 14, 18, 27, 31, 35, 52, 66).

Example: No. 27.
Which tells best just what a horse is?
1 It has a tail, 2 a live thing, 3 a thing that works and eats, 4 a large, fourlegged animal, 5 something to pull a wagon.

\[ ...............( ) \]

\[
\begin{array}{ccc}
\text{TABLE XVI} \\
\text{COMPARISON OF CATEGORY (3)} \\
\text{H.E. AND L.E. GROUPS.} \\
\hline
& \text{H.E.} & \text{L.E.} \\
\hline
\text{N} & 30 & 29 \\
\text{Otis II I.Q.} & 109.2 & 110.6 \\
\text{W/A} & 29.5\% & 28.2\% \\
\text{Sex (Boys)} & 16 (53\%) & 25 (86\%) \\
\text{(Girls)} & 14 (47\%) & 4 (14\%) \\
\text{S.E.Status} & 5.5 & 5.7 \\
\text{Error Ratio (3)} & 4.4 & 1.1 \\
\text{Error Ratio (1)} & 4.3 & 3.8 \\
\text{Error Ratio (10)} & 2.6 & 4.5 \\
\hline
\end{array}
\]
These figures indicate that:-

(a) The two groups are similar in mean I.Q. and W/A.

(b) Mean error ratios for this category show a difference of 3.3 which is significant well beyond the 1% level.

(c) Sex differences are clearly present in the L.E. group, boys making up 86% of the total group.

(d) There is no significant difference in socio-economic status.

(e) The difference between the category (1) H.E. and L.E. groups was also evident on category (3) (significant on Otis I). Figures given above are based on Otis II and a small (but no significant) difference of 0.5 is indicated for these category (3) groups on category (1). Comparable figures for these same groups on Otis I are 4.3 and 3.9. It is clear, therefore that the differences between H.E. and L.E. on (1) and (3) are not reciprocal despite the fact that they are both based on definitions of words or phrases.

(f) Category (10) (Number series) again stands out as showing a difference in the opposite direction.  

\[ t \text{ for } 57 \text{ df at } 0.02 \text{ level is } 2.39, \text{ and at } 0.01 \text{ level is } 2.67, \text{ thus the difference is significant between the } 1\% \text{ and } 2\% \text{ levels of confidence. On Otis I the difference is significant at the } 5\% \text{ level.} \]
of this reversal of error tendency will be reserved until category (10) results are presented. In order to test the possibility that this reversal also carries over to Intermediate D category (4), number series, these error ratio scores were compared.

H.E. group on Otis (3) - 2.48 on Intermediate D (4)
L.E. group on Otis (3) - 3.05 on Intermediate D (4)

the difference is in the expected direction but is not significant. 'N' in the P.M.A. is quite different in structure and purpose from number series, nevertheless a comparison of the two groups on 'N', taking the positive percentiles (errors are included in the positive score because of deductions for wrong answers) resulted in the following:

H.E. group on Otis (3) - Q.=101.8; N=54.0%ile
L.E. group on Otis (3) - Q.=101.6; N=43.6%ile

This difference (although not significant) is in the expected direction - i.e. high error group gaining a higher positive score than the L.E. group.

The final analysis of category (3) - namely distractor distribution of each item - is given in the following table.
<table>
<thead>
<tr>
<th>Table XVII</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distribution of Specific Item Errors</strong> For Category (3).</td>
</tr>
<tr>
<td><strong>Column</strong></td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>21</td>
</tr>
<tr>
<td>27</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>23</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>66</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>
Item 14 - both groups have a strong bias towards distractor (5) - H.E. 87%, L.E. 100%.

Item 18 - H.E. 59% on (2), L.E. 50% on (5).
"What is the most important reason that motor cars have displaced horses and carriages?"
1 Horses were getting scarce, 2 Horses often run away, 3 Motors save time and "time is money," 4 Motors are cheaper than carriages, 5 Motors cost less to repair than carriages.

The difference between the choices of the two groups may reflect sex difference because the L.E. group is 86% boys, and boys may be more attracted by the possibilities of (5) than by (2). An analysis of the H.E. response to 18 in terms of sex gives some support to this, for whereas an equal number of boys and girls chose (2) - 60% and 57%, 20% of the boys chose (5) - girls 0%, whereas more girls chose (1) than boys - 29% and 13%.

Item 27 - both have a strong bias towards distractor (5), H.E. 64%, L.E. 100%.

Item 31 - distractor (1) carries the greatest percentage for both H.E. 41%, L.E. 100% - H.E. 28% on (3) is not matched by any L.E. group errors.

Item 35 - In this case the L.E. group is divided more equally between (4) and (5) than the H.E. group - i.e. L.E. 38% and 38%; H.E. group 27% and 60%.

Item 52 - Greatest proportion of errors for both groups on (3) - H.E. 42%, L.E. 71%.

Item 66 - Both groups have a strong bias for (3) - H.E. 93%, L.E. 89%.

Although the patterns of errors of the two groups show some discrepancies, the 'biases' of the H.E. group tend to be repeated in the L.E. group.
(d) Category (4) Forming a Concept from Given Examples.
(Items 1, 9, 25, 41, 58.)

Example: No. 25. 
"Which one of the five things below is most like these:
  snake, cow, sparrow?"
  1 tree, 2 doll, 3 pig, 4 feather, 5 skin ..........( )

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>35</td>
<td>39</td>
</tr>
<tr>
<td>Otis II I.Q.</td>
<td>106.4</td>
<td>107.0</td>
</tr>
<tr>
<td>W/A</td>
<td>32.6</td>
<td>30.9</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>22 (63%)</td>
<td>24 (62%)</td>
</tr>
<tr>
<td></td>
<td>(Girls)</td>
<td></td>
</tr>
<tr>
<td>Sex (Girls)</td>
<td>13 (37%)</td>
<td></td>
</tr>
<tr>
<td>S.E. Status</td>
<td>5.5</td>
<td>5.2</td>
</tr>
<tr>
<td>Error Ratio (4)</td>
<td>6.06</td>
<td>0.63</td>
</tr>
<tr>
<td>Error Ratio (1)</td>
<td>3.86</td>
<td>5.22</td>
</tr>
</tbody>
</table>

(a) Groups are similar in mean I.Q. and W/A.

(b) Difference between the two groups on category (4) is significant well beyond 1% level (CR 14).

(c) No sex differences in the proportions in each group.

(d) Slight but not significant difference of 0.3 in S.E. status.

(e) On category (1) these two groups show a difference of 1.36, which is significant at beyond the 1% level. This difference is reversed, as would be expected from the results of the category (1) analysis where there was also a significant difference favouring the H.E. group. It is not
at all clear why an H.E. group on word meaning should make fewer errors than the L.E. group on 'forming a concept'.

The nearest approximation to the Otis Category (4) type question in the Intermediate D is (2) - Classification (2,5,9,16,20,58), e.g.: 9. Four of the following words are alike. What is the other word? (1) ice (2) cream (3) dew (4) hail (5) rain. A comparison of performance on Intermediate D (2) yields mean error ratio scores of H.E. 3.14, L.E. 3.46. Thus the difference does not transfer to a similar type of item in another test.

**TABLE XIX**

**DISTRIBUTION OF SPECIFIC ITEM ERRORS FOR CATEGORY (4).**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total Om's E'rs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>HE</td>
<td>6 75</td>
<td>2 25</td>
<td></td>
<td></td>
<td>1 8</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>HE</td>
<td>3 12</td>
<td>3 12</td>
<td>15 58</td>
<td>5 18</td>
<td>2 26</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>HE</td>
<td>2 4</td>
<td></td>
<td>13 27</td>
<td>33 69</td>
<td>1 48</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>HE</td>
<td>6 12</td>
<td>8 15</td>
<td>14 27</td>
<td>24 46</td>
<td>1 52</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>1 2</td>
<td></td>
<td>1 1</td>
<td>7 7</td>
<td>1 11</td>
</tr>
</tbody>
</table>

As the L.E. group errors are confined to items 41 and 58, only a limited comparison is possible, although in both
cases the pattern of errors is very similar to the H.E. group pattern. In the H.E. group there is evidence of strong bias towards certain distractors.

(e) Category (5) Mixed Sentences. (Items 12, 28, 32, 49, 55, 60).

Example: No. 12. "If the words below were rearranged to make a good sentence, with what letter would the last word of the sentence begin?" (Make the letter like a printed capital). usually are made of tables wood .................................

TABLE XX

COMPARISON OF CATEGORY (5) H.E. AND L.E. GROUPS.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>34</td>
<td>28</td>
</tr>
<tr>
<td>Otis II I.Q.</td>
<td>107.9</td>
<td>110.7</td>
</tr>
<tr>
<td>W/A</td>
<td>27.9</td>
<td>27.2</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>65% (22)</td>
<td>50% (14)</td>
</tr>
<tr>
<td>(Girls)</td>
<td>35% (12)</td>
<td>50% (14)</td>
</tr>
<tr>
<td>S.E. Status</td>
<td>5.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Error Ratio (5)</td>
<td>7.3</td>
<td>1.4</td>
</tr>
<tr>
<td>Error Ratio (1)</td>
<td>3.7</td>
<td>4.1</td>
</tr>
</tbody>
</table>

(a) There is a small (2.8 points) but not significant difference in I.Q. between the groups, although the W/A scores are similar.

(b) Difference between the two groups on category (5) is well beyond the 1% level of significance.

(c) There is a sex difference in the L.E. group, proportionately fewer boys than girls falling into the L.E. group.
(d) No socio-economic status difference between the groups.

(e) The H.E. and L.E. groups of Category (1) showed a significant difference on category (5), at least in Otis II, and this difference can be checked in the table above. Although the difference is in the expected direction, i.e. the H.E. group making fewer errors on (1) than the L.E. group, it is not significant, thus no interpretation is warranted.

In an earlier analysis (pp.137-139 above) it was shown that category (5) had a much higher mean error ratio than was expected, and that this was due in large part to the unusual number of omissions. As most of the omissions for the whole test were made on this category, a comparison of the two groups in terms of total omissions would merely reflect the category (5) difference and throw no light on the possibility that the H.E. group tends to omit more items overall. This difficulty can be overcome if the total omissions on a similar test (i.e. Intermediate D) are compared, and relevant figures for such a comparison are given below.

**TABLE XXI**

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>33</td>
<td>25</td>
</tr>
<tr>
<td>I.Q. (Int.D)</td>
<td>102.3</td>
<td>105.8</td>
</tr>
<tr>
<td>W/A (Int.D)</td>
<td>39.4</td>
<td>32.1</td>
</tr>
<tr>
<td>Omissions(Int.D)</td>
<td>7.3</td>
<td>4.6</td>
</tr>
<tr>
<td>(Raw scores)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E. diff. (omissions)</td>
<td>1.37, CR 1.97 t for 56 df at 0.05 = 2.00.</td>
<td></td>
</tr>
</tbody>
</table>
The H.E. group on category (5) shows a greater tendency to omit items of all types and while the difference is on the borderline of significance, it should be noted that the groups are not equal on Intermediate D, I.Q. or W/A, and these inequalities (although not significant) reduce the value of the figures obtained for 'omissions'. These results may be taken as suggestive but by no means conclusive.

One further possibility arises from a comment of Thurstone's in his interpretation of the fact that a mixed sentences test had a loading of 0.46 on P - perceptual speed - "...the best subjects sense the meaning of each sentence without explicit rearrangement of the word order." Although there is no measure of perceptual speed available in the tests used, a related test which may have some bearing on "rearrangement", is the P.M.A. test of 'spatial relations', and it is possible that the H.E. group on (5) would score poorly on the P.M.A. (S) when compared with the L.E. group, for one source of error may be an inability to visualize the words in different positions.

---

Far from supporting the possibility mentioned above, the figures show a trend in the opposite direction, although it is not a significant difference, and furthermore, the groups are not equated on 'Q'.

The analysis of the error distributions for each item is slightly different in the case of (5), because there are no distractors.

### TABLE XXII

COMPARISON OF CATEGORY (5) H.E. AND L.E. GROUPS ON P.M.A. Q. and S.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>P.M.A. (Q)</td>
<td>102.6</td>
<td>108.7</td>
</tr>
<tr>
<td>(S)</td>
<td>44.61%ile</td>
<td>38.7%ile</td>
</tr>
</tbody>
</table>

### TABLE XXIII

DISTRIBUTION OF SPECIFIC ITEM ERRORS FOR CATEGORY (5).

<table>
<thead>
<tr>
<th>Item</th>
<th>D</th>
<th>W</th>
<th>S</th>
<th>O</th>
<th>T</th>
<th>8</th>
<th>L</th>
<th>Wrote Sentence</th>
<th>Total Om's Er's</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 HE</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>30 13</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1 5</td>
</tr>
<tr>
<td>The 26</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26 6</td>
</tr>
<tr>
<td>28 HE</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 0</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 HE</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td>18 41</td>
</tr>
<tr>
<td>LE</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>Th</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 HE</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>42 2</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5 1</td>
</tr>
<tr>
<td>53 HE</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>28 8</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>1 8</td>
</tr>
<tr>
<td>60 HE</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>23 7</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 4</td>
</tr>
</tbody>
</table>
Of the total errors and omissions taken together, omissions accounted for 69.1% in the H.E. group, and 22.0% in the L.E. group. In both groups the greatest percentage of errors (H.E. 53% of total, L.E. 44% of total) occurred on item 32, and of omissions (H.E. 24% of total, L.E. 56% of total) on item 49. At this stage the interpretation is centered around differences between the two groups, so that the problem of explaining the different reactions of both groups to items 32 and 49 must await a fuller analysis. It does appear, however, that whereas the H.E. group error ratio score is mainly due to avoidance of the question (69.1% omissions), the L.E. group error ratio score is due not so much to avoidance (22.0% omissions) as to erroneous attempts.\(^{19}\)

When the actual errors for the two groups are compared, it can be seen that most of the errors occur within a limited range of answers - e.g. -

Item 12: D and T - 62% of H.E. errors, 60% of L.E.
Item 52: R and G - 56% of H.E. errors, 71% of L.E.
Item 53: B, C and T - 63% of H.E. errors, 63% of L.E.
Item 60: 12 and 5 - 57% of H.E. errors, 50% of L.E.

Despite this broad similarity between the groups there are some differences of emphasis, e.g. -

\(^{19}\) There is no way of estimating, from this data, whether the children who omit these items could do them correctly if they tried. This problem will be dealt with in a later section.
Item 12: H.E. D = 6/13, L.E. D = 1/5. The question reads: -
"If the words below were arranged to make a good sentence, with what letter would the last word of the sentence begin?" (Make the letter like a printed capital). usually are made of tables wood .................. ( ).

The answer 'D' may depend upon a very superficial reading of the instructions in which the words 'letter', and 'last word' are prominent, leading to the selecting of 'D'. If this is so, then the difference between the two groups is possibly a difference in attention to instructions, rather than a difference in the ability to rearrange words. This is not supported from the answers to item 32. The question reads: -
"If the words below were rearranged to make a good sentence, with what letter would the third word of the sentence begin?" (Make the letter like a printed capital). honey bees clover gather red from .................. ( ).

The answer most likely to occur with a superficial reading, i.e. ignoring the word 'rearranged', would be 'C', namely the beginning letter of the third word of the sentence as it stands, but whereas only 3/41 (7%) of the H.E. group made this error, 4/14 (29%) of the L.E. group did so. Furthermore, the three answers R, G and C, could all be interpreted as due to superficial reading. One possible explanation of the L.E. group choice of 'C', which would overcome the 'superficial reading' explanation, is that a reasonable rearrangement of the sentence into "from red clover bees gather honey",
would make 'C' the correct answer.

To sum up for category (5): there is some evidence that in the H.E. group, children who omit this type of question may also tend to omit more items overall. Of those who attempt the item, many make errors which are similar for the two groups, although it is difficult to distinguish errors due to faulty reading of the question and those due to faulty rearrangement of the words. The similarities between the patterns of error of the two groups are more outstanding than the differences.

(f) Category (6) Classifying and arranging elements of a concept in a given order. (Items 6, 10, 20, 29, 43, 55, 64, 69, 73).

Example: No. 20.
"Which one of the words below would come first in the dictionary?"
1 tramp, 2 saint, 3 razor, 4 quart, 5 grass, 6 night, 7 paint

and No. 43
"If the following words were arranged in order, with what letter would the middle word begin?"
Eight Ten Six Nine Seven

TABLE XXIV

COMPARISON OF CATEGORY (6) H.E. AND L.E. GROUPS.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Otis II I.C.</td>
<td>109.2</td>
<td>108.1</td>
</tr>
<tr>
<td>W/A</td>
<td>27.1</td>
<td>30.5</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>16 (57%)</td>
<td>15 (54%)</td>
</tr>
<tr>
<td>(Girls)</td>
<td>12 (43%)</td>
<td>13 (46%)</td>
</tr>
</tbody>
</table>
(a) Groups are of similar mean I.Q., with a 3.4% difference in W/A (not significant).

(b) On category (6) the two groups differ significantly (well beyond 1% level).

(c) No sex or socio-economic status differences.

(d) As the essential operation involved is rearrangement of elements and then selection according to instructions, the scores of these groups on category (5) are given. Although the difference is in the same direction as the category (6) H.E. and L.E. groups, it is not significant.

(e) Items in category (10) - number series, bear a superficial similarity to items in this category, although the figures given above show that the small difference is in the opposite direction (and not significant).

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S.E. Status</td>
<td>5.3</td>
<td>5.3</td>
</tr>
<tr>
<td>Error ratio (6)</td>
<td>4.97</td>
<td>1.01</td>
</tr>
<tr>
<td>(5)</td>
<td>5.1</td>
<td>4.4</td>
</tr>
<tr>
<td>(10)</td>
<td>3.08</td>
<td>3.86</td>
</tr>
</tbody>
</table>
### TABLE XXV.

**Distribution of Specific Item Errors for Category (6).**

<table>
<thead>
<tr>
<th>2</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>LE</td>
<td>Ers</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**K, A, B, S, T, G**

<table>
<thead>
<tr>
<th>2</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>LE</td>
<td>Ers</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>76</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Additional Columns:**

- **W %**: 4
- **H %**: 5
- **M %**: 1
- **S %**: 3
- **P %**: 1
- **S %**: 1
- **C %**: 1
- **G %**: 1
- **E %**: 5

---

**Note:** The table represents a distribution of errors for specific categories, with columns detailing the number of errors and omits for different categories and conditions.
The usual pattern of H.E. group errors is apparent - i.e. concentration on a few particular answers (e.g. item 10, 76% on K). As there are very few L.E. group errors, comparison is difficult, although the same general bias as in the H.E. group is evident, e.g. items 10, 29, 43, 55, 69 and 73. Errors in the two groups thus appear to be of the same type.

(g) Category (7) Analogies (Items 4, 7, 11, 15, 19, 22, 26, 30, 34, 38, 42, 46, 50).

Example: No. 19.
"Coal is to a railway engine as what is to a motor car?"
1 motor cycle, 2 smoke, 3 wheels, 4 petrol, 5 horn .......

### TABLE XXVI.

**COMPARISON OF CATEGORY (6) H.E. AND L.E. GROUPS.**

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Otis II I.Q.</td>
<td>106.0</td>
<td>107.9</td>
</tr>
<tr>
<td>W/A</td>
<td>28.9</td>
<td>30.8</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>11 (44%)</td>
<td>18 (72%)</td>
</tr>
<tr>
<td>(Girls)</td>
<td>14 (56%)</td>
<td>7 (28%)</td>
</tr>
<tr>
<td>S.E. Status</td>
<td>5.3</td>
<td>5.5</td>
</tr>
<tr>
<td>Error Ratio (7)</td>
<td>3.04</td>
<td>0.83</td>
</tr>
<tr>
<td>(1)</td>
<td>3.18</td>
<td>4.26</td>
</tr>
<tr>
<td>(2)</td>
<td>1.43</td>
<td>3.50</td>
</tr>
<tr>
<td>(8)</td>
<td>4.40</td>
<td>4.50</td>
</tr>
<tr>
<td>(10)</td>
<td>2.72</td>
<td>5.06</td>
</tr>
<tr>
<td>(11)</td>
<td>3.69</td>
<td>4.67</td>
</tr>
</tbody>
</table>

(a) The two groups are of similar I.Q. and W/A.

(b) Difference between the groups on (7) is signifi-
cant well beyond the 1% level.

(c) There is a clear sex difference, fewer boys proportionately, falling into the H.E. group and more into the L.E. group.

(d) The very small socio-economic status difference is not significant.

(e) Of the five other categories listed, (2) and (10) show significant differences (at 0.01 level in each case), while (1) is on the borderline of significance (at the 0.05 level). It was expected that (8) 'verbal reasoning' might show a difference in the same direction, but the two means are almost identical. Category (2) is 'opposites' and (10) is 'number series', in both cases the direction of the difference being reversed, while (11) - 'arithmetical reasoning' shows a difference in the same direction as (10) although not significant. It is not clear why children who have a high error ratio in analogies should have low error ratio in number series, unless the children find the verbal structure of the analogies confusing and the numerical structure of number series more straightforward. In this case the differences shown in categories (1) and (2) do not make sense. There are obvious dangers in attempting to interpret on the basis of such figures as the above, therefore all that can be said is that apparently there are differences in the way some children react to different types of items, without at this stage attempt-
ing to account for these differences.

Both analogies and number series items occur in the Intermediate D, so that a comparison of the two groups on these categories should reveal any "carry over".

**TABLE XXVII**

**COMPARISON OF CATEGORY (7) H.E. AND L.E. GROUPS ON INTERMEDIATE D SCORES.**

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>Int. D. I.Q.</td>
<td>101.8</td>
<td>100.5</td>
</tr>
<tr>
<td>%/A</td>
<td>40.3</td>
<td>43.4</td>
</tr>
<tr>
<td>(1) Analogies</td>
<td>4.44</td>
<td>4.40</td>
</tr>
<tr>
<td>(4) Number Series</td>
<td>3.43</td>
<td>2.97</td>
</tr>
</tbody>
</table>

Although the two groups were chosen from Otis analogies to represent the extremes in error ratio scores, yet on Intermediate D they have almost identical mean error ratio score; thus there is no carry over, and the difference on Otis analogies may be due to the particular structure and content of the Otis items and not to characteristics of analogies items as such. In number series the small difference is not significant and is not even in the expected direction.

In the analysis of the distractor distributions for each item, boys and girls are shown separately, in view of the fact that there is a sex difference in the percentages falling into each group. (This separation is not made in the L.E. group because of the few errors involved).
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total Om's Er's</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 HE B G T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>7 HE B G T</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>LE</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>11 HE B G T</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td></td>
<td>6</td>
<td>55</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>15 HE B G T</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>LE</td>
<td>5</td>
<td></td>
<td>9</td>
<td>69</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>19 HE B G T</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>22 HE B G T</td>
<td></td>
<td>1</td>
<td></td>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td></td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>26 HE B G T</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td></td>
<td>6</td>
<td>75</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>30 HE B G T</td>
<td></td>
<td>3</td>
<td></td>
<td>5</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>LE</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Cont'd.
(a) The patterns of errors for boys and girls are very similar, with a few exceptions such as item (11) "Hat is to head as thimble is to what?", where of the 11 errors, 8 are made by boys. This difference possible reflects the unfamiliarity of boys with 'thimble', compared with girls. In the same way on item (22), "A motorcar is to wagon as a motor cycle is to a what?", of the 9 errors, 6 are made by girls, due possibly to unfamiliarity, or unwillingness to analyse the question and hence the choosing of a super-

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
<th>Om's Er's</th>
</tr>
</thead>
<tbody>
<tr>
<td>34 HE B</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>G</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>16</td>
<td>52</td>
<td>12</td>
<td>39</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 HE B</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>12</td>
<td>46</td>
<td>6</td>
<td>23</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 HE B</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>G</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>13</td>
<td>76</td>
<td>3</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46 HE B</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>G</td>
<td>6</td>
<td>14</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>8</td>
<td>24</td>
<td>21</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 HE B</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>14</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE XXVIII. Cont'd.
ficially related word. On the whole, however, the boys and girls show the same bias in choosing a distractor.

(b) Concentrations of three or more errors, in the L.E. group, occur in the same distractors as the predominant errors made by the H.E. group. What little evidence there is points again to the similarity in types of error made by the two groups. It is noticeable that the only omissions made by the L.E. group occur on item 38 - "A king is to kingdom as a president is to what?". This item may be difficult for New Zealand children just because of lack of knowledge of the functions of a president, and it is interesting to find that some children in the L.E. group, in the face of unknown words, omit the question rather than attempt an answer - (4 omissions to 5 errors), while in the H.E. group many children make some attempt rather than omit - (3 omissions to 26 errors).

(c) As with the other categories, there is evidence of a strong bias towards certain distractors.

(h) Category (8) Verbal Reasoning (Items 36, 37, 45, 48, 51, 57, 62, 63).

Example: No. 37
"If George is taller than Frank, and Frank is taller than James, then George is (?) James?"
1 taller than, 2 shorter than, 3 just as tall, 4 (cannot say which).
Example: No. 45.
In a foreign language, boy = Kolo
good boy = Kolo Daak
The word that means good begins with what letter? .......( )
TABLE XXIX

COMPARISON OF CATEGORY (8) H.E. AND L.E. GROUPS.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>27</td>
<td>25</td>
</tr>
<tr>
<td>Otis II I.Q.</td>
<td>107.2</td>
<td>108.0</td>
</tr>
<tr>
<td>W/A</td>
<td>33.3%</td>
<td>28.5%</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>17 (63%)</td>
<td>14 (56%)</td>
</tr>
<tr>
<td>(Girls)</td>
<td>10 (37%)</td>
<td>11 (44%)</td>
</tr>
<tr>
<td>S.E. Status</td>
<td>5.6</td>
<td>5.5</td>
</tr>
<tr>
<td>Error Ratio (8)</td>
<td>6.94</td>
<td>1.36</td>
</tr>
<tr>
<td>(1)</td>
<td>4.36</td>
<td>4.40</td>
</tr>
<tr>
<td>(7)</td>
<td>1.82</td>
<td>1.93</td>
</tr>
<tr>
<td>(10)</td>
<td>4.67</td>
<td>4.20</td>
</tr>
<tr>
<td>(11)</td>
<td>4.40</td>
<td>4.60</td>
</tr>
</tbody>
</table>

(a) The two groups are not well equated for W/A (difference of 4.8% - not significant) and this reduces the reliability of the results.

(b) Difference on (8) is significant at well beyond 1% level.

(c) A comparison of the two groups on other categories which might be expected to show a difference in the same direction, shows that there are in fact no differences.

In order to compare these groups on related categories in other tests, the scores on P.M.A. (R) and Intermediate D (Verbal Reasoning - (7)) were calculated.
### TABLE XXX

**COMPARISON OF CATEGORY (8) H.E. AND L.E. GROUPS ON P.M.A. AND INTERMEDIATE D SCORES.**

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.M.A. N =</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>Q =</td>
<td>104.0</td>
<td>101.2</td>
</tr>
<tr>
<td>R =</td>
<td>57.7</td>
<td>54.5</td>
</tr>
<tr>
<td>Int. D. N =</td>
<td>24</td>
<td>21</td>
</tr>
<tr>
<td>I.Q. =</td>
<td>102.4</td>
<td>101.9</td>
</tr>
<tr>
<td>W/A =</td>
<td>42.0</td>
<td>39.5</td>
</tr>
<tr>
<td>Error ratio</td>
<td>5.98</td>
<td>4.94</td>
</tr>
</tbody>
</table>

There is only a slight difference in Reasoning (R), while the difference in Intermediate D verbal reasoning, although in the expected direction is not significant. As there is little overlap between the types of items asked in the Otis and Intermediate D, and none at all between Otis and P.M.A. (R), it is not surprising that there is very little or no 'carry over'.

### TABLE XXXI.

**DISTRIBUTION OF SPECIFIC ITEM ERRORS FOR CATEGORY (8).**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>On's Er's</td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>2</td>
<td>6</td>
<td>17</td>
<td>52</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>9</td>
<td>7</td>
<td>64</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In the four items with distractors (36, 37, 48, 51) the two groups show the same bias in the main, although in 36, "The son of my father's sister is my(?)"
1 brother, 2 nephew, 3 cousin, 4 uncle, 5 grandson ......( ) the H.E. group chose (5) relatively more often than the L.E. group. This may be due to the fact that 'grandson' has a superficial resemblance to 'son' in the statement, and while the L.E. group like the H.E. group shows a high percentage of
errors on the more justifiable choice 'nephew', there may be less of a tendency to make a superficial association.

Distractor 4 in items 37, 48 and 51 is "cannot say which," and the H.E. group tended to chose this more often than the L.E. group (30% over all 3 items for H.E. group and 10% for L.E.), although the major distractor choice in each item is the same in both groups. It is unlikely that such a large percentage of the H.E. group chose distractor 4 because it was thought that there is not enough evidence in the statement to warrant a conclusion, but it is quite likely that, as in the case of 'grandson', many chose an easy way out of a difficult question. This raises the question of whether the children who chose 4 were really satisfied with the choice - whether in fact the mere making of a choice reduced or eliminated the 'tension' set up by the item regardless of whether or not it was a sensible choice from the child's point of view.\textsuperscript{20} In other words, how strong is the 'set' to complete an item satisfactorily, and is there anything in particular types of item structure which influences this 'set'?

The other four items, without distractors, show the same major bias for the two groups, except in 57 where there are too few L.E. group errors for any useful comparison. In the H.E. group the spread of errors is much greater, although

\textsuperscript{20}Cf Hamid (see p98 above) who used this explanation to account for the tendency to choose the first distractor.
this may be due to the greater number of errors recorded.

(i) Category (9) Proverbs (Items 33, 59, 65, 70).

Example: No. 33.
There is a saying, "A stitch in time saves nine." This means (?)
1 A little sewing may save nine shillings. 2 It pays to attend to troubles before they get worse. 3 Work hard and save as much as you can. 4 You can save time by sewing. .............( )

This category is unsatisfactory from the point of view of analysis because of the small number of items and their position in the test. So many cases had to be eliminated that the resulting groups were too uneven for the usual comparison. In order to make some comparison possible, further cases had to be dropped and the resulting groups were well above the mean I.Q. for the whole sample.

| TABLE XXXII |
| COMPARISON OF CATEGORY (9) H.E. AND L.E. GROUPS. |

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Otis II I.Q.</td>
<td>115.5</td>
<td>114.9</td>
</tr>
<tr>
<td>W/A</td>
<td>26:5</td>
<td>22.2</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>13 (65%)</td>
<td>11 (55%)</td>
</tr>
<tr>
<td>(Girls)</td>
<td>7 (35%)</td>
<td>9 (45%)</td>
</tr>
<tr>
<td>S.E. Status</td>
<td>5.1</td>
<td>5.3</td>
</tr>
<tr>
<td>Error Ratio (9)</td>
<td>7.5</td>
<td>0.6</td>
</tr>
<tr>
<td>(1)</td>
<td>4.4</td>
<td>3.0</td>
</tr>
</tbody>
</table>

The only interesting figures in the table are the category (1) mean error ratios which show a difference of
1.4 (C.R. 2.8, t for 38 df = 2.71 at 0.01 level). Thus children in the H.E. group tend to make more errors in word meaning than those in the L.E. group, a condition which would be understandable if success in both categories depends upon a varied linguistic background, although the difference cannot be given much weight due to the small size and the non-representative nature of the groups.21

Proverbs occur in the Intermediate D - namely items 27, 34, 40, 46 (i.e. Category (8)) and a check on the performance of the two groups on this category provides the following figures.

<table>
<thead>
<tr>
<th>TABLE XXXIII.</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPARISON OF CATEGORY (9) H.E. AND L.E. GROUPS ON INTERMEDIATE D. SCORES.</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>H.E.</td>
</tr>
<tr>
<td>L.E.</td>
</tr>
</tbody>
</table>

S.E. diff. 0.90, CR 3.1, t for 35 df 2.72 at 0.01 level.

High or low error on 'proverbs' items thus appears to carry over to a similar test, although the actual setting out

21 The category (1) H.E. and L.E. groups showed a difference on category (9) in the expected direction (not significant) of 0.99 on Otis I and 0.54 on Otis II.
of the items is different. Little confidence can be attached to this difference due to the small numbers, and the large discrepancy in the W/A score.

TABLE XXXIV.

DISTRIBUTION OF SPECIFIC ITEM ERRORS FOR CATEGORY (9)

<table>
<thead>
<tr>
<th></th>
<th>1%</th>
<th>2%</th>
<th>3%</th>
<th>4%</th>
<th>5%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>33 HE</td>
<td>8</td>
<td>35</td>
<td>9</td>
<td>39</td>
<td>6 26</td>
<td>3 23</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>59 HE</td>
<td>8</td>
<td>35</td>
<td>8</td>
<td>35</td>
<td>7 30</td>
<td>3 23</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>65 HE</td>
<td>4</td>
<td>19</td>
<td>1</td>
<td>5</td>
<td>14 67</td>
<td>2 9</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 21</td>
</tr>
<tr>
<td>70 HE</td>
<td>10</td>
<td>56</td>
<td>2</td>
<td>11</td>
<td>6 33</td>
<td>2 18</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

There are too few errors made by the L.E. group to warrant comparison. In the H.E. group only 2 of the items show distractor bias, i.e. 65 and 70, the other two having a more even spread of errors.

(j) Category (10) Number Series (Items 17, 21, 47, 67, 74).

Example: No. 21
"One number is wrong in the following series. What should that number be?"
1 7 2 7 3 7 4 7 5 7 6 7 8 7
## TABLE XXXV

COMPARISON OF CATEGORY (10) H.E. AND L.E. GROUPS.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Otis II I.Q.</td>
<td>108.2</td>
<td>106.9</td>
</tr>
<tr>
<td>W/A</td>
<td>32.2</td>
<td>31.8</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>17 (59%)</td>
<td>19 (63%)</td>
</tr>
<tr>
<td>(Girls)</td>
<td>12 (41%)</td>
<td>11 (37%)</td>
</tr>
<tr>
<td>S.E. Status</td>
<td>5.5</td>
<td>5.4</td>
</tr>
<tr>
<td>Error Ratio (10)</td>
<td>8.0</td>
<td>0.6</td>
</tr>
<tr>
<td>(1)</td>
<td>4.0</td>
<td>4.4</td>
</tr>
<tr>
<td>(3)</td>
<td>2.3</td>
<td>3.3</td>
</tr>
<tr>
<td>(7)</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>(11)</td>
<td>3.8</td>
<td>3.9</td>
</tr>
</tbody>
</table>

(a) The two groups show no significant difference in sex or socio-economic status.

(b) On (10) the difference is significant at well beyond the 0.01 level (CR 17.7).

(c) One categories (1), (3) and (7) the differences are in the expected direction - (i.e. H.E. group lower than L.E. group) but are not significant.²² As the differences between H.E. and L.E. groups in categories (1), (3) and (7) on category (10) are not reciprocal (at least at a level of statistical significance) when category (10) H.E. and L.E. groups are compared on (1), (3) and (7), it is not possible

²²The nearest to significance is (3) with a CR of 1.47.
to make a firm statement about the relation of number series to other categories. On the other hand there is at least a suggestion that number series may be a unique and somewhat complex type of item and may be worth studying from a subjective point of view - i.e. the attitudes and methods which arise when an individual is faced with this type of item.

(d) On category (11), arithmetical reasoning, there is little difference between the two groups. This is surprising in view of the very great difference between the H.E. and L.E. groups on number series, yet confirms the suggestion that there is something unique about number series as a type of item.

In Intermediate D there is a number series, category (4) (items 1, 7, 10, 15, 18, 21, 26, 36, 38, 45, 59). Figures for the two groups on Intermediate D are as follows:

<table>
<thead>
<tr>
<th>TABLE XXXVI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COMPARISON OF CATEGORY (10) H.E. AND L.E. GROUPS ON INTERMEDIATE D SCORES.</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Int. D. I.Q.</td>
<td>102.8</td>
<td>102.1</td>
</tr>
<tr>
<td>W/A</td>
<td>40.8</td>
<td>41.7</td>
</tr>
<tr>
<td>Error ratio(4)</td>
<td>2.7</td>
<td>2.1</td>
</tr>
</tbody>
</table>

The difference (0.6) is not significant (CR 1.2) although in the expected direction, hence it appears that the
very great difference between the two groups on the Otis category (10) is due to some characteristic of the Otis number series, rather than to number series as such.23

**TABLE XXXVII**

**DISTRIBUTION OF SPECIFIC ITEM ERRORS FOR CATEGORY (10)**

<table>
<thead>
<tr>
<th></th>
<th>39</th>
<th>35</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Om's Ers's</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>HE</td>
<td>40</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>8</td>
<td>8-7</td>
<td>7-7</td>
</tr>
<tr>
<td>HE</td>
<td>23</td>
<td>66</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>16</td>
<td>57</td>
<td>2</td>
</tr>
<tr>
<td>HE</td>
<td>2</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>14</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>HE</td>
<td>3</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td>10</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>5</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>HE</td>
<td>2</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td>3</td>
<td>15</td>
<td>1</td>
</tr>
</tbody>
</table>

23On P.M.A. number, the two groups scored:—H.E.Q 103.5, N. 57; L.E. Q 101.0, N. 49, a difference (not significant) in the opposite direction to the category (10) difference. The re is, however, little in common between the two tests.
In the first four items the H.E. group errors show a considerable bias, especially on item 17, while in the L.E. group the only item with at least three errors shows the same bias as the H.E. group.

(k) Category (11) Arithmetical Reasoning (Items 13, 39, 56, 71, 72).

Example: No. 56
"If a man walked east from his home for 7 miles and then walked west 4 miles, how many miles is he from home?"

TABLE XXXVIII

COMPARISON OF CATEGORY (11) H.E. AND L.E. GROUPS.

<table>
<thead>
<tr>
<th></th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Otis II I.Q.</td>
<td>108.5</td>
<td>107.6</td>
</tr>
<tr>
<td>W/A</td>
<td>31.5</td>
<td>29.8</td>
</tr>
<tr>
<td>Sex (Boys)</td>
<td>14 (50%)</td>
<td>20 (71%)</td>
</tr>
<tr>
<td></td>
<td><em>(Girls)</em></td>
<td></td>
</tr>
<tr>
<td>S.E. Status</td>
<td>5.5</td>
<td>5.6</td>
</tr>
<tr>
<td>Error ratio (11)</td>
<td>7.4</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>*(8)</td>
<td></td>
</tr>
</tbody>
</table>

(a) Apart from the very great difference between the two groups on category (11), the only figures of interest are those for sex - fewer boys falling into the H.E. group, and fewer girls into the L.E. group.

(b) On a comparable category - i.e. (8) verbal reasoning, there is little difference between the groups, although in both cases the items are couched in verbal terms, and at this stage it is not possible to decide how many failures on
(11) were due to failure to understand the meaning of the instructions.

On the similar category in Intermediate D (6) arithmetical reasoning \((6,17,29,33,44,53,61,64,66,68,70,73,75)\), the two groups scored 6.0 and 5.8 respectively \((I.Q.'s \ 102.6 \text{ and } 101.1; \ W/A 40.4 \text{ and } 42.4)\). There is apparently no 'carry over' to Intermediate D.

**TABLE XXXIX**

**DISTRIBUTION OF SPECIFIC ITEM ERRORS FOR CATEGORY (11)**

<table>
<thead>
<tr>
<th>Category</th>
<th>6</th>
<th>288</th>
<th>24</th>
<th>48</th>
<th>E1-4</th>
<th>28</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>6</td>
<td>46</td>
<td>2</td>
<td>15</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE(G)</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>8</td>
<td>22</td>
<td>3</td>
<td>25</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td>9</td>
<td>2</td>
<td>5</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>HE(G)</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>29</td>
<td>81</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td>7</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE(G)</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE(G)</td>
<td>50</td>
<td>20</td>
<td>1</td>
<td>25</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>7</td>
<td>50</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>HE(G)</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

**Note:** The table above shows the distribution of specific item errors for Category (11). The data includes percentages for each error type and counts for the total number of errors.
Where comparison is possible (e.g. items 39, 56) the
two groups show the same general bias. In the H.E. group,
girls and boys also show the same general bias, hence the sex
difference is not reflected in type of error made. It is
noticeable that on the easiest item, in terms of position,
there are 8 omissions (against 13 errors), a situation which
may indicate a lack of confidence on the part of some children
when faced with a straight-forward arithmetical problem.

III. - SEX DIFFERENCES OVER ALL CATEGORIES.

In the total sample from which the various H.E. and
L.E. groups were drawn, the percentage of boys to girls was
59% to 41%. Below, is given a table of the percentages of
boys to girls in each H.E. and L.E. group over the eleven
categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>H.E.</th>
<th>L.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B %</td>
<td>G %</td>
</tr>
<tr>
<td>1</td>
<td>66</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>61</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>53</td>
<td>47</td>
</tr>
<tr>
<td>4</td>
<td>63</td>
<td>37</td>
</tr>
<tr>
<td>5</td>
<td>65</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>43</td>
</tr>
<tr>
<td>7</td>
<td>44</td>
<td>56</td>
</tr>
<tr>
<td>8</td>
<td>63</td>
<td>37</td>
</tr>
<tr>
<td>9</td>
<td>65</td>
<td>35</td>
</tr>
<tr>
<td>10</td>
<td>59</td>
<td>41</td>
</tr>
<tr>
<td>11</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
In the H.E. groups, categories (7) and (11) show the greatest differences while in the L.E. groups, categories (3), (7), (11), (2) and (5) in that order. Because of the small size of the groups upon which these percentages were based (mean = 29, excluding category (9) because of the difficulties in selecting H.E. and L.E. groups) the differences are not likely to be statistically significant, in fact by the chi square test for percentages the only significant difference is on category (3) in the L.E. group ($P/2 = .01$), and although the difference on the H.E. group is in the expected direction, it is not significant. Two other categories which show a considerable, but not significant, difference in the L.E. groups (and confirming reversal of differences on the H.E. groups), are (7) and (11).

As a check on these differences, which suggest that on these three categories the girls tend to make more errors than boys, the mean differences between the mean error ratios of all boys and girls in I.Q. groups 2-7 were calculated. Table XLI gives these mean differences for Otis I and II separately, then summed in order to reduce inconsistent differ-

---


25 There were too few girls in Group I to include this group.
ences and emphasise differences in the same direction. In each case the figure given represents the higher error tendency for either boys or girls.

**TABLE XLI**

**MEAN DIFFERENCES BETWEEN MEAN ERROR RATIO FOR BOYS AND GIRLS.**

<table>
<thead>
<tr>
<th>Categ.</th>
<th>Otis I B</th>
<th>Otis I G</th>
<th>Otis II B</th>
<th>Otis II G</th>
<th>Summed B</th>
<th>Summed G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.05</td>
<td></td>
<td>.20</td>
<td></td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.71</td>
<td>.48</td>
<td>.18</td>
<td>.85</td>
<td>.53</td>
<td>1.33</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>.20</td>
<td></td>
<td>.25</td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td>4</td>
<td>.33</td>
<td>.30</td>
<td></td>
<td>.63</td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td>5</td>
<td>.25</td>
<td>.23</td>
<td></td>
<td>.48</td>
<td></td>
<td>.36</td>
</tr>
<tr>
<td>6</td>
<td>.05</td>
<td>.11</td>
<td>.25</td>
<td>.30</td>
<td></td>
<td>.16</td>
</tr>
<tr>
<td>7</td>
<td>.31</td>
<td>.25</td>
<td>.15</td>
<td>.20</td>
<td></td>
<td>.89</td>
</tr>
<tr>
<td>8</td>
<td>.27</td>
<td>.51</td>
<td></td>
<td>.38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There is some evidence from these mean differences to support the results of the H.E. and L.E. percentage differences on categories (3), (7) and (11) - however, as the methods of selection and analysis were mainly directed at the problem of comparing types of errors made by H.E. and L.E. groups rather than sex differences, these results cannot be regarded very seriously. A further check can be made on types of errors made by boys and girls on these categories, when the error choices of the whole group on each item are listed.
IV. SUMMARY OF THE ANALYSIS OF H.E. AND L.E. GROUPS.

A method was devised for picking out a high error and a low error group for each category in which the extreme deviation did not depend on I.Q. or overall error tendency. In each category the groups were checked for I.Q. and W/A, and comparisons were made on variables which, it was thought, might throw some light on the causes of the deviations.

The results are mainly negative.—

(a) Socio-economic status ranks did not differ significantly for the two groups in any of the categories, in fact in most cases the difference was negligible.

(b) Although in many cases the groups showed a similar or opposite deviation on other categories (of varying degrees of statistical significance), these were not consistent enough to warrant interpretation. The only categories which showed a consistent, statistically significant difference were (1) on (4), and (4) on (1), i.e. word meaning and forming a concept from given examples. In this case the H.E. group on one category had a lower mean error ratio than the L.E. group, when both were compared on the other category. It is difficult to suggest a reason for this reversal as both categories are verbal in form, and the different structure of the items does not give a clear contrast. Possibly children in the L.E. group on "word meaning", approach the "forming a
concept" items with a disposition to look for meanings or synonyms for the individual words given and thus tend to overlook the implications of the instruction "which one... is most like these three."

Another category which showed a significant difference on three other categories (with direction of difference reversed) was (10) number series. There may be particular characteristics of this type of item which predispose some children to make a type of error not made on other categories. A more detailed analysis of the actual errors made may throw some light on this problem.

(c) Although there were very great differences between the H.E. and L.E. groups on the Otis categories, there were few significant differences on similar categories in Intermediate D, and the P.M.A. subtests. This suggests that the differences may have been due mainly to features of the Otis items and that the errors made may have a close relationship to not only the content, but the form of the items.

(d) On the whole, the two groups tended to make the same type of error, for although the comparisons of distractor choices or responses could not be very accurate (because of the small number of errors in the L.E. groups) the "bias" on particular responses was usually similar in the two groups, where comparisons were possible.
(e) One of the interesting outcomes of this analysis was the confirmation that on many items, errors cluster around one or two particular responses. This points to the need for an analysis of the error distributions of the total sample, and a more intensive analysis of the particular processes used by children in arriving at incorrect answers.

(f) A slight tendency for girls to make more errors than boys on three categories was noted.
CHAPTER VI.

INCONSISTENCY OF ERROR RATIOS, AND 'CAREFULNESS-CARELESSNESS' RATINGS IN RELATION TO ERROR.

In the outline of methods of analysis given in chapter III (p130), two questions were posed, namely - "Are there cases of marked inconsistency in error ratios from Otis I to Otis II?" and "How do children rated at the extreme of the scale (i.e. of carefulness-carelessness) compare on I.Q., errors, omissions etc.?" This chapter is concerned with the presentation and discussion of evidence relevant to these two questions.

I.- INCONSISTENCY OF ERROR RATIOS.

The preceding analyses were based upon data drawn from the results of two Otis tests, and the deviations for each category were classified as high or low only if they were consistent from one test to the next. In quite a number of cases the deviations were excluded because of inconsistency, and although these cases do not contribute to the type of analyses carried out above, they may reveal some features of interest in themselves, and should therefore be examined. To this end the table of deviations on each record sheet (see specimen in Appendix D) was inspected and where marked inconsistencies occurred, sex, category, Otis I and II I.Q.,
error ratio (unadjusted) and omissions were noted. 1

Out of the 248 children who completed Otis I and II, 108 or 43.5% showed a marked inconsistency in one or more categories. The total number of inconsistencies noted was 147, this figure being made up in the following manner:-

<table>
<thead>
<tr>
<th>Category</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of inconsistencies</td>
<td>13</td>
<td>5</td>
<td>15</td>
<td>9</td>
<td>25</td>
<td>19</td>
<td>16</td>
<td>18</td>
<td>5</td>
<td>14</td>
<td>8</td>
</tr>
</tbody>
</table>

One figure that stands out is the frequency of 25 for category (5) - Mixed Sentences.

As the purpose of this selection was to examine the extent and nature of inconsistency rather than to compare categories, the procedure adopted was to classify the inconsistency in terms of increase/decrease of error ratios and I.Q.'s over Otis I and II.-

- e.g.- a particular boy increased his I.Q. from 97 to 104, while on category (10) his error ratio was 0/3 for Otis I and 4/4 for Otis II, i.e. increased I.Q., increased error ratio; thus although he attempted three items in category (10) in Otis I and made no mistakes, on Otis II he attempted 4 items and got them all wrong despite the fact that his I.Q. had increased by 7 points.

In the classifications set out below, typical examples have been given to illustrate the nature and extent of the

1Deviations were classified as "marked" if there was a positive deviation from the I.Q. group mean on one adjusted error ratio score and a negative deviation on the other, where at least one (or both) of the adjusted deviations was equal to or greater than the M.V.
inconsistencies.

(1) Same or almost the same I.Q. (within 2 points)

(a) Error reduced markedly (N=6)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Otis I I.Q.</th>
<th>Otis II I.Q.</th>
<th>Categ.</th>
<th>Error Ratio I</th>
<th>Error Ratio II</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>135</td>
<td>133</td>
<td>9</td>
<td>4/4</td>
<td>0/4</td>
</tr>
<tr>
<td>B</td>
<td>108</td>
<td>110</td>
<td>5</td>
<td>6/6 (6 om's)</td>
<td>1/5</td>
</tr>
</tbody>
</table>

(b) Error increased markedly (N=30)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Otis I I.Q.</th>
<th>Otis II I.Q.</th>
<th>Categ.</th>
<th>Error Ratio I</th>
<th>Error Ratio II</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>107</td>
<td>107</td>
<td>4</td>
<td>0/4</td>
<td>4/5 (3 om's)</td>
</tr>
<tr>
<td>B</td>
<td>107</td>
<td>108</td>
<td>7</td>
<td>3/13</td>
<td>6/13</td>
</tr>
<tr>
<td>B</td>
<td>113</td>
<td>114</td>
<td>8</td>
<td>0/8</td>
<td>5/8</td>
</tr>
<tr>
<td>B</td>
<td>92</td>
<td>93</td>
<td>5</td>
<td>2/4 (1 om'n)</td>
<td>6/6 (4 om's)</td>
</tr>
<tr>
<td>G</td>
<td>117</td>
<td>118</td>
<td>7</td>
<td>0/13</td>
<td>4/13</td>
</tr>
</tbody>
</table>

(2) Higher I.Q. on Otis II (3 or more points difference)

(a) Error reduced markedly (N=44)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Otis I I.Q.</th>
<th>Otis II I.Q.</th>
<th>Categ.</th>
<th>Error Ratio I</th>
<th>Error Ratio II</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>114</td>
<td>125</td>
<td>8</td>
<td>4/5 (3 om's)</td>
<td>1/8</td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>106</td>
<td>2</td>
<td>3/3 (1 om'n)</td>
<td>0/4</td>
</tr>
<tr>
<td>B</td>
<td>115</td>
<td>118</td>
<td>9</td>
<td>4/4 (1 om'n)</td>
<td>0/4</td>
</tr>
<tr>
<td>B</td>
<td>105</td>
<td>110</td>
<td>5</td>
<td>5/5 (3 om's)</td>
<td>2/6 (2 om's)</td>
</tr>
<tr>
<td>G</td>
<td>91</td>
<td>96</td>
<td>3</td>
<td>6/8</td>
<td>2/8</td>
</tr>
</tbody>
</table>

(b) Error increased markedly (N=37)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Otis I I.Q.</th>
<th>Otis II I.Q.</th>
<th>Categ.</th>
<th>Error Ratio I</th>
<th>Error Ratio II</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>113</td>
<td>116</td>
<td>1</td>
<td>0/6</td>
<td>5/8</td>
</tr>
<tr>
<td>G</td>
<td>95</td>
<td>103</td>
<td>11</td>
<td>0/2</td>
<td>3/3 (1 om'n)</td>
</tr>
<tr>
<td>B</td>
<td>97</td>
<td>104</td>
<td>10</td>
<td>0/3</td>
<td>4/4</td>
</tr>
<tr>
<td>G</td>
<td>91</td>
<td>96</td>
<td>5</td>
<td>2/5 (2 om's)</td>
<td>5/6 (3 om's)</td>
</tr>
<tr>
<td>B</td>
<td>109</td>
<td>119</td>
<td>10</td>
<td>1/3</td>
<td>3/4 (1 om'n)</td>
</tr>
</tbody>
</table>

(3) Lower I.Q. on Otis II

(a) Error reduced markedly (N=4)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Otis I I.Q.</th>
<th>Otis II I.Q.</th>
<th>Categ.</th>
<th>Error Ratio I</th>
<th>Error Ratio II</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>111</td>
<td>106</td>
<td>8</td>
<td>3/6</td>
<td>1/7</td>
</tr>
<tr>
<td>B</td>
<td>106</td>
<td>98</td>
<td>9</td>
<td>4/4</td>
<td>1/4</td>
</tr>
</tbody>
</table>
A few further illustrations have been drawn from the figures noted for categories (5) and (10), because of the large number of omissions on (5) and the peculiar relation of the H.E. and L.E. groups of other categories, to (10).

**Category (5)**

<table>
<thead>
<tr>
<th>Sex</th>
<th>Otis I</th>
<th>Otis II</th>
<th>Categ.</th>
<th>Error Ratio</th>
<th>Error Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.Q.</td>
<td>I.Q.</td>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>(a) Error reduced markedly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>110</td>
<td>123</td>
<td>5</td>
<td>5/6</td>
<td>1/6</td>
</tr>
<tr>
<td>B</td>
<td>105</td>
<td>110</td>
<td>5</td>
<td>5/5 (3 om's)</td>
<td>2/6 (2 om's)</td>
</tr>
<tr>
<td>G</td>
<td>104</td>
<td>101</td>
<td>5</td>
<td>6/6 (6 om's)</td>
<td>2/5 (1 om'n)</td>
</tr>
<tr>
<td>B</td>
<td>108</td>
<td>110</td>
<td>5</td>
<td>6/6 (6 om's)</td>
<td>1/5</td>
</tr>
<tr>
<td>B</td>
<td>85</td>
<td>88</td>
<td>5</td>
<td>6/6 (6 om's)</td>
<td>3/6</td>
</tr>
<tr>
<td>(b) Error increased markedly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>106</td>
<td>113</td>
<td>5</td>
<td>1/3</td>
<td>5/6 (5 om's)</td>
</tr>
<tr>
<td>G</td>
<td>111</td>
<td>110</td>
<td>5</td>
<td>1/5</td>
<td>4/6 (2 om's)</td>
</tr>
<tr>
<td>G</td>
<td>106</td>
<td>102</td>
<td>5</td>
<td>1/4</td>
<td>4/5 (2 om's)</td>
</tr>
<tr>
<td>B</td>
<td>102</td>
<td>101</td>
<td>5</td>
<td>5/5 (2 om's)</td>
<td>6/6 (6 om's)</td>
</tr>
<tr>
<td>B</td>
<td>102</td>
<td>98</td>
<td>5</td>
<td>0/5</td>
<td>5/6 (2 om's)</td>
</tr>
</tbody>
</table>

**Category (10)**

(a) Error reduced markedly

<table>
<thead>
<tr>
<th>Sex</th>
<th>Otis I</th>
<th>Otis II</th>
<th>Categ.</th>
<th>Error Ratio</th>
<th>Error Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I.Q.</td>
<td>I.Q.</td>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>(a) Error reduced markedly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>129</td>
<td>142</td>
<td>10</td>
<td>4/5 (4 om's)</td>
<td>1/5</td>
</tr>
<tr>
<td>B</td>
<td>101</td>
<td>120</td>
<td>10</td>
<td>3/3 (1 om'n)</td>
<td>0/5</td>
</tr>
<tr>
<td>G</td>
<td>111</td>
<td>110</td>
<td>10</td>
<td>3/3</td>
<td>0/3</td>
</tr>
<tr>
<td>G</td>
<td>101</td>
<td>109</td>
<td>10</td>
<td>2/3 (2 om's)</td>
<td>0/4</td>
</tr>
<tr>
<td>B</td>
<td>118</td>
<td>125</td>
<td>10</td>
<td>3/4</td>
<td>1/5</td>
</tr>
</tbody>
</table>
An inspection of these examples shows that there are some drastic changes in the error ratios, and these are particularly significant where the I.Q. change is in the same direction as the error ratio change. The changes may be due to a number of factors, including the following:

(a) **Chance.** It could be argued that with so few items in each category and over only two tests, the chances of an inconsistency in the error ratios are high. While admitting this possibility it might also be argued that as the items are spread throughout the test at varying levels of difficulty, the chances of, for example, a change from 6/6 to 1/5 are low, even if the child were guessing on all items, and it is not reasonable to assume that guessing plays such a large part in the answers.

(b) **Growth.** During a period of about six months the child's intellectual capacity will have increased, thus the error ratio should change. There is, however, no reason to believe that this change would be magnified in one or two
particular categories only - on the contrary it should be
spread more or less equally over all the categories ("more
or less" because of the different numbers of items and their
positions in the test). Furthermore, the change should be
in the nature of a reduction in the error ratio, and yet many
changes are in the opposite direction - e.g. - B. I.Q. 113/114,
Category (8), 0/8 - 5/8.

(c) Special training. It is possible that during the six
month's period some children had direct training on types of
problems very similar to those in the Otis. This would ex-
plain a reduction in error ratio but it would not explain an
increase. There is no reason to believe that any direct
coaching had been given, because there would be nothing to
gain, as the Otis is not normally given once the child is in
Form I, and the teachers did not know that a second Otis would
be given until the day of testing. Furthermore, the teachers
knew that any testing done was purely for research purposes.

(d) Practice. Some reduced error ratios could be due to the
effects of practice on Otis I, which might result in a learn-
ing situation of the following nature. A type of item may
be omitted or the child may choose an answer which he knows is
not satisfactory. This may occur a number of times during the
test, resulting in either, insight into the nature of the
question or the method of handling it, or the setting up of a
tension as a result of the effort to understand - perhaps akin to an interrupted task or incomplete closure. In the first case the effects of learning may not show until the end of the test, or may not show at all on Otis I because the learning having been established on easy material may not be adequate to cope with the more difficult items in the latter part of the test. In the second case, the incomplete closure may persist consciously or unconsciously after the test and thus create a condition in which learning could take place before Otis II. This could account for a drastic change in error ratio - e.g. B I.Q. 115/118, Category (9), 4/4 - 0/4.

(e) Temporary sets. One of the difficulties involved in making use of the preceding explanations, is the occurrence of increase in error ratios. This may be attributed to such things as fatigue, poor health, emotional disturbance, lack of motivation or a negative attitude towards the test situation, but while these may explain a poor performance on the test as a whole, they do not throw any light on the reasons why one or two particular categories should show marked increases. It seems more likely that the changes are due to attitudes or methods of working which may arise within the test situation itself, and which for convenience may be called "temporary sets". These sets could arise and operate in the following manner.-
1. On reading the first item of a category the child may fail to see the full implications and may distort it to fit in with something familiar. When the second item is attempted the similarity of instructions may result in the application of the already distorted schema, while by the time several items have been attempted a glance at the general outline or form of the item may be sufficient to set the schema in motion. This type of set would be encouraged when, as in the Otis, the child is working to a time limit.2

2. Where the general form of the question remains similar but the instructions change slightly, the child may answer the first item of the category correctly, but may carry over the instructions from this item to other items with slightly different instructions, on the assumption that they are all of the same general type.

3. Having read and attempted a problem and found it difficult to answer, the child may omit it or make only a half-hearted attempt to answer other items of the same type, on the grounds that it is not worth wasting time on this type of item - i.e. a loss of confidence.

4. A modification of 3 would be the case in which a type of item may be attempted in Otis I and found difficult (al-

---

2Cf. the 'task displacement' of Selz, 'subsumption' of Fortes and 'difficulty reduction' of Hildreth. See chapter II pp 68, 71 and 84.
though with some success) while in Otis II a recollection of the difficulty of the particular type of item may result in the whole or a substantial part of the series being omitted.

e.g.- B I.Q. 102/101 Categ.(5) 3/5 (2 om's) 6/6 (6 om's)
G I.Q. 92/84 Categ.(6) 1/5 7/9 (4 om's)
G I.Q. 107/107 Categ.(4) 0/4 4/5 (3 om's)
B I.Q. 106/113 Categ.(5) 1/3 5/6 (5 om's)

These "temporary sets" operating in either Otis I or II may account for inconsistency, but where they operate in both tests the result would be a consistent high error ratio which would place the child in one of the H.E. groups already analysed. They are not themselves fully explanatory, because the question arises as to the conditions which produce such sets - i.e. their relation to temperamental characteristics, motivation, particular forms of item structure, setting out of the item; or characteristic problem solving processes.

II.- RATINGS ON CAREFULNESS-CARELESSNESS IN RELATION TO ERROR.

It is a common characteristic of teachers to refer to their children as "careful" or "careless", and to attribute errors to carelessness or to refer to a child as a "careful worker". While the teachers concerned may not be able to give a very good analysis of the psychological variables which underlie these designations, at the same time they persistently refer to these characteristics of children's work as though they are describing a trait which is both identifiable and
reasonably persistent.³

In view of the possibility of some such factor as carefulness operating in test situations,⁴ and of the teachers' descriptions of a similar trait operating in school work, it was decided to attempt to relate the two in an exploratory manner. This required, firstly, some measure of the "school work" trait and secondly a measure or measures of the "test" factor. As a measure of the former a rating scale was constructed similar to that described by Vernon⁵ and the eight teachers of the Form I classes were given a cyclostyled explanatory sheet with illustrations of the trait to be measured - i.e. a trait represented by carefulness at one extreme and carelessness at the other. Further verbal explanations were given in order to ensure uniformity of approach, and the teachers were then asked to place a tick in one of the boxes along an eleven point scale, for each child in the class.⁶

---

³Informal discussions with a number of teachers confirmed these observations, along with the further observation that many teachers are prepared to pick out children from their classes as representing careful or careless types.

⁴See discussion on relevant factorial studies in chapter II, pp99-100.


⁶The explanatory sheet and an example of a completed form are given in Appendix F.
It was not assumed that teachers could discriminate eleven degrees of the trait, however, it was felt that the rather flexible scale, with descriptions of five points along it, might make it easier for teachers to handle, and also easier for purposes of analysis. The results of the ratings by each teacher are given below:

TABLE XLII
TEACHER RATINGS ON A "CAREFULNESS-CARELESSNESS" SCALE.

<table>
<thead>
<tr>
<th>Teachers</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>G</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>35</td>
</tr>
</tbody>
</table>

Total 17 36 34 46 31 29 31 23 21 12 3 283

% 6.0 12.7 12.3 16.3 10.9 10.2 10.9 8.2 7.5 4.2 1.1

An inspection of these figures indicates a variability in the distribution of ratings among the teachers, as would be expected, so that it is not possible to combine any sets of ratings in terms of 1, 2, 3 etc...11. On the other hand, teacher variability may be reduced by taking a constant proportion from the two ends of each teacher's scale, e.g.
20-25%, and assuming that although, for example, teacher 'D'
spread his ratings over the 11 points more widely than tea-
cher 'G', this was not due to a real difference in the distri-
bution of the trait in the two classes, but merely to a
difference in the way it had been estimated. This assumption
cannot be made with any great degree of confidence, however, and
there is no way of checking it, unless the trait is first de-
 fined in specific terms, and its measurement then validated
against some external criterion.

In order to explore the possibility that this trait may
be related to some aspect of test performance, two contrast-
ing groups were selected - i.e. approximately 20-25% at each
end of the scale. This procedure was followed for two rea-
sons, firstly because of the reduction in teacher variability
by this means, and secondly because although ratings as a whole
may be most unreliable, they are more likely to be reliable
with a clear contrast between the selected points on the scale.
As it was not always possible to cut off exactly 20-25% the
cutting points were modified in the direction of ensuring a
reasonable balance of numbers at each end of the scale for
each teacher. This resulted in the following selection:

Careful : A7, B6, C12, D11, E10, F5, G8, H6, : Total = 65

Careless: A6, B6, C10, D10, E9, F6, G8, H6, : Total = 61

Total : 13 12 22 21 19 11 16 12 : Total = 126
Of these 126 children, 89 had completed both Otis tests, the P.M.A. tests and the Intermediate D, 43 in the "careful" group and 46 in the "careless", and these two groups were used in the comparisons which follow.

The choice of measure upon which to compare the two groups was based on the following hypotheses:-

(a) That the careful group would attempt fewer items; would omit more items (i.e. leaving those where there was some doubt about the answer - or in other words being "cautious") would have a lower error score (i.e. ratio of errors to attempted, or even gross errors).

(b) That the careful group would repeat more item errors from Otis I to Otis II, and that where items were repeated, the actual distractor choice or wrong answer would be more constant than for the "careless" group. This hypothesis was based on the argument that if a child rated "careful" makes an error, it is more likely to be more deliberate than an error made by a child rated "careless".

(c) That the careful group would show less I.Q. variability from Otis I to II due, as in (b), to the more deliberate and cautious approach.

Along with these measures it was necessary to check the equivalence of the two groups in terms of (a) Mean I.Q., for unless the two groups were reasonably similar in mean I.Q. the other measures would be useless because of their positive correlation with I.Q., (b) age, for if the groups were of similar mean I.Q. but differed in mean age, then the higher age group would necessarily need to attempt more items or make fewer errors in order to be of equal I.Q. with the lower age group, (c) sex, as the trait may be related to sex differences in test behaviour or may be linked in the minds of teachers with certain sex characteristics.
The measures used were -

1. Otis II I.Q.
2. Difference between Otis I and II (i.e. for variability).
3. Percentage of total items in test actually covered.
4. Percentage of total items in test attempted.
5. Percentage of number covered which were omitted.
6. Percentage of number attempted which were wrong.
   (These last four all on Otis II).
7. Percentage of items wrong in Otis I which were also wrong in Otis II.
8. Percentage of items wrong in both tests which had identical errors.
9. Intermediate D, I.Q.
10. Intermediate D, percentage of number attempted which were wrong.
11. Number of omissions in Intermediate D.
12. P.M.A., Q Score.
13. P.M.A. number of errors on V, S, R and N (the fluency test could not be scored for errors in the same way as the other tests).
14. P.M.A., number of errors on N alone (on the grounds that this test appears on the surface to be a good indicator of carefulness as contrasted to carelessness).
15. Age as at 16/11/53.
16. Sex.

Comparisons between the two groups on each of the above measures resulted in the following figures:-

<table>
<thead>
<tr>
<th>Measure</th>
<th>Careful</th>
<th>Careless</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otis 1. I.Q.</td>
<td>108.0</td>
<td>110.2</td>
</tr>
<tr>
<td>2. Diff. I and II</td>
<td>+3.3</td>
<td>+3.7</td>
</tr>
<tr>
<td>3. % covered</td>
<td>92.3</td>
<td>93.5</td>
</tr>
<tr>
<td>4. % attempted</td>
<td>88.2</td>
<td>89.9</td>
</tr>
<tr>
<td>5. % omitted</td>
<td>4.4</td>
<td>5.2</td>
</tr>
<tr>
<td>6. % wrong</td>
<td>30.3</td>
<td>28.8</td>
</tr>
<tr>
<td>7. % repeat items</td>
<td>66.0</td>
<td>63.2</td>
</tr>
<tr>
<td>8. % identical errors</td>
<td>54.7</td>
<td>63.3</td>
</tr>
</tbody>
</table>
Apart from 16. - sex, there is only one measure which shows a clear difference, namely 8. - percentage of items wrong with identical errors (difference approaching the .05 level of confidence) but this difference is in the opposite direction to that predicted in the hypothesis! One interpretation would be that careful or careless behaviour as judged overall, does not influence the way a test is approached and worked. Another, and much more probable interpretation is that either teachers' ratings of this trait are invalid; or the trait is not really a consistent and identifiable characteristic of behaviour. The great difference in the sex percentages in the two groups - i.e. many more girls than boys in the "careful" group and vice versa in the "careless" group, suggests that the teachers' ratings may have been based on factors little related to actual performance of school
work but possibly to appearance of school work. There is, of course, the further possibility that if the trait were consistent and if teachers' judgements were valid, the difference might show up in some characteristic of test performance not measured in these particular comparisons - e.g. type of error made.

On the whole this attempt to relate such a factor as "carefulness" to a general trait alleged by many teachers to be strongly present in some children, has proved unsatisfactory, and points to the need for a re-examination of the factor itself in order to judge the appropriateness of its description. If "carelessness" produces error in intelligence test items, and "carefulness" reduces error, then these two factors need to be defined in terms of specific concepts already established, and may in fact turn out to be misleading terms until so defined.
CHAPTER VII.

DISTRIBUTIONS OF WRONG RESPONSES FOR ALL CASES.

This chapter summarises certain measures of error, taking the whole group and using the Otis I records. The measures were calculated in order to answer the following questions.

(1) Is there a tendency for more errors to fall in one particular distractor position, than in others? This question arises from the reports of Hamid and others that the first position was favoured over the others.

(2) Is there confirmation of the observation made in comparing H.E. and L.E. groups, namely, that the wrong responses tended to cluster around particular distractors?

(3) Are the slight sex differences reported in chapter V, reflected in the total error frequencies for boys and girls in each category?

I. DISTRACTOR POSITION.

Each error was classified in terms of distractor position, and the resulting distributions of responses, over all items of the multiple choice type, were summed and averaged. This

1See discussion of Otis II below.

2See above, p98.
involved 3,057 errors over 42 items with 4 possible wrong responses, and 5 with 3 possible wrong responses. Such averages would be meaningless unless allowance were made for the positions of the right responses, and in order to allow for this, the number of errors on any given distractor position was divided by the number of times that position was available as a wrong response. This eliminated the effects of uneven correct response positions throughout the items. Table XLIV gives the means of the distributions and the percentage of errors on each position.

**TABLE XLIV.**

MEAN NUMBER OF ERRORS AND PERCENTAGES OVER FIVE DISTRACTOR POSITIONS.

<table>
<thead>
<tr>
<th>Distractor Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Errors</td>
<td>19.8</td>
<td>17.6</td>
<td>14.4</td>
<td>12.5</td>
<td>19.0</td>
</tr>
<tr>
<td>Percentage</td>
<td>23.8</td>
<td>21.1</td>
<td>17.3</td>
<td>15.0</td>
<td>22.8</td>
</tr>
</tbody>
</table>

3 The percentages given above, were based on total number of errors, and could therefore be distorted if one item had a particularly large number of errors concentrated on one distractor. In order to check the figures given, all error distributions on the 47 items were expressed as percentages for each item, thus ruling out total differences of errors for each item. The resulting proportions for each distractor position were - (1) 24.0%, (2) 20.1%, (3) 18.0%, (4) 14.4%, (5) 23.5% - i.e. allowing for the fact that the original percentages were rounded to the nearest whole number, these figures confirm the percentages calculated from the actual error scores.
It can be seen that there is no clear cut advantage for distractor (1) - for the first, second and fifth positions are all within a narrow range of percentages. The third and fourth positions have a lower percentage of responses, although the greatest difference, i.e. that between the fourth and first - 8.8% is not great unless it is assumed that errors occur at random, or that the only cause of a particular distractor choice is position, in which case it would be reasonable to expect the middle positions to be chosen less frequently because of the greater perceptual distinctiveness of end positions. Although this may be one influence on distractor choice, it is not likely to be a very potent one, as already indicated by the concentrations of errors reported in the H.E. and L.E. groups.

Hamid noted that whereas the dull subjects tended to mark off the first of the alternatives, the bright subjects discovered the correct answer wherever it was placed. It was therefore decided to break up the total group into three broad I.Q. groups and calculate the percentage of errors falling in the distractor positions for each group. The results are given in the following table.

4See above, p98.
TABLE XLV.

PERCENTAGE OF ERRORS IN DISTRACTOR POSITIONS FOR THREE I.Q. GROUPS.

<table>
<thead>
<tr>
<th>I.Q. GROUP</th>
<th>N</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) 114-142</td>
<td>57</td>
<td>26.6</td>
<td>21.9</td>
<td>18.0</td>
<td>15.6</td>
<td>18.0</td>
</tr>
<tr>
<td>(b) 96-113</td>
<td>136</td>
<td>22.8</td>
<td>21.9</td>
<td>18.6</td>
<td>14.0</td>
<td>22.6</td>
</tr>
<tr>
<td>(c) 70-95</td>
<td>55</td>
<td>24.2</td>
<td>19.0</td>
<td>14.7</td>
<td>16.7</td>
<td>25.4</td>
</tr>
</tbody>
</table>

Although the general pattern is not seriously disturbed, (Cf Table XLIV) it appears that there are more first position and fewer fifth position choices for the high I.Q. group, compared with more fifth position and fewer third position choices for the low group. This type of analysis does not, however, have much value, as overall measures such as these do not lend themselves to interpretation in the same way as differences in the performance of I.Q. groups on specific items - such differences will be dealt with in the course of the analysis of individual errors.

II. CONCENTRATION OF WRONG RESPONSES.

In the H.E. and L.E. comparisons it was found that the errors were not spread evenly over the distractors but were often concentrated on one or two particular wrong responses - a characteristic noted by Fortes and Eells\(^5\) and interpreted by

\(^5\)See above pp97-98.
Brigham⁶ in his comment that "errors are merely characteristic solutions of the problems set". As an indication of the extent to which such concentrations of errors on multiple choice items occurred, the distributions of the percentage of errors on each distractor for each item are given in Table XLVI. Altogether there were 47 items and 183 possible distractor choices; N. for each column represents the available number of wrong response distractors in each position.

### TABLE XLVI.

**DISTRIBUTIONS OF PERCENTAGES OF ERRORS ON 47 MULTIPLE CHOICE ITEMS.**

<table>
<thead>
<tr>
<th>Percentage of Errors</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 - 99</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>80 - 89</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>70 - 79</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>60 - 69</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>50 - 59</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>40 - 49</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>30 - 39</td>
<td>8</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>20 - 29</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>10 - 19</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>12</td>
<td>10</td>
<td>43</td>
</tr>
<tr>
<td>0 - 9</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>15</td>
<td>3</td>
<td>46</td>
</tr>
</tbody>
</table>

N | 42 | 32 | 34 | 39 | 36 | 183

These five distributions are not symmetrical, e.g. - if 70% of the errors for a given item are concentrated on a particular distractor, then for this one high percentage there must be three low percentages on the other distractors, hence

⁶Brigham C.C. *A Study of Error*, 1932, (College Entrance Examination Board) p. 84.
the high frequency of the lower percentages. A significant feature of this table is the fact that in 28 of the 47 items, 50% or more of the errors were concentrated on one particular distractor and of these 28, 7 items had concentrations of 70% or more of the errors.

It could, of course, be argued that the test was constructed in such a way that certain distractors were bound to attract many of the wrong responses, and this may be true (although it has important implications for test construction and measurement which will be discussed later). If it is the particular structure of the distractors offered which causes such concentrations, then they should not appear on items of the open answer type. The table of distributions below gives some evidence on this point, being based on 28 items, four items (Category (6) - Classification) which have a wide range of choices: 10, 7, 8 and 7 respectively, and 24 which are of the "open answer" type.

**TABLE XLVII.**
DISTRIBUTIONS OF PERCENTAGES OF ERRORS ON 24 "OPEN ANSWER" ITEMS AND 4 MULTIPLE CHOICE ITEMS.

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 - 99</td>
<td>1</td>
</tr>
<tr>
<td>80 - 89</td>
<td>4</td>
</tr>
<tr>
<td>70 - 79</td>
<td>1</td>
</tr>
<tr>
<td>60 - 69</td>
<td>0</td>
</tr>
<tr>
<td>50 - 59</td>
<td>1</td>
</tr>
<tr>
<td>40 - 49</td>
<td>12</td>
</tr>
<tr>
<td>30 - 39</td>
<td>10</td>
</tr>
<tr>
<td>20 - 29</td>
<td>13</td>
</tr>
<tr>
<td>10 - 19</td>
<td>41</td>
</tr>
<tr>
<td>0 - 9</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>183</strong></td>
</tr>
</tbody>
</table>
Although the mean number of wrong responses for these 28 items was 6.5, it is clear that many of these responses on a particular item were given by only a few children, as indicated by the frequency of 100 for the 0-9% interval. More significant in view of the discussion above, is the fact that 19 of the 28 items have 40% or more errors concentrated on one wrong response, while for 6 of these 19, the percentage is over 70. In other words, the characteristic error solutions do not depend merely on the distractors offered, but also upon the actual question asked, or at least the interpretation of that question.

III.- SEX DIFFERENCES.

As a check on the small sex differences noted in the analyses of H.E. and L.E. comparisons, the total errors for boys and girls were calculated - boys representing 59% of the sample accounted for 57.5% of all errors, and girls (41% of sample) for 42.5% of all errors. The comparative percentages for each of the categories are given in the following table.
TABLE XLVIII.

PERCENTAGE OF TOTAL ERRORS ON EACH CATEGORY FOR BOYS AND GIRLS.

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage Boys</th>
<th>Percentage Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>58.8</td>
<td>41.2</td>
</tr>
<tr>
<td>(2)</td>
<td>63.2</td>
<td>36.8</td>
</tr>
<tr>
<td>(3)</td>
<td>53.9</td>
<td>46.1</td>
</tr>
<tr>
<td>(4)</td>
<td>57.6</td>
<td>42.4</td>
</tr>
<tr>
<td>(5)</td>
<td>61.1</td>
<td>38.9</td>
</tr>
<tr>
<td>(6)</td>
<td>58.9</td>
<td>41.1</td>
</tr>
<tr>
<td>(7)</td>
<td>56.5</td>
<td>43.5</td>
</tr>
<tr>
<td>(8)</td>
<td>55.4</td>
<td>44.6</td>
</tr>
<tr>
<td>(9)</td>
<td>58.1</td>
<td>41.9</td>
</tr>
<tr>
<td>(10)</td>
<td>61.8</td>
<td>38.2</td>
</tr>
<tr>
<td>(11)</td>
<td>54.8</td>
<td>45.2</td>
</tr>
</tbody>
</table>

Of the four categories where girls made more than 42.5% of the errors, three correspond to the categories noted in the comparisons of H.E. and L.E. groups - i.e. (3), (11) and (7), however, the difference for (7) is negligible. A more fruitful approach is the examination of sex differences in individual items, for example on item 24, girls made only 26.8% of the errors while on item 14 they were responsible for 59% of all errors. Such differences can be dealt with when the errors on specific items are being discussed.

IV.- COMPARISON OF OTIS I AND OTIS II.

As a check, the Otis II error distributions were calculated for a selection of items representing both multiple and open answer types (7 of each), some with very marked con-
centration on one response and some with a more even spread of errors over two or more responses, and covering a range of categories.

This check was made in order to avoid the very lengthy procedure of calculating Otis II error distributions in the same way as was done for Otis I. A complete analysis of Otis II was not warranted, (a) because of the indication in the H.E. - L.E. comparisons (in which both Otis I and Otis II errors were originally listed) of the similarity of the error distributions, and (b) because of the limited value of the overall figures derived from such an analysis.

In only one of the 14 items checked was there a discrepancy in the general pattern of error frequencies, i.e. in item 53, a mixed sentence item of the open answer type:


Although the percentages vary over the 1st, 3rd and 4th answers, in both tests three responses account for 70% or more of the errors.

From this check, and the experience with Otis I and II errors on the H.E. - L.E. comparisons it is safe to assume that in general, the error distributions for the two tests are very similar.
SECTION C.

ANALYSES BASED ON INTERVIEWS WITH CHILDREN AND SUMMARY AND CONCLUSIONS.
CHAPTER VIII.

INDIVIDUAL ANALYSES OF ERRORS:
SELECTION OF GROUP AND PROCEDURE.

In chapter III a broad outline of the methods to be followed was presented, and reference was made to the questioning of individual children in order to throw some light on such problems as:- "How did this child arrive at this particular error on this item, as suggested by his own explanation and by relevant questions? What is his attitude towards his own error?" Before presenting the results it is necessary to discuss the basis of selection of subjects, the items to be re-worked, the method of re-working, conditions and rapport, and the most effective method of presenting the results.

I.- SELECTION OF SUBJECTS.

The criteria set up for the selection of a sample were-

(i) children who had taken both Otis tests,

(ii) a wide range of I.Q. so that differences in types of errors between I.Q. groups would have a chance of being exposed; also a "rectangular" distribution of subjects over the 7 I.Q. groups, for the same reason,

(iii) an equal number of boys and girls in order to expose any sex differences,

(iv) children who would co-operate and who were not nervous in an "interview" situation,
(v) a large enough number of children to allow for the fact that many items previously wrong would be answered correctly.

In the light of these criteria it was decided to choose 6 boys and 6 girls from each of the 7 I.Q. groups (using Otis II I.Q.) - i.e. 84 children in all.\textsuperscript{1} Apart from the criteria listed, the only basis of selection was availability (e.g. some children had cafeteria duties etc.), and where several children of the required I.Q. group and sex were available, those selected were picked at random.

II.- SELECTION OF ITEMS.

Originally it was intended to re-work only those items which were wrong or omitted on both Otis tests, but after preliminary trials, it was found that this restriction cut out useful information from those cases where an item was wrong or omitted on only one of the Otis tests, and yet wrong when re-worked during the interview. As a result, all items which were wrong or omitted on one or both of the Otis tests were re-worked, together with a few correct items in order to provide a check and also to reduce the possibility of a series of difficult items (for a particular child) resulting in a loss of confidence.

\textsuperscript{1}Thus there were 12 children in each of the following I.Q. groups - 123+; 117-122; 111-116; 105-110; 99-104; 93-98; 92-.
III. - METHOD OF RE-WORKING.

In the preliminary trials it was found that merely asking the child to work the question aloud was quite inadequate, because it meant that the child had to try and formulate his thoughts in words, as well as carry through a type of introspection. On the other hand questioning could result in the asking of leading questions, perhaps unintentionally from the point of view of the interrogator, or even misinterpreted as such by the child. As a compromise the following plan was adopted as a general guide:

(a) The child worked the question silently and then gave his answer. In no case was any adverse comment made on the answer, although occasional encouragement was given.

(b) Following this, the child was asked to explain how he got his answer, and although some children gave very full and insightful accounts of the steps leading up to the answer, most had to be prompted with such questions as "why did you decide to do that...or choose that?"

(c) If the explanation was inadequate, or if a check was needed, the child was asked to read the question aloud - in some cases where this was of little help in distinguishing the source of the error, the question was read aloud for him. It was characteristic of children who, after rereading a question, saw that they had made a mistake, to comment spontaneously e.g. "Oh! - I thought etc. etc...." - and one of the difficulties
in presenting written reports of these interviews, is the impossibility of recording the signs of insight shown by the children e.g.: smiles or nods, drawing in of breath, dropping of the jaw in surprise etc.

(d) Where still further information was required, questions were asked such as:- "Which is the important part of the question?...Where does it tell you what to do?...What is worrying you about this question?...What is the meaning of this word?...Are you quite sure that this is the best answer?" These questions were also asked on items where there were no errors, so that the children would not associate them with failure.

(e) As a check on the interpretation of the type of error involved, some hints were given in a few cases - eg. "What do you think is the connection between this and that...?"

IV.- CONDITIONS AND RAPPORT.

Conditions were ideal for most of the interviews, as a small sitting room in the homecraft "flat" was made available; when this was not available, an empty classroom was used. The average time taken for each interview was 1\(\frac{1}{4}\) hours - the maximum 1\(\frac{3}{4}\) hours, the minimum 25 minutes (i.e. for a few of the brightest children). As a rule it was found possible to interview only 4 children a day, because of the nature of the school timetable, so that the 84 children were covered over four school weeks.
Rapport was good, as far as could be judged, because the children were used to seeing the investigator at the school - e.g. in the classrooms testing during the year, occasionally at morning assembly, in the playground etc. In order to ensure that there was no anxiety, each child was called for, in his classroom, and as he walked over to the "flat" he was encouraged to talk about his interests etc. The first items worked were always easy ones on which he had made no mistakes, there was no sense of pressure to complete within a given time, and it was made clear that the interview was not an examination of the child, but something to help the investigator. There was, in fact, quite a lot of enthusiasm and eagerness to be chosen, so that there is no reason to believe that the children did not co-operate fully. In only two cases was there shyness or anxiety which interfered with the interviews, and the records of these children were not used, two other children being interviewed in their places.

These details of rapport have been given, because the success of this method depends almost entirely on the atmosphere of the interview and the way in which the questions are asked. While such a method can give valuable insights into the processes operating to produce error, it shares the defects of all such methods - e.g. subjectivity, and to this extent the results need to be interpreted with caution.
V.—METHODOLOGY OF PRESENTING THE RESULTS.

The interviews resulted in the accumulation of a considerable amount of material, and difficulties arose in deciding how to present this material most effectively, the following methods being considered:

(a) A straightforward tabulation of the types of errors noted. This method would fail to convey the authenticity of the first-hand reports of the children's explanations. Furthermore, it would not show the variations, for a given type of error, from item to item.

(b) Full reports of each interview. Apart from the unwieldy nature of this method, it would not result in a workable clarification and classification of recurrent errors.

(c) The presentation and illustration of types of errors for each item (by categories) with a summarised classification for each category. This method would have the advantage of showing as clearly as possible the sort of working records upon which classifications were based.

Method (c) was chosen as being most likely to order the evidence presented into convenient groupings, without losing any of the authenticity of first-hand reports. Unfortunately, this resulted in a very lengthy documentation of source material, which while too detailed for inclusion in the body of the thesis, is yet necessary as illustrating the variety and pertinency of the explanations recorded. Accordingly, this material is presented as Appendix H, while a summary of the types of errors found is given in the next chapter.
CHAPTER IX.

INDIVIDUAL ANALYSES OF ERRORS: RESULTS AND DISCUSSION.

The following classifications of errors are based on types of responses illustrated in appendix H, and the form of presentation, i.e. by categories, follows the plan adopted for the appendix. While this necessarily results in some overlap and repetition from category to category, at the same time it allows for comment on particular characteristics of the errors for a given category or item. As explained in the introduction to appendix H, an attempt to give exact frequencies for the various types of errors noted, would be misleading. It is not claimed that these classifications give a complete picture of the conditions underlying the production of responses to intelligence test items, but at least they illustrate some of the "methodological" aspects of intelligence which function in a test situation.

I.- CLASSIFICATIONS OF ERRORS BY CATEGORIES.

(a) Category (1) - Word Meaning. (See Appendix H. pp 306-316)

In view of the analysis of H.E. and L.E. groups on category (1), it was not expected that the investigation of errors on this category would yield much of significance, over and above evidence that some children "know" fewer words than others.\(^1\) Of all the items it would seem that this type is

\(^1\) Cf. Feifel's analysis of vocabulary errors p96 above.
subject to the straightforward condition - either the child knows the word, or he does not. Although this is a more appropriate explanation in this category than in most others, the analysis revealed that this "either-or" condition is a simplification of the processes involved in choosing a response.

Errors noted were due to the following conditions:

(1) Failure to read the whole sentence, usually in the form of neglect of some essential qualification, ranging from one word to half of the sentence. Also misplaced emphasis on the wording of the sentence.

(2) Failure to give due consideration to all the choices offered, or even to read them all through, e.g.:- seeing a relevant word at or near the beginning of the line, and choosing this without having read the others.

(3) Passing over the correct answer, even although fully understood, in favour of a distractor with particularly potent qualities or associations, which displace analysis of the other words, including the correct one.

(4) Understanding the meaning of the correct answer but rejecting it as wrong in the light of common usage.

(5) Mis-reading, or inability to pronounce a word which was, however, understood aurally.

(6) Failure to distinguish between a word and its opposite - i.e. an amount of understanding not adequate for the correct answering of the item, but considerably in advance of failure to understand the word at all.

(7) Overlooking what is not understood or what complicates the problem.

(8) Adding to the data given in the item, or reconstructing the item to fit in with an assumption about what is required.

This classification indicates that apart from a lack of know-
ledge of the word being tested, and even in spite of a correct understanding of the word, the actual response made, may be due to one or more conditions arising from the child's appraisal of, and approach to the test item. That these conditions are not confined to the lower I.Q. groups, but are evident over a wide range of I.Q.s, is clear from the records in appendix H.

Such a condition as failure to take into consideration all the data, either in terms of the given sentence or the choices offered - may be caused by many factors, e.g.:- inertia, haste, or low level of aspiration, yet the circumstances under which the items were reworked, were such as to reduce these factors to a minimum. It is more likely that this condition could result from the child's learning experiences in a variety of school subjects - e.g.:- being in the habit of assuming that there is only one correct answer to a question, and not having had to choose between partially correct answers. Such an assumption would explain why the sighting of a partially correct or relevant answer, led to the ignoring of other possible answers or interpretations.

Thus, even on such a straightforward type of item as the word meaning items, the error may be due not only to lack of verbal ability - for example, the quantity and quality of the child's vocabulary - but also to the interaction of the particular item structure, the work habits of the child, and
the assumptions he makes about the nature of the item.

(b) **Category (2) - Opposites** (See Appendix H pp 317-320).

Although the structure of these items appears to be quite simple, as in category (1), the errors resulted from a range of conditions influencing the response made:

1. Failure to consider all the choices available, usually coupled with a particularly prepotent distractor. In some cases the distractor was attractive because of its unfamiliarity, even although the correct answer was known.

2. Failure to formulate the nature of the task. Although there were cases of complete failure to understand the item despite the fact that in a more familiar context the correct opposite could be elicited, these were confined to the lowest I.Q. group. A more common form of confusion was the giving of a synonym, i.e. the ignoring of the crucial word "opposite".

3. Misreading and misinterpretation of words.

4. Unfamiliarity of the correct word, although with clear understanding of the given word and the relation required. Most of these responses showed that attempts were made to deal sensibly, with at least those parts of the problem with which the children were familiar.

Although on items 24 and 63 the boys made substantially more errors than the girls (73% and 65%), the re-working did not throw any light on this situation.

That children should persistently misread a simple question such as - "Which word means the opposite of...," even after more than one reading, indicates the extent to which assumptions can distort what is read. It would be easy to class such an error as due to "carelessness", yet when the child is working through a test in which the items are of
different but intermittently recurrent types, it is quite conceivable that the constant restructuring of the task may lead to distortion or incomplete analysis of the instructions.

(c) **Category (3) - Word or Phrase Attribute or Definition.**
(See Appendix H. pp 320-330)

The items in this category are of the "which tells best" or "which is the most important reason" type - thus they are much less straightforward than those of the preceding two categories, and this was reflected in the conditions which produced errors:

(1) Inadequate formulation of the crucial aspect of the task, or reduction of a specific requirement to a more general requirement.

The response made to the item hinges on the interpretation of "tells best" and "most important reason," and with the former phrase, particularly, some children had difficulty in deciding from what point of view "tells best" should be interpreted. They were not clear as to the implications of "tells best," although having made a choice they were prepared to defend it. This overlaps with the second type of error.

(2) Reading a wrong, but not completely excluded interpretation into the sentence.

(3) Rejection of the correct answer on reasonable grounds.

(4) Failure to consider all the available choices.

(5) Misreading and mispronunciation.
Girls made proportionately more errors than boys on this category, although this was mainly due to errors on item 14, a possible interpretation being the tendency for more girls to choose a distractor which was related to the appearance of doors and gates, while the correct answer required an appreciation of the principle relating to doors and gates.

One item (35) demonstrated the dilemma which can result from the inclusion of an unfamiliar word in a phrase, i.e. "a malicious false statement"; for while some children rejected this (the correct answer) because of the unfamiliar word "malicious", others accepted it despite the unfamiliar word.

On the whole it appears that this type of item is too vaguely structured, as the child may make errors which from his point of view, are quite reasonable. The evidence from categories (1) and (2) however, makes it clear that such conditions as failure to consider all the choices, or reformulation of the instructions, do not depend entirely upon the complexity of the item structure, but also on the assumptions and work habits which the child brings to the test situation.

(d) Category (4) - Forming a Concept From Given Examples, (See Appendix H. pp 330-337).

In these items the crucial phrase is "most like", and the child must decide on a basis of classification, then choose another word which can be classified in the same way. Errors
were classified as due to the following conditions:

(1) Failure to clarify the requirements of the problem - e.g. the phrase "most like" was interpreted as "something which goes with"; or a word was chosen because it was a common characteristic of the three given objects. Allied to this was the substitution of a completely different problem due either to perseveration from earlier items of a different category, or to an assumption arising from the arrangement of the words in the problem.

(2) Ignoring a disturbing or unknown element in the data, or failure to use all the essential parts.

(3) Narrowing of the possible field of search by concentrating on one particular characteristic to the exclusion of other characteristics. This type of "false direction" was difficult to break and rather than change the basis of classification chosen, there was a tendency for some children to distort the item or ignore those parts which did not fit.

(4) Superficial reading - as acknowledged by the children themselves when re-working. This underlies many of the conditions listed for all categories, although it is a useful description of the cases where the children, after having given the wrong answer, corrected themselves spontaneously by recognizing that they had misread the problem.

(5) Confusion over the meaning of words used in the items.

Underlying all the specific misconceptions and misinterpretations, is the general condition of failure to understand the structure of the problem, coupled with failure to give due consideration to all the data presented. This has already been illustrated in the earlier categories and it will be seen that it is a characteristic evident in all the category classifications.
(e) Category (5) - Mixed Sentences (See Appendix H, pp 337-348)

It was clear from the manner in which the children reacted to these items that they disliked them and had little confidence in their ability to solve them. This was borne out by the very high rate of omissions in both Otis I and II. Despite this lack of confidence, when asked to attempt the problems, the children in many cases were able to give the correct re-arrangement, although as can be seen from the first condition given below, this was no guarantee that the correct answer would be given.

(1) Correct re-arrangement of the sentence, but failure to express the answer in the specific terms required by the instructions.

This type of error occurred repeatedly, although there was no pressure of time, and the wrong answer was sometimes still given, even after one or more re-readings of the instructions aloud, by the child or investigator. Such persistent failure to follow specific directions cannot be explained in terms merely of "haste" during a timed test. It may, however, spring from the reading and work habits of children, habits built up in working problems where the actual method of expressing the answer is implicit in the structure of the problem. Doubtless under the conditions of a timed test, these habits would be even more likely to appear.
(2) Difficulty in "recentering" - i.e. a lack of flexibility in shifting a word from one context to another, or in breaking up a word combination with a strong "gestalt". Examples of the former were the interpretation of "sum" as a problem in arithmetic instead of as addition, and "letter" as a written communication instead of a unit of the alphabet. A word combination which caused great difficulty was "honey bees" - and many children attempted to rearrange the words in several ways to fit in with this pair, they could not break such a strong combination. On the other hand, a combination which was resisted strongly, was the unusual coupling - "birch bark".

Not only did these difficulties set the children off in the wrong "direction" in attempting to rearrange the sentences, but they resulted in "sets" which were strongly resistant to any restructuring.

(3) Failure to use all the data given i.e. ignoring words in the disarranged sentence, or reading only part of the sentence. Allied to this was the opposite condition of adding to the data given in order to solve the problem of rearrangement.

(4) Failure due to the unacceptibility of the correct solution when this had been produced. There were some very reasonable objections to the sentences, which led to hesitation in choosing an answer or to omission of the item.

(5) Confusion as to the general requirements of the task - e.g. failure to rearrange the sentence coupled with an attempt to answer the specific requirements; attempting to join up the disarranged words with a part of the instructions. Related to this was the displacement of the task by a similar task worked previously, e.g.:- attempting to arrange the words into some sort of classification.

(6) A condition which appeared, along with many of those noted above, was satisfaction with an incomplete or poor solution, even although the child was aware of this limitation of his answer.

Although one obvious cause of error not specifically mentioned above, was inability to make a satisfactory sentence
from the words given, the conditions listed influenced the actual answer given, whether the words were rearranged correctly or not. Thus this type of item, has a much more complex structure, than is apparent on the surface.

(f) Category (6) - Classifying and Arranging Elements of a Concept in a Given Order.

(See Appendix H. pp 349-357)

There were more omissions on this category than on any other, except category (5) - i.e. 13.9% of the total omissions. This may have been due to the similarity of the items to mixed sentences, in that the data had first to be rearranged, and then a specific answer given based on this rearrangement. As would be expected the errors were due to conditions which were very similar to those found in category (5).

(1) Failure to follow the specific instructions of the task - e.g. rearranging the words correctly and then picking the wrong word or letter. As in category (5), some children could not correct this elementary mistake, even after rereading the instructions.

(2) Failure to understand the general requirements of the task - e.g.: attempting to give an answer without rearranging the elements, or using a wrong or irrelevant principle of arrangement.

(3) Mis-interpreting the instructions, e.g.: the words "tell by letter" in the instructions confused children at all I.Q. levels, because whereas the test constructor meant this to indicate the mode of expressing the answer, these children interpreted it as meaning "find out by".

(4) Prepotency of part of the data displacing attention from the rest, and leading to the ignoring of crucial elements,
especially in the middle of a series. Also failure to read all the data.

(5) Losing the "task direction" or distorting the structure of the problem, in concentrating on a detail, or the comparison of two elements.

(6) Correct rearrangement of elements, but with the specific answer expressed in terms drawn from items of a different type.

(7) Misreading, and miscounting, e.g. correct rearrangement followed by miscounting to find the middle word.

(8) Lack of knowledge of the elements, in particular - army ranks.

On the whole it may be claimed that one of the major difficulties in answering items of this type, was not the correct rearrangement of the words, but the translation of the correct order into the specific answer required, despite the fact that the translation was of an elementary nature.

(g) Category (7) - Analogies (See Appendix H. pp 357-370).

Although this is one of the categories on which girls made more errors than boys, this tendency was not consistent over all the items, for although girls made 55% of all the errors on item 38, they made only 31% of the errors on item 11. The latter item, is based on the use of a thimble on the finger for sewing, while the former, requires the understanding of the position of "president" in a country. Girls would be more likely to relate thimble-finger, although it is not clear why boys should have had less difficulty in relating
"president-republic". Much more important than these sex differences, was the common tendency for both boys and girls, to misinterpret the nature of these analogies items. This major source of error is discussed in the first condition listed below.

(1) Failure to understand the structure of the items, in particular, failure to appreciate the significance of the illustration in the first half of the sentence, and the crucial words "is to a".

Although analogies items can be constructed to test a carefully controlled series of relations, it is essential for the child to realize what the specific structure of the item implies, otherwise the whole point of the item is lost. If this failure was confined to the lower I.Q. groups, then it could be argued that appreciating the structure of the item is a good criterion to use. That this is not necessarily the case, is illustrated in the records presented, for many children, at all levels, reformulated the item into a different, and less exacting structure.

The particular response made was not the result of a strong association displacing eduction, but of an inadequate relation being educed, e.g.:- a general relation such as "this goes with that" rather than a specific relation such as "this is a particular function of that". From the explanations and questioning, it was clear that many children were quite capable of educing the correct relation, then finding a correlate, but they did not stop to consider the implications of the
relation illustrated in the first half of the sentence. In other words, these children read the question as a whole, then made an unchecked assumption about what was required. They did not make explicit the pattern of relations implied in the illustration, but concentrated on the given word and its most common characteristic.

(2) Failure to read all the words given, either in the sentence or in the choices offered. Once a word had been found which satisfied the child as relevant in some respect, later words in the line were not even read.

(3) Misinterpretation of an uncommon word; also unfamiliarity or mispronunciation of a commonly used word.

(4) Alteration in the "task direction" - i.e., reversing the relation established through the illustration, when applying it to the given word.

(5) Substitution of a different task previously worked - e.g.: attempting to classify the words in a series.

(6) Rejection of the correct answer on reasonable grounds.

The overall impression gained from this analysis of "analogies" items, was that the children were more concerned with the finding of a solution, rather than with "explicating" (in Duncker's sense of the term) the data, in order to find the specific requirements of the correct solution. They did not appreciate the fact, that although there may be several words which bear some relation to the problem, there is only one which is the best, and fits the structure of the item neatly. This may be described as a type of "sensitivity to logical necessity", partially dependent on specific learning.
(h) **Category (8) - Verbal Reasoning** (See Appendix H. pp 371-380).

Of the 8 items in this category, 4 are multiple choice, while 4 are of the open answer type, and in the summaries below the two types are presented separately.

(A) **Multiple Choice Items.**

It was difficult to establish the cause of error in many cases, because the children found it difficult to explain in words what they had done, without getting confused. Some of these children were encouraged to explain what they were attempting to do, by drawing, and this often resulted in a clarification of the problem for the child, leading to a correct answer. Conditions noted were:-

(1) Failure to note all the requirements of the problem, including reversal of the specific requirements, e.g.: getting the correct relationship, but expressing it from the wrong point of view.

(2) Confusion in keeping in mind the requirements of the task, e.g.: losing the "task direction" in concentrating upon a particular part of the problem.

(3) Unfamiliarity with, misinterpretation of, or failure to read a crucial word. The last of these was in most cases due to the particular way the items were set out in the test booklet.

(4) Misplaced emphasis on a part of the problem, e.g.: mistaking the key phrase.

(5) Lack of confidence. This was widespread, and although it did not lead to many omissions, it resulted in a tendency to choose an answer (including the phrase "cannot say which") without attempting to clarify the nature of the task.
(B) "Open-Answer" Items.

(1) Failure to clarify the requirements of the task, including the ignoring or forgetting of crucial aspects of the data. In some cases, this led to the ignoring of everything except the final instructions, a situation explained by one boy when he said knowingly, "...the top part might be just to fool you!"

(2) False assumptions about what was allowable, preventing the restructuring of the data.

(3) Persistence of an inappropriate "set" from previous items of a similar type.

(4) "Difficulty reduction" in a complex verbal situation.

Although verbal reasoning requires the educing of relations and correlates, there was much more involved in the incorrect responses than the displacing of eduction by reproduction. One of the major difficulties, appeared to be the clarification of what was required, i.e. the building up of a schema which could give direction to the various stages in the solution, and which was based on all the data provided. As in the other categories, the evidence points to a tendency for children to under-estimate the necessity for being clear as to exactly what is given, and exactly what is required.

(i) Category (9) - Proverbs. (See Appendix H. pp 380/384)

It was clear that most of the children who made errors had not heard these proverbs before, and did not know what a "saying" was. The actual choice of distractor was governed by several conditions:-
(1) Partial understanding of the task, with the ignoring of qualifying details, and the prepotency of words which were repeated in the distractors.

(2) Failure to read or consider all the statements given.

(3) Literal interpretations of the proverb, leading to a reasonable choice, (allowing for the limitations of the interpretation).

The analysis of errors was not very fruitful, because of the childrens' ignorance of the proverbs, and their failure to understand the "generalized" nature of a "saying''.

(j) Category (10) - Number Series. (See Appendix H. pp 384/390).

There were more omissions on this category, (9.6% of total), than on all others except (5) and (6), possibly due to lack of confidence after having attempted the first of the number series items, because whereas there were no omissions on item 17, on the next item-21, there were over half as many omissions as errors. Conditions noted were:-

(1) Failure to note the specific requirements of the task - in particular, the recording of the number wrong in the series, instead of its correction.

(2) Partial structuring of the data, coupled with the ignoring of other parts which did not fit this structure.

(3) Juxtaposition of inconsistent principles, even although recognized as such.

(4) Inability to restructure the problem once a particular principle had been chosen, coupled with the prepotency of certain properties of the numbers, e.g.:- the attraction of consecutive numbers.

(5) Satisfaction with an inferior, but consistent solution.
(6) Minor slips in applying a correctly established principle.

The analysis of this category illustrated very clearly, the tendency, already noted, to pay little attention to the exact form of the instructions. It also confirmed the observation that some children, although capable of re-working the item correctly when asked to read the item again, are quite satisfied to give a response based on only a part of the data, and to ignore disturbing elements.

As suggested by the number of omissions, and confirmed during the interviews, many children approached these items with hesitation and lack of confidence.

(k) Category (11) - Arithmetical Reasoning. (See Appendix H. pp 391-398).

This category contains items very similar to problems worked in classroom arithmetic, hence the analysis of the errors made is of some significance. If the errors are similar to those of the other categories, then there are good grounds for interpreting errors made on test items, in terms of classroom habits and attitudes. That the errors are in fact typical can be seen from the following conditions:-

(1) Failure to clarify the requirements of the task, due to superficial reading of the problem, or the prepotency of particular elements. This could be interpreted as "difficulty reduction" i.e., reducing the problem to something quite elementary, and answering this reformulated problem (with a certain amount of satisfaction). This could not be due merely to misreading, for even after children had re-read the items aloud, they still distorted them.
(2) Miscounting, or minor slips in the fundamental processes, after having established the correct principle.

(3) Inability to structure the problem, either because it was beyond the experience of the children, or because of lack of confidence preventing an analysis of the data. Coupled with this was the application of irrelevant principles.

(4) Misreading of particular words in the problem.

One of the major conditions producing error is illustrated in this, as in other categories, namely the distortion of the item in the direction of an easier, more familiar, or expected problem.

Before passing on to a more general discussion of these results, it should be pointed out that in almost all of the items, there were examples of correct answers given for wrong or inadequate reasons. In some cases these answers resulted from either a superficial or a devious train of reasoning, in others they were due to the ambiguous form of the structure of the item. Examples of these answers have been given in the records for each item.

The opposite condition, listed in some of the classifications, was rejection of the correct answer on reasonable grounds. Some of the examples given in Appendix H, illustrate how difficult it is to present the child with a clear cut problem - that is, "clear cut" from the child's point of view.

Some children were hesitant in answering, because they could see the weakness in the correct answer - e.g. the re-arranged sentence "bees gather honey from red clover" was questioned,
on the grounds that bees do not gather "honey", but "pollen"!

This weighing up of the meaning and intention of the item, was summed up by the comment of: one boy who agreed, reluctantly, to the reasonableness of the correct answer - "...if you look at it that way"!

II.- DISCUSSION.

In chapter I an attempt was made to throw some light on the nature of intelligence, by bringing together studies which could specify some of the underlying processes. This led to the postulation of a "methodological" component of intelligence, and a description of some of the concepts, including those from studies of reasoning and problem solving, which give greater meaning to factorial interpretations, and which are consistent with more general theories of cognitive functioning.

Some of these concepts were stressed as being particularly valuable, in that they add to Spearman's noegenetic principles, such notions as closure, control or direction, flexibility and globalization. It was argued that these concepts, which lead to a clarification of the processes involved in ordering relevant relations in the light of the particular task, give a clearer understanding of the "methodological" component. Moreover, in the light of the theories of Hebb and Piaget, and the experiments on reasoning and problem solv-
ing, it was claimed that the role of experience in the development of intelligence, could be understood more fully, by studying the ways in which learning enters into reasoning and problem solving.

In chapter II this claim was reduced to a specific study, i.e. the analysis of cognitive error, and it was shown that the positive concepts developed in chapter I, especially the emphasis on the order and structure of the task, were confirmed and elaborated, through a classification of errors, which was nearer to the "task displacement" of Selz than the "reproduction" of Spearman.

Although the chapters in section II produced little evidence of value, and eliminated, rather than confirmed, certain hypotheses set up, they at least resulted in the emphasizing of the structure of the item, on the one hand, and the attitudes and methods of the children, on the other. In this chapter, these two influences on test performance, have been examined more closely, along with an attempt to classify the conditions which led to the production of errors.

At this point, the investigation can be focussed on the classifications of cognitive error developed in chapter II, in order to note the extent to which the errors classified in this chapter, correspond to the original classification.
(a) Spearman, Selz and Relevant Experiments.

It is clear, from the descriptions of the errors made on the Otis test, that although reproduction plays a part in producing error, (as in fact it must do in producing the correct response), it is not the fundamental process involved. If the conditions involved were to be summed up under one heading, then the "task displacement" of Selz would seem to be the most appropriate description, i.e. a related problem serves instead of the one actually set. Furthermore the condition which produces "task displacement" - namely "the law of part efficiency", operating through the dropping of some characteristics, the prepotency or low dispositional effects of some elements, and the readiness to attend to the more general characteristics of the task - are just those conditions noted in the classifications above. The children answered the items, in many cases, on the basis of a "residual complex" - i.e. that part of the item which still functioned after the operation of the "law of part efficiency."

The errors found by Fortes, in his analysis of perceptual intelligence tests, can be matched quite closely with those described in this chapter, e.g.:

---

2See pp68-69 above.

3See pp70-71.
(1) the errors of "subsumption" leading to the distorting of a figure by supplementation, are of the same type as those described above, as the substitution of a different problem, due to a particular assumption about what is required;

(2) the displacement of energy from the total situation to a part only, i.e. certain elements prepotent, is illustrated in most of the classifications above;

(3) the correct general inference detached from the specific context, is paralleled by the failure to meet the specific requirements of the task, despite the establishing of the correct principle;

(4) the displacement of the crucial relation by an easier or more familiar relation, is paralleled by the displacement of a specific characteristic by a more general or common characteristic.

Fortes asked the question - "...is this a purely cognitive phenomenon...?" and although it is not possible to answer the question accurately, the interpretation suggested above in several instances - namely that many of the errors could be accounted for in terms of work habits and attitudes, developed in school work - may give some clue, to at least one of the important conditions involved.

Brigham's statement that, "although the problems may be presented and the answers recorded simply, yet the psychological structure of the situations may be exceedingly complex" has been fully confirmed, and furthermore, one of the clearest outcomes of this analysis, is the fact that errors are "not capricious, but orderly".  

4 See full quotation p72.
5 See p73.
(b) The Gestalt Account of Error.

Kohler's 3 types of errors were repeatedly observed, i.e. -

(1) "good" errors - Cf. rejection of the correct answer on reasonable grounds;
(2) complete lack of comprehension of the conditions of the task - Cf. failure to understand the general requirements of the task;
(3) stupidities arising from habit - Cf. the transfer of an irrelevant but "overlearned" procedure in approaching the task.

The conditions described by Duncker and Wertheimer, are very relevant to the conditions classified above, e.g.:-

(1) attempts to escape from the original setting of the problem - Cf. reformulation of the requirements of the task; 
(2) "single track" responses which leave out of consideration all other factors - Cf. partial structuring of the task, and the ignoring of disturbing elements;
(3) confusion of the original problem for another for which the correct solution is known - Cf. substitution of a different task previously worked;
(4) inadequate view of the whole situation - Cf. incomplete formulation of the task;
(5) losing breadth of view in concentrating on details - Cf. confusion in keeping in mind the requirements of the task;
(6) overextended view - Cf. adding words to the data given;
(7) premature or short cut closure - Cf. ignoring crucial aspects of the data, and giving an answer based on only a part of the problem.

---

6 See p74.
7 See pp75-76.
"Habitual directions" and "set", described by Maier and Luchins, were evident, as illustrated in the following conditions: losing task direction, false assumption as to what is allowable, wrong direction preventing restructuring, perseveration of earlier item structure, inability to break a strong combination of words or figures, failure to "re-centre" a word into a difficult context.

(c) **Errors Found in the Literature on Problem Solving.**

It is not necessary to attempt to match each error reported in the literature, with errors noted on the Otis test, as the general classification given, covers the major types of errors or error producing conditions.

(1) **Inadequate formulation of the problem.** This is the most frequent general condition of error reported in the literature, and as can be seen from the conditions described above, it applies also to the errors noted for each category. There is some formulation of the problem, but, as Piaget pointed out, the child does not ask himself the question as we ask it. It is not merely a case of his being unable to ask the right question, for the illustrations in the records make it clear, that in most cases, the child did not take the trouble to make

---

8 See p41 and p76.

9 See pp80-95 for the major groups of errors and the specific references discussed in this section.
sure that his interpretation was a correct or adequate one.

The various description of this condition; eg.:- insufficient knowledge of directions, failure to study the description of the problem fully enough to appreciate all the limiting factors, confusion about the requirements of the problem, failure to see the problem clearly or to consider all its aspects - are all applicable to the conditions listed for each category.

(2) Distortion or partial operation of the task. This is really a consequence of the first condition, and Piaget's explanation - that a schema is formed, which is unconsciously applied to whatever will "more or less" fit into it, is particularly valuable in pointing to two characteristics - the unconscious operation of the process, and the fitting of the data given, to the schema constructed.

Whether it be described as rubricizing, levelling and sharpening, attending to only a few clues or substitution of a manageable problem, the process is clearly recognizable in the analysis of the children's errors, on almost every item in the test. Hildreth's description of this tendency as "difficulty reduction" is very appropriate, and her comment that the subject "remakes" the problem in order to make it meaningful, points to an explanation of why so many children appear to be satisfied, with an inferior solution. Having distorted the schema, to fit whatever aspects of the problem
have been attended to, the children are satisfied, because they have constructed a manageable problem, and are unconscious of the elements ignored or misinterpreted.

In the light of the conditions described for each category, the following statement from chapter II\(^{10}\) is most pertinent - "...the actual problem the subject sets about answering may be more general, more specialized, simplified, distorted or only partially interpreted".

(3) Dislocation or disrelation of the elements in a problem. Failure in analysis due to an "intuitional" approach, the overlooking of major variables, lack of awareness of the given elements in the problem, lack of careful deliberation, under or over potency of elements (including the unfamiliar as well as the familiar), - these conditions are all evident in the various classifications given above. One of the most important of these conditions - lack of awareness of the given elements - can be associated with misreading, although a significant outcome of the interviews was the observation of a persistent contradiction between what the child read, and what he thought he had read.

(4) Influence of "set". This has already been illustrated in dealing with the Gestalt account of error. Particular aspects of "set" described in the literature - e.g.- inability

\(^{10}\)p85.
to see a particular idea in a different function, inability to change a relationship once it has been established, transfer of the wrong principle, failure to find an effective technique for starting the problem, loss of "task set" - have all been used in the analysis, as appropriate descriptions of the conditions leading to the production of errors, in the test items.

(5) Influence of attitude, emotional involvement and personal preference. It is very difficult to judge what attitudes the children have towards the test as a whole, and towards particular test items. Lack of confidence, is an obvious case of an attitude influencing performance, yet there are more subtle attitudes, revealed occasionally through a remark, a gesture or tone of voice. During the interviews, it was possible to observe such attitudes as over-confidence, eagerness to get the right answer, disinterestedness, hesitation and doubt even although the question was answered quite correctly, as well as broad "analytic" or "intuitional" approaches. Some children gave an answer, then sat back waiting for the next item to be reworked, others gave an answer, then kept scanning over the item as though they were not satisfied, or wanted to make sure. The "either-or" attitude, noted by Bloom and Broder, was a characteristic of some children who took one look at the item, and announced that they could not do that sort of problem.
From observations made while supervising the class testing, it seems likely that, especially in the middle and lower I.Q. groups, there are considerable variations in the amount of effort and interest stimulated by the test situation.11

(d) Specific Errors in Intelligence Test Items.

There was little evidence in this chapter, relevant to the points dealt with in chapter II12 under this heading. One general comment, concerning concentrations of distractor choices, is that the type of analysis attempted in chapter VII, (i.e. distributions of wrong responses for each item), is of little value unless accompanied by a more detailed investigation of the reasons for particular choices. As illustrated in appendix H, the same distractor may be chosen for very different reasons, thus it can be misleading to read into error frequencies, an a priori interpretation, based on the probable process leading to that particular error, from the point of view of the "test sophisticated" investigator.

(e) Carefulness.

Many of the errors described above could have been avoided, if the children had checked over their reading of


12See pp95-96.
the items, the choices offered, and the actual responses made. In chapter II\textsuperscript{13} it was pointed out that the factor called "carefulness" may not be a temperamental trait, but a learned method, developed through the many experiences of school work. One of the significant features of the classifications and discussion in this chapter, is the evidence that errors can be interpreted in terms of certain specific processes, which recur in many different types of problems; thus the crucial question is not - can children exercise more care in general? but - in what specific ways should care be exercised?

In view of the fact that the conditions producing errors in the Otis test items, are in general, the same conditions which have been found to operate in the wider cognitive domain, it can be concluded, that there is some evidence to support an interpretation of the "methodological" component of intelligence, in terms of concepts developed in studies of reasoning and problem solving.

\textsuperscript{13}See p100.
CHAPTER X.

SUMMARY AND CONCLUSIONS.

I. SUMMARY.

The significance of method in intellectual tasks is a general topic of such broad scope and widespread implications, that it embraces many specific topics and gives rise to many diverse problems. This thesis arose from an interest in one particular aspect of the general topic, namely work habits or methods of children in relation to performance in school work. An attempt to understand the fundamental problems involved in work methods and performance, led back to a more basic study of intellectual development, and to the nature of the processes underlying intelligence.

Section A.

(1) In the first chapter, a review of factorial studies led to the conclusion that Spearman's noegenetic principles are inadequate as a description of 'g', and that along with formal analyses of relations and their fundamentals, it is necessary to add concepts such as closure, direction or control, flexibility, and synthesis or globalization. It was shown that these concepts can be understood in terms of the processes involved in reasoning and problem solving, and that these latter are developed through learning. In order to allow for the interaction of the innate and the learned aspects of
intelligence, a "methodological" component was postulated. That such a component is consistent with a general cognitive theory was illustrated by showing its operation in the complementary theories of Hebb and Piaget.

(2) In order to examine the characteristics of this "methodological" component more closely, it was decided to approach the problem from the point of view of mal-functioning, through a study of cognitive error. Two contrasting theories of error were discussed - those of Spearman and Selz - and evidence bearing directly upon these, was interpreted as favouring the type of theory developed by Selz. A review of errors reported in the literature on problem solving and reasoning, supported this choice, and at the same time provided a classification based upon diverse experimental studies. A "carefulness" factor, the influence of socio-economic status, and some characteristics of distractor distributions, were also noted as being relevant to the study of error.

Section B.

(3) At this point, an experimental design was described, which aimed at the investigation of the nature of the errors made by children on an intelligence test. This step was taken in order to judge whether the theoretical and experimental investigations on error, described in chapter II, were applicable to intelligence test errors. If so, then it could be claimed that some of the characteristics of the
"methodological" component, had been clarified. As a background to the main method of investigation chosen - namely the analysis of errors made by individual children, through reworking and questioning - several subsidiary questions were posed, which could be answered through analyses of the children's test results.

(4) These analyses provided evidence for the following conclusions.

(a) Repetition of errors on Otis I and II: In comparing the errors made on the two tests, it was found that an average of 64% of the errors were made on the same items, while of these repeated item errors, an average of 58% were identical wrong responses.

(b) Consistency of errors in, and difficulty of categories: The proportion of errors in each of the eleven categories (into which the items were classified), remained constant from Otis I to Otis II. Although the difficulty of the items in each category - as judged through the amount of error in each - was close to the expected or predicted difficulty, there was a discrepancy on category (5) - mixed sentences. This was due to the concentration of omissions on this category, a feature interpreted as resulting from an attitude of avoidance of the unfamiliar, or lack of confidence.

(c) Comparison of high error and low error groups over all categories: For each category, two groups were selected
matched for I.Q. and overall error tendency, but showing a marked contrast in the number of errors made on the particular category. Several hypotheses were advanced, and the evidence presented resulted in the following conclusions:

i. Socio-economic status ranks did not differ significantly for the two groups, on any of the categories.

ii. When the two groups for a given category were compared on the other categories, a priori predictions of expected differences were not confirmed.

iii. Comparisons of the performance of the two groups, on similar categories or sub-tests in the A.C.E.R. Intermediate D general ability test, and the Thurstone Primary Mental Abilities test (S.R.A. edition), resulted in the repetition of significant differences on only two categories, (1)-word meaning, and (2)-proverbs. The results were interpreted as suggesting that the differences between high error and low error groups on a given category, may be due mainly to the particular structure of the items.

iv. Although it was expected that the two groups would make different types of errors - as judged by particular distractor choices or wrong responses on each item - the evidence pointed to the great similarity of the types of errors made by each group.

v. There was a slight tendency for girls to make more errors than boys on three categories, but on only one of these three (category (3)) was the difference significant.

(d) Inconsistency of error ratios: Individual cases of marked inconsistency in the number of errors made on any category, from Otis I to II, were illustrated. These inconsistencies were interpreted in terms of learning or the development of temporary sets, during the course of working the test.

(e) Carefulness-carelessness ratings in relation to error:
Two groups were selected on the basis of teachers' ratings, as representing the extremes of a trait described as carefulness-carelessness. It was predicted that these two groups would show significant differences on such measures as overall error tendency, repetition of errors, omissions, and variability in I.Q. from Otis I to Otis II. The predicted differences did not appear, although a substantial sex difference was noted - girls being rated as "careful" more often than boys. In view of this difference the value of the ratings obtained was questioned.

(f) Distributions of wrong responses over all cases: The distributions were interpreted as indicating the following:

i. The tendency, noted by other investigators, for the first distractor position to be favoured over all others, was not confirmed, although the third and fourth positions carried slightly lower percentages of the total wrong responses.

ii. Wrong responses were not evenly distributed in either multiple choice or "open answer" type items, many items having 50% or more of the errors concentrated on one distractor or wrong response.

iii. In general, the distributions of errors were similar for high, average and low I.Q. groups.

iv. Sex differences in total errors made for each category, were interpreted as pointing to the desirability of examining specific item differences rather than category differences.

v. From a sampling of Otis II error distributions, it was concluded that, in general, the distributions are consistent over the two tests.
Section C.

(5) Eighty-four children were selected for interview, the method adopted being a combination of reworking aloud the items on which errors had been made, and questioning by the investigator, in order to check steps which were not clear from the child's explanation of his reworking.

(6) A classification of the conditions leading to error, was made for each of the eleven categories in the test, based on an analysis of the errors on each item within any one category. These classifications were then compared with the original classifications discussed in chapter II. It was shown that the same conditions were responsible for the production of error, in the diverse studies included in the general review, and the specific analysis of intelligence test errors.

(7) In the light of these findings it was concluded that there are good grounds for accepting the hypothesis developed in the first chapter, namely, that the postulation of a "methodological" component in intelligence, not only assists in clarifying the relation between the innate and the learned characteristics in intelligence, but also leads to the enrichment of the concept of intelligence through incorporation of concepts developed in studies of reasoning and problem solving.
II. CONCLUSIONS.

There are certain implications of the results reported above, which may be discussed briefly under two headings - firstly, the implications for intelligence test construction, and secondly, the implications for classroom teaching.

(a) Intelligence test construction.

Although the value of multiple-choice items in objective testing cannot be questioned, the concentration of wrong responses on particular distractors, indicates that many of the choices offered play very little part in the answering of the item. On the other hand the inclusion of "strong" distractors may weaken the test by directing attention away from the actual question asked, and leading to an undue dependence on the comparing of alternative answers.

From the records of the interviews, it is clear that the actual setting out and wording of the item, may have a considerable influence on the answers given,\(^1\) and to ensure that the instructions are quite clear, they should be checked by taking a group of children, and asking each child to say what he thinks the question really means. Such phrases as "which tells best", "which one is most like", or the crucial

\(^1\)A summary of the specific weaknesses in the Otis items, has been added as appendix I.
words "is to a" in analogies items, may lead either to superficial answers on the part of those who skip over the instructions, or to wrong but reasonable answers, on the part of those who read more into the instructions than is intended.

The individual analysis of errors, indicated that it can be misleading to identify the statistical criterion of "difficulty", with a psychological criterion. An item may carry a high proportion of errors for several reasons, the least of important of which may be the complexity of the relations involved, while the most important, may be the influence of a particular characteristic of the way the instructions are worded.

One general comment, arising from the discussion in chapter I, and the records of individual analyses, concerns the type and range of items included in a test such as the Otis. Although it is in many ways economical and convenient, to test intelligence through a sample of 75 short questions, at the same time this sort of test, with a clear right or wrong for each question, does not allow for degrees of "rightness" in the answers. The individual explanations show what great differences there are in the quality of the wrong answers. Even a scheme of weighting each answer in terms of the response chosen, would still fall short, for the same wrong response may be given for different reasons. Although it may be more difficult to score, and take longer to administer, it might be
profitable in the long run, to develop tests in which, for example, four or five problems relevant to children's interests were presented, a number of lines of action suggested, provision made for the evaluation of each element contained in the data, in the light of the action chosen, and a list of reasons offered from which some may be selected to support the choice made. Such tests could lead to the use of test results in a diagnostic as well as predictive capacity, and throw more light on qualitative differences, at present obscured through the use of a single test score.

While it is necessary to improve the setting out and wording of test items (and test constructors are well aware of this) one of the implications of the evidence resulting from the individual analyses, is that there is a limit to the value of such improvements, a limit imposed by the work habits and attitudes of the children. This leads to the second, and much more important part of this consideration of implications.

---

2 This is merely a variation of a type of test illustrated in several different contexts, i.e. Guilford and Lacey (see ref. to judgement and integration tests p21 above), Smith and Tyler (especially part I see ref. p30 above), Lafitte (see ref. p21), Ashby (Ashby E. Challenge to Education, 1946, Sydney, Angus and Robertson), and Boyd et al (Boyd E. Call D, Copsey B, "So You Think You Can Think," Progressive Education, 1952, March, pp170-174).
(b) Classroom teaching.

There has been a great deal of discussion on conditions which influence test scores, particularly the influence of practice, coaching and test sophistication. It could be argued that the classifications of errors given in the previous chapter, merely specify certain weaknesses which could be eradicated through coaching. Doubtless such classifications would be of some use to the teacher interested in increasing the mean I.Q. of his class by a few points, for some particular competitive purpose, but the whole argument developed in the thesis is based on a more fundamental point of view.

This point of view can be summed up in the following way:

i. Cognitive processes which underly that characteristic of behaviour known as intelligence, are dependent on a certain level of neurological complexity, and certain types of experiences which call forth, and provide the context for the development of, these processes.

ii. This context determines not only the material upon which these processes operate, but also, by way of particular habits, methods, and the types of material involved, the form in which they will develop.

iii. Because of the considerable uniformity of the contexts in which most civilized children grow up, there is a considerable uniformity in the ways in which these processes operate, and the sort of material in which it is assumed that they will be most clearly demonstrated.

iv. Because there has been a tendency to overlook the influence of context on the development of intelligence, teachers have been led to underestimate the influence of the attitudes,
habits and consciously formulated methods, which are built up in the child throughout the school years.

v. This underestimation has been strengthened, because of typical experiments, in which a period of training has resulted in either a small increase in I.Q., or a larger increase which has, however, been temporary only. Such experiments have been interpreted as showing that training may increase the I.Q., but that the increase is not due to a fundamental change in intelligence, but merely to the greater familiarity of the subjects with the form of the test items, and to the reducing of certain incidental factors to a minimum (e.g. knowing what to expect in the instructions, familiarity with the methods of expressing answers, not wasting too much time on one item).

vi. While this interpretation is probably adequate for such training experiments, it does not thereby preclude the influence of long term training when this is thought of in terms of all the educational experiences, over all of the child's formative years, in school and out of school.

vii. These educational experiences enter into the development of intelligence, and what is measured by an intelligence test is the result of the interaction of the original neurological foundation and the context within which it is developed, including the experiences provided by the school.

As was pointed out at the beginning of the thesis, only a small segment of this very complex problem could be investigated, thus the results merely point to one aspect of educational experience which bears on the general problem. This aspect is concerned with the development of methods of approaching and handling problem situations. The implication which follows from the general argument and the specific results presented, is that teachers should be conscious of, and should attempt to make children conscious of such methods of approaching and handling intellectual tasks as, for example, the following:-
i. Careful analysis of all the data, with checks in order to ascertain that at least all elements have been accounted for.

ii. Looking at the problem as a whole, in order to determine the crucial part and the qualifying parts.

iii. Being prepared to give reasons for specific choices where several plausible answers are possible.

iv. Trying several methods - i.e. a conscious flexibility of approach.

v. Verification of an answer against all parts of the data.

vi. Following directions accurately - in particular making sure that the question as formulated is the same as the actual question asked.

There is nothing original about these examples - good teachers have always attempted to do these things, but what is important, is that teachers at all levels in the school system should realize that they cannot afford to leave these methods to develop automatically, but must constantly train children in them, whenever appropriate subject matter can be used in this way. It is quite clear from the results of this investigation for example, that children are not given enough direct training in reading instructions or following specific directions - yet this is a matter of major importance, not just in relation to school work, but even more so in later occupational experience.

This implies that teachers should present their subjects, and set examinations, in such a manner that they will focus attention on the ways in which problems are handled, as well
as on the content of the subjects, and the rightness or wrongness of particular answers. A formal course in "problem solving" would be unlikely to achieve the desired results; the approach would need to be through the actual subject matter on which the children were working, especially in such subjects as mathematics, the sciences and social studies.

The classifications of the previous chapter provide a foundation, not only for remedial or preventative training, but also for the development of attitudes, habits and methods of work which enter into, and are partly responsible for determining, the level of functioning of the child's intelligence.
BIBLIOGRAPHY.
BIBLIOGRAPHY.


Ashby E., Challenge to Education, 1946, Sydney, Angus and Robertson.


Bartlett F.C., Remembering, 1932, Cambridge University Press.


Brigham C., A Study of Error, 1932, N.Y.: College Entrance Examination Board.


Humphrey G., Thinking, 1951, London: Methuen.


Roethlisberger F.J. and Dickson W.J., Management and the Worker, 1939, Harvard Univ. Press.


A Factorial Study of Perception, 1944, Psychomet. Mon. No. 4.


and Thurstone G., Factorial Studies of Intelligence, 1941, Psychomet. Mon. No. 2.


Wolfle D., Factor Analysis to 1940, 1940, Psychomet. Mon. No. 3.

APPENDICES.
APPENDIX A.

MEAN ERROR RATIO SCORES BY CATEGORY, SEX, AND I.Q. GROUP, FOR OTIS I.

<table>
<thead>
<tr>
<th>Group</th>
<th>Category</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Too few girls in this group to warrant separate treatment)</td>
</tr>
<tr>
<td>T</td>
<td>3.1</td>
<td>1.8</td>
<td>1.4</td>
<td>1.9</td>
<td>2.3</td>
<td>2.2</td>
<td>1.3</td>
<td>1.7</td>
<td>3.9</td>
<td>1.6</td>
<td>1.2</td>
<td>2.04</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>3.0</td>
<td>2.3</td>
<td>2.0</td>
<td>1.8</td>
<td>2.8</td>
<td>2.7</td>
<td>1.6</td>
<td>2.4</td>
<td>3.6</td>
<td>3.6</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>2.9</td>
<td>1.5</td>
<td>2.6</td>
<td>1.4</td>
<td>3.8</td>
<td>2.2</td>
<td>1.4</td>
<td>1.8</td>
<td>4.6</td>
<td>2.3</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>3.0</td>
<td>1.9</td>
<td>2.2</td>
<td>1.6</td>
<td>3.2</td>
<td>2.5</td>
<td>1.5</td>
<td>2.2</td>
<td>4.1</td>
<td>3.0</td>
<td>2.6</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>3.3</td>
<td>2.1</td>
<td>2.2</td>
<td>2.8</td>
<td>3.9</td>
<td>2.6</td>
<td>1.6</td>
<td>3.4</td>
<td>5.2</td>
<td>3.8</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>3.7</td>
<td>1.6</td>
<td>3.4</td>
<td>3.3</td>
<td>4.6</td>
<td>3.1</td>
<td>1.7</td>
<td>4.3</td>
<td>5.4</td>
<td>3.7</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>3.4</td>
<td>1.9</td>
<td>2.8</td>
<td>3.0</td>
<td>4.2</td>
<td>2.8</td>
<td>1.6</td>
<td>3.7</td>
<td>5.3</td>
<td>3.8</td>
<td>3.3</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>4.3</td>
<td>1.8</td>
<td>3.0</td>
<td>3.1</td>
<td>6.2</td>
<td>3.0</td>
<td>2.0</td>
<td>4.6</td>
<td>6.4</td>
<td>4.2</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>4.0</td>
<td>2.2</td>
<td>4.0</td>
<td>2.7</td>
<td>4.8</td>
<td>2.6</td>
<td>2.4</td>
<td>5.5</td>
<td>4.2</td>
<td>3.9</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>4.2</td>
<td>2.0</td>
<td>3.4</td>
<td>2.9</td>
<td>5.5</td>
<td>2.8</td>
<td>2.2</td>
<td>4.9</td>
<td>5.5</td>
<td>4.1</td>
<td>4.2</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>4.9</td>
<td>2.3</td>
<td>4.5</td>
<td>3.6</td>
<td>6.7</td>
<td>4.3</td>
<td>2.5</td>
<td>5.4</td>
<td>6.7</td>
<td>4.5</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>4.6</td>
<td>2.0</td>
<td>3.8</td>
<td>3.1</td>
<td>6.3</td>
<td>3.8</td>
<td>3.0</td>
<td>5.7</td>
<td>6.0</td>
<td>4.2</td>
<td>5.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>4.7</td>
<td>2.3</td>
<td>4.1</td>
<td>3.3</td>
<td>6.4</td>
<td>3.9</td>
<td>2.7</td>
<td>5.6</td>
<td>6.3</td>
<td>4.3</td>
<td>5.0</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>5.3</td>
<td>4.1</td>
<td>3.8</td>
<td>4.9</td>
<td>8.3</td>
<td>4.4</td>
<td>3.4</td>
<td>6.2</td>
<td>7.7</td>
<td>3.5</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>5.0</td>
<td>3.1</td>
<td>4.7</td>
<td>4.0</td>
<td>6.5</td>
<td>3.8</td>
<td>3.1</td>
<td>6.2</td>
<td>5.9</td>
<td>4.4</td>
<td>4.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>5.1</td>
<td>3.6</td>
<td>4.3</td>
<td>4.4</td>
<td>7.3</td>
<td>4.1</td>
<td>3.3</td>
<td>6.2</td>
<td>6.7</td>
<td>4.0</td>
<td>4.9</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>6.3</td>
<td>5.4</td>
<td>5.9</td>
<td>4.5</td>
<td>8.5</td>
<td>5.6</td>
<td>4.1</td>
<td>8.4</td>
<td>8.5</td>
<td>5.5</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>6.6</td>
<td>3.8</td>
<td>5.8</td>
<td>5.0</td>
<td>8.4</td>
<td>5.6</td>
<td>4.3</td>
<td>6.6</td>
<td>10.0</td>
<td>4.8</td>
<td>8.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>6.4</td>
<td>4.8</td>
<td>5.8</td>
<td>4.6</td>
<td>8.5</td>
<td>5.6</td>
<td>4.1</td>
<td>7.8</td>
<td>9.0</td>
<td>5.3</td>
<td>6.9</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.3</td>
<td>2.6</td>
<td>3.3</td>
<td>3.1</td>
<td>5.3</td>
<td>3.4</td>
<td>2.4</td>
<td>4.6</td>
<td>5.8</td>
<td>3.7</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B.

EXAMPLE OF RECORD SHEET FOR MARKING ERRORS ETC.

<table>
<thead>
<tr>
<th>Name. M... W....</th>
<th>Sex. B</th>
<th>Age, 10-7 at. 18-11-53.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributing School.</td>
<td>Addington</td>
<td></td>
</tr>
<tr>
<td>Present School.</td>
<td>South Intermediate Class. Form I D.</td>
<td></td>
</tr>
<tr>
<td>Fathers Occupation.</td>
<td>Electrician</td>
<td></td>
</tr>
<tr>
<td>Rating.</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>I. Q. Group.</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Otis Int.I. I. Q. 108</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score. 34 Wrong. 16</td>
</tr>
<tr>
<td>Wrong/Attempted. 16/50(32%)</td>
</tr>
<tr>
<td>Last Attempted. 54(72%) (21)</td>
</tr>
<tr>
<td>Total Attempted. 50(67%)</td>
</tr>
<tr>
<td>Omitted. 4(7%)+50+21</td>
</tr>
<tr>
<td>Check. 75</td>
</tr>
</tbody>
</table>

Wrong - (a = item number; b = actual wrong response).

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>a</th>
<th>b</th>
<th>a</th>
<th>b</th>
<th>a</th>
<th>b</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>D</td>
<td>15</td>
<td>2</td>
<td>25</td>
<td>5</td>
<td>30</td>
<td>5</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>36</td>
<td>5</td>
<td>38</td>
<td>1</td>
<td>41</td>
<td>5</td>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>1</td>
<td>46</td>
<td>2</td>
<td>47</td>
<td>7</td>
<td>54</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Distractor Frequency.

| (1) 4 | (2) 3 | (3) 2 | (4) 0 | (5) 4 |

Omissions.

| 50 | 51 | 52 | 53 |

Classification of Wrongs and Omissions by Categories.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Error Ratios

<table>
<thead>
<tr>
<th>2</th>
<th>0</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>1</th>
<th>7</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>2.9</td>
<td>5.0</td>
<td>4.0</td>
<td>2.0</td>
<td>5.4</td>
<td>4.0</td>
<td>10.0</td>
<td>3.3</td>
<td>0</td>
</tr>
</tbody>
</table>
### APPENDIX C.

**I. MEAN ERROR RATIO SCORES BY CATEGORY, SEX, AND I.Q. GROUP, FOR OTIS II.**

| Group 1 | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | Total
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 B</td>
<td>2.4</td>
<td>1.9</td>
<td>1.2</td>
<td>1.0</td>
<td>1.6</td>
<td>1.7</td>
<td>0.6</td>
<td>1.4</td>
<td>2.0</td>
<td>1.0</td>
<td>2.2</td>
</tr>
<tr>
<td>G</td>
<td>2.5</td>
<td>1.9</td>
<td>1.7</td>
<td>1.2</td>
<td>1.5</td>
<td>2.1</td>
<td>0.5</td>
<td>1.1</td>
<td>0.9</td>
<td>1.6</td>
<td>2.7</td>
</tr>
<tr>
<td>T</td>
<td>2.5</td>
<td>1.9</td>
<td>1.3</td>
<td>1.1</td>
<td>1.6</td>
<td>1.8</td>
<td>0.6</td>
<td>1.3</td>
<td>1.7</td>
<td>1.2</td>
<td>2.3</td>
</tr>
<tr>
<td>2 B</td>
<td>3.4</td>
<td>1.7</td>
<td>1.4</td>
<td>1.3</td>
<td>2.1</td>
<td>2.0</td>
<td>1.3</td>
<td>2.0</td>
<td>2.2</td>
<td>2.9</td>
<td>2.9</td>
</tr>
<tr>
<td>G</td>
<td>2.5</td>
<td>2.2</td>
<td>2.7</td>
<td>2.0</td>
<td>3.3</td>
<td>1.7</td>
<td>1.2</td>
<td>1.8</td>
<td>3.6</td>
<td>3.8</td>
<td>2.8</td>
</tr>
<tr>
<td>T</td>
<td>3.2</td>
<td>1.8</td>
<td>1.7</td>
<td>1.5</td>
<td>2.5</td>
<td>1.9</td>
<td>1.3</td>
<td>1.9</td>
<td>2.6</td>
<td>3.1</td>
<td>2.9</td>
</tr>
<tr>
<td>3 B</td>
<td>3.6</td>
<td>2.3</td>
<td>2.1</td>
<td>2.1</td>
<td>3.2</td>
<td>2.7</td>
<td>1.0</td>
<td>3.3</td>
<td>4.1</td>
<td>3.4</td>
<td>3.0</td>
</tr>
<tr>
<td>G</td>
<td>4.0</td>
<td>2.1</td>
<td>2.3</td>
<td>2.3</td>
<td>3.6</td>
<td>2.9</td>
<td>1.6</td>
<td>2.5</td>
<td>3.2</td>
<td>2.8</td>
<td>3.3</td>
</tr>
<tr>
<td>T</td>
<td>3.8</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>3.4</td>
<td>2.8</td>
<td>1.3</td>
<td>2.9</td>
<td>3.7</td>
<td>3.1</td>
<td>3.4</td>
</tr>
<tr>
<td>4 B</td>
<td>4.0</td>
<td>1.9</td>
<td>2.7</td>
<td>2.8</td>
<td>4.5</td>
<td>3.4</td>
<td>1.8</td>
<td>3.8</td>
<td>4.6</td>
<td>3.9</td>
<td>3.6</td>
</tr>
<tr>
<td>G</td>
<td>4.4</td>
<td>3.3</td>
<td>3.4</td>
<td>3.3</td>
<td>4.2</td>
<td>2.9</td>
<td>1.8</td>
<td>3.8</td>
<td>5.1</td>
<td>3.3</td>
<td>4.3</td>
</tr>
<tr>
<td>T</td>
<td>4.2</td>
<td>2.5</td>
<td>3.0</td>
<td>3.0</td>
<td>4.4</td>
<td>3.2</td>
<td>1.8</td>
<td>3.8</td>
<td>4.8</td>
<td>3.7</td>
<td>4.1</td>
</tr>
<tr>
<td>5 B</td>
<td>4.6</td>
<td>2.5</td>
<td>2.5</td>
<td>3.1</td>
<td>6.3</td>
<td>3.3</td>
<td>2.2</td>
<td>5.3</td>
<td>7.7</td>
<td>4.9</td>
<td>5.5</td>
</tr>
<tr>
<td>G</td>
<td>4.4</td>
<td>2.2</td>
<td>3.4</td>
<td>3.0</td>
<td>5.3</td>
<td>3.2</td>
<td>2.5</td>
<td>4.9</td>
<td>6.1</td>
<td>3.9</td>
<td>5.8</td>
</tr>
<tr>
<td>T</td>
<td>4.5</td>
<td>2.3</td>
<td>3.0</td>
<td>3.1</td>
<td>5.7</td>
<td>3.3</td>
<td>2.4</td>
<td>5.0</td>
<td>6.8</td>
<td>4.3</td>
<td>5.7</td>
</tr>
<tr>
<td>6 B</td>
<td>4.4</td>
<td>2.9</td>
<td>3.0</td>
<td>3.6</td>
<td>6.6</td>
<td>4.0</td>
<td>1.8</td>
<td>5.8</td>
<td>6.1</td>
<td>3.4</td>
<td>4.9</td>
</tr>
<tr>
<td>G</td>
<td>5.4</td>
<td>3.5</td>
<td>4.4</td>
<td>3.8</td>
<td>5.7</td>
<td>3.7</td>
<td>2.9</td>
<td>6.2</td>
<td>7.7</td>
<td>5.7</td>
<td>5.6</td>
</tr>
<tr>
<td>T</td>
<td>4.9</td>
<td>3.2</td>
<td>3.7</td>
<td>3.7</td>
<td>6.2</td>
<td>3.8</td>
<td>2.3</td>
<td>6.0</td>
<td>6.9</td>
<td>4.5</td>
<td>5.2</td>
</tr>
<tr>
<td>7 B</td>
<td>5.6</td>
<td>5.4</td>
<td>5.3</td>
<td>4.6</td>
<td>8.9</td>
<td>5.1</td>
<td>3.9</td>
<td>7.7</td>
<td>8.2</td>
<td>5.7</td>
<td>7.1</td>
</tr>
<tr>
<td>G</td>
<td>6.1</td>
<td>4.4</td>
<td>5.9</td>
<td>4.6</td>
<td>7.7</td>
<td>4.7</td>
<td>3.5</td>
<td>7.2</td>
<td>8.3</td>
<td>5.1</td>
<td>7.0</td>
</tr>
<tr>
<td>T</td>
<td>5.8</td>
<td>5.1</td>
<td>5.5</td>
<td>4.6</td>
<td>8.5</td>
<td>5.0</td>
<td>3.8</td>
<td>7.6</td>
<td>8.2</td>
<td>5.5</td>
<td>7.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>6.1</strong></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>4.1</td>
<td></td>
<td>2.7</td>
<td>2.9</td>
<td>2.7</td>
<td>4.6</td>
<td>3.1</td>
<td>1.9</td>
<td>4.1</td>
<td>5.0</td>
<td>3.6</td>
</tr>
</tbody>
</table>

**II. No. of Cases in 7 I.Q. Groups Based on Approximately Half S.D. (6 Points) for OTIS II.**

<table>
<thead>
<tr>
<th>Group</th>
<th>I.Q. Range</th>
<th>No. of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>123+</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>117-122</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>111-116</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>105-110</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>99-104</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>93-98</td>
<td>27</td>
</tr>
<tr>
<td>7</td>
<td>92-</td>
<td>38</td>
</tr>
</tbody>
</table>
APPENDIX D.

ADJUSTMENT OF ERROR RATIO SCORES.

A description and discussion of the method used will be clearer if it is based upon an actual example, e.g. - the adjustments made to the scores of "W. McC."

Mean Error Ratio Score 2.9
I.Q. Group Mean 2.5 (Adj. = .4)

<table>
<thead>
<tr>
<th>Categ.</th>
<th>Error Ratio</th>
<th>Deviation</th>
<th>Adjusted Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td>.5</td>
<td>.9</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>.6</td>
<td>.2</td>
</tr>
<tr>
<td>3</td>
<td>2.5</td>
<td>.3</td>
<td>.1</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>1.6</td>
<td>2.0</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
<td>1.5</td>
<td>1.9</td>
</tr>
<tr>
<td>6</td>
<td>3.3</td>
<td>.8</td>
<td>.4</td>
</tr>
<tr>
<td>7</td>
<td>3.8</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td>8</td>
<td>5.0</td>
<td>2.8</td>
<td>2.4</td>
</tr>
<tr>
<td>9</td>
<td>2.5</td>
<td>1.6</td>
<td>2.0</td>
</tr>
<tr>
<td>10</td>
<td>6.0</td>
<td>3.0</td>
<td>2.6</td>
</tr>
<tr>
<td>11</td>
<td>2.0</td>
<td>.6</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Otis II, I.Q. 121, (I.Q. Group 2).
Mean Error Ratio Score 2.1
I.Q. Group Mean 2.2 (Adj. = .1).

<table>
<thead>
<tr>
<th>Categ.</th>
<th>Error Ratio</th>
<th>Deviation</th>
<th>Adjusted Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td>.7</td>
<td>.6</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>.7</td>
<td>.8</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>1.7</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>2.0</td>
<td>.5</td>
<td>.6</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>2.5</td>
<td>2.4</td>
</tr>
<tr>
<td>6</td>
<td>2.2</td>
<td>.8</td>
<td>.4</td>
</tr>
<tr>
<td>7</td>
<td>2.3</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>8</td>
<td>1.3</td>
<td>.6</td>
<td>.5</td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td>2.6</td>
<td>2.5</td>
</tr>
<tr>
<td>10</td>
<td>8.0</td>
<td>4.9</td>
<td>5.0</td>
</tr>
<tr>
<td>11</td>
<td>2.0</td>
<td>.9</td>
<td>.8</td>
</tr>
</tbody>
</table>

Sum of Otis I and II, Adjusted Deviations

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>1.9</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>1.8</td>
<td>4.5</td>
</tr>
</tbody>
</table>
Step (1).

Error ratio scores for Otis I and II listed; (i.e. these are the ratios of number of errors and omissions made in any category over total number of items in that category covered by the individual: e.g. Categ. (1) 2/3 = 2.5).

Step (2).

Deviation of each error ratio score from the mean error ratio score of the I.Q. group into which the individual falls (Group 2 in both tests). This is done separately for each category and each test, using the figures in Appendices A and C. This step should do two things: it should reduce, although it cannot entirely eliminate, (a) the effects of I.Q. on the error ratio score, because each individual's deviation is expressed in terms of his own I.Q. group; (b) the differences in error ratio scores due to differences in the difficulty of the items in the various categories, because the deviation now becomes a deviation from each category mean. In the above example the 'Deviation' columns show the result of this step. (A good example of the necessity of considering category means being shown in the error ratio scores on categories (6) and (7) compared with their deviations).

Step (3).

The mean error ratio for this boy is 2.9 (on Otis I) while the I.Q. group mean error ratio (Otis I) is 2.5 (from Appendix A). In other words this boy makes more errors
overall than his I.Q. group as a whole, this being due to some such characteristic as a tendency to attempt more items but get more wrong, or to age (i.e. a younger boy can make more errors yet remain in the same I.Q. group as an older boy.) While this does not distort a comparison between categories for this particular boy, it does distort any comparison with the deviations of other children whose mean error ratio scores are different from his.

By subtracting 0.4 (i.e. 2.9 - 0.4 = 2.5) from each category deviation, the pattern of errors (i.e. relative strengths or weaknesses) is not altered, but is moved down as a whole so that the pattern of I.Q. group mean error ratio scores passes through the middle. If each individual pattern of deviations is thus adjusted up or down so that the I.Q. group means pass through the middle, then the individual deviations on any category become directly comparable (i.e. the overall error differences are ruled out). In the case of Otis II above, the adjustment is in the opposite direction, i.e. moving up the whole pattern of deviations by adding 0.1.

In applying the above adjustments, two assumptions are made:- (a) That if this boy had made fewer errors on Otis I the reduction would be distributed equally over all categories. In fact the whole amount of 4.4 overall (0.4 x 11) may have been reduced on one category only, or 2.2 on two categories or 1.1 on four categories etc. There is no way of predicting
accurately how this reduction would have been distributed, but it is safer to assume that it would follow the pattern of the group as a whole. It is better to depress the deviations in the direction of the group pattern and thus perhaps eliminate some cases from consideration later on in "high error (H.E.) and "low" error (L.E.) groups, than to increase some deviations and thus perhaps include some spurious cases in the H.E. and L.E. groups.

(b) That even granted an equal distribution of the mean error ratio 'difference' over all categories, the distribution is still unaffected by differences in category means. This assumption cannot be upheld because the pattern of group means shows considerable variations, e.g. Otis II, Group 4, Category (1), Mean error ratio = 4.2; Group 4, Category (7), Mean error ratio = 1.8. Thus if, for example, an individual pattern of deviations has to be adjusted upwards, i.e. an amount added to each category deviation, then clearly category (1) should be allotted a greater proportion of the total amount than category (7). Ideally each total adjustment should be broken up, so that the amount added to, or subtracted from each category, would be directly proportional to the size of the mean error ratio score for that category, in the particular I.Q. group into which the case falls. A trial distribution from 0.1 to 1.0 showed that the differences between a 'uniform' adjustment and a 'proportional' adjustment
were so small in adjustments of 0.9 and below that the added accuracy was not justified by the considerable amount of time involved in making such proportional adjustments. On the other hand, for adjustment of 1.0 and above, the use of uniform adjustments could result in an error of up to 0.9 (on category (9)). In order to reduce such errors, a correction table was used (see "correction" table at the end of this appendix), in which all proportional adjustments involving a difference (over uniform adjustments) of over 0.4 were listed. There were, however, only 22 corrections of 0.5 or over out of a possible 154 (i.e. 11 categories x 2 Otis tests x 7 groups), and these corrections were applied to the adjustments whenever appropriate. The size of adjustments made is indicated by the means of these adjustments - i.e. for Otis I, 0.88; for Otis II, 0.67.

Step (4).

At this stage the two adjusted deviations (Otis I and II) were added to give the last column (see example above). This step resulted in an increase, where deviations were in the same direction, and a decrease or zero score where the deviations were in the opposite direction, thus providing for a safeguard against the inclusion of inconsistent deviations.
CORRECTION TABLE.

TO CORRECT UNIFORM ADJUSTMENTS WHERE THESE WOULD BE DISTORTED 0.5 OR MORE BY ABSOLUTE DIFFERENCES IN THE MEAN ERROR RATIOS OF THE CATEGORIES.

a = correction for 1.0 - 1.4,  
b = correction for 1.5 - 1.9.

<table>
<thead>
<tr>
<th>Categ.</th>
<th>Otis</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>+.5</td>
<td>+.7</td>
<td>+.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>+.6</td>
<td>+.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I</td>
<td></td>
<td></td>
<td>-.6</td>
<td>-.6</td>
<td>-.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No correction</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+.6</td>
<td>+.6</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+.5</td>
<td>+.7</td>
<td>+.6</td>
</tr>
<tr>
<td>5</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No correction</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No correction</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td>-.5</td>
<td>-.7</td>
<td>-.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I</td>
<td></td>
<td>-.6</td>
<td>-.9</td>
<td>-.6</td>
<td>-.6</td>
<td>-.9</td>
<td>-.6</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No correction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I</td>
<td></td>
<td>+.9</td>
<td>+1.3</td>
<td>+.6</td>
<td>+.9</td>
<td>+.9</td>
<td>+.9</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td>+.6</td>
<td>+.6</td>
<td>+.6</td>
<td>+.6</td>
<td>+.6</td>
</tr>
<tr>
<td>10</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+.6</td>
<td>+.6</td>
<td>+.6</td>
</tr>
<tr>
<td>11</td>
<td>I</td>
<td></td>
<td>-.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td>+.5</td>
<td>+.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The plus sign indicates that the uniform adjustment has been too small; minus sign, that it has been too large. Thus, for example, if on Otis I the difference between the overall mean error score for a group 3 subject and the overall mean error score of his group is 1.0, then after the uniform adjustment of 1.0, added or subtracted according to whether he is below or above the group mean, a further correction of .6 must be made on category (9) in the same direction as the original adjustment. On the other hand, under the same conditions the correction for category (7) would be .5 in the opposite direction to the original adjustment.
### APPENDIX E.

**M.V.s FOR OTIS I AND II**

**ERROR RATIO SCORES.**

<table>
<thead>
<tr>
<th>Group</th>
<th>Otis 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>.8</td>
<td>1.2</td>
<td>1.3</td>
<td>.7</td>
<td>1.4</td>
<td>2.5</td>
<td>1.7</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>1.3</td>
<td>.8</td>
<td>1.4</td>
<td>1.1</td>
<td>.8</td>
<td>.6</td>
<td>1.0</td>
<td>1.9</td>
<td>1.4</td>
<td>1.3</td>
</tr>
<tr>
<td>2</td>
<td>1.7</td>
<td>1.4</td>
<td>1.2</td>
<td>1.5</td>
<td>1.7</td>
<td>1.5</td>
<td>.8</td>
<td>1.5</td>
<td>3.2</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.7</td>
<td>1.2</td>
<td>.8</td>
<td>1.4</td>
<td>2.4</td>
<td>2.3</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>1.4</td>
<td>1.6</td>
<td>1.3</td>
<td>1.7</td>
<td>1.9</td>
<td>1.6</td>
<td>.8</td>
<td>2.3</td>
<td>3.7</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>1.5</td>
<td>1.2</td>
<td>.6</td>
<td>1.9</td>
<td>2.8</td>
<td>2.2</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>1.7</td>
<td>1.9</td>
<td>1.5</td>
<td>2.1</td>
<td>2.4</td>
<td>1.7</td>
<td>.7</td>
<td>1.8</td>
<td>3.8</td>
<td>2.5</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>1.7</td>
<td>1.3</td>
<td>2.1</td>
<td>2.2</td>
<td>1.7</td>
<td>.9</td>
<td>1.3</td>
<td>3.1</td>
<td>2.7</td>
<td>1.8</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>1.7</td>
<td>1.8</td>
<td>1.8</td>
<td>2.6</td>
<td>1.6</td>
<td>1.0</td>
<td>2.2</td>
<td>4.1</td>
<td>3.1</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>1.6</td>
<td>1.4</td>
<td>1.9</td>
<td>1.9</td>
<td>1.7</td>
<td>.8</td>
<td>1.9</td>
<td>3.1</td>
<td>2.4</td>
<td>2.4</td>
</tr>
<tr>
<td>6</td>
<td>1.4</td>
<td>2.3</td>
<td>1.4</td>
<td>2.0</td>
<td>1.9</td>
<td>1.7</td>
<td>.9</td>
<td>1.8</td>
<td>4.3</td>
<td>2.4</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>2.3</td>
<td>1.9</td>
<td>2.0</td>
<td>2.0</td>
<td>1.6</td>
<td>.8</td>
<td>1.6</td>
<td>3.3</td>
<td>2.9</td>
<td>2.3</td>
</tr>
<tr>
<td>7</td>
<td>1.5</td>
<td>3.2</td>
<td>1.5</td>
<td>2.3</td>
<td>1.6</td>
<td>1.4</td>
<td>1.1</td>
<td>1.7</td>
<td>1.8</td>
<td>2.7</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>2.4</td>
<td>1.3</td>
<td>2.2</td>
<td>1.4</td>
<td>1.6</td>
<td>.9</td>
<td>1.9</td>
<td>2.7</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

**N of MVs**

<table>
<thead>
<tr>
<th>Group</th>
<th>Otis 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.5</td>
<td>1.9</td>
<td>1.4</td>
<td>1.8</td>
<td>1.9</td>
<td>1.5</td>
<td>.9</td>
<td>1.8</td>
<td>3.3</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td>II</td>
<td>1.3</td>
<td>1.7</td>
<td>1.3</td>
<td>1.7</td>
<td>1.7</td>
<td>1.4</td>
<td>.8</td>
<td>1.6</td>
<td>2.8</td>
<td>2.3</td>
<td>1.9</td>
</tr>
</tbody>
</table>
INSTRUCTIONS AND RECORD SHEET FOR RATINGS ON CAREFULNESS - CARELESSNESS.

A RATING SCALE OF 'CAREFULNESS - CARELESSNESS'.

This is part of an investigation concerned with the way children approach and carry out various tasks. The particular information you are asked to provide is a rating, for each child in your class, on a characteristic called carefulness and its opposite carelessness.

In talking of particular children, teachers often express the opinion that "so and so is a careless worker". or "he's not one of my brightest children, but he's a careful worker". or "he's very slapdash in his approach to work". or "you can always rely on him, he's careful in everything he does."

Do not assume that carefulness necessarily goes with high intelligence and carelessness with low intelligence; a child at any level of intelligence may stand out as being particularly careful or careless.

When making your judgments try to form a picture of the child's typical way of doing things, covering not only school work but any other activities which may reveal a careful or a careless approach to tasks.

Although it is difficult to make accurate ratings, it is assumed that a child can be placed somewhere along a scale ranging from very careful at one end to very careless at the other. The actual scale is divided into eleven parts and the descriptions beneath the scale are merely an aid to defining, roughly, distances along it. Try to form a general impression of the child, then place a small tick in one of the eleven boxes.

Here is the scale, with a sample tick:—

| ☑ |   |   |   |   |   |   |   |   |   |   |
---|---|---|---|---|---|---|---|---|---|---|
Stands out as being careful | On the whole careful, but needs an occasional reminder | Does not stand out as careless but responds temporarily | On the whole careful or careless | Stands out as being careless in practically all situations.
Now think of a child who would represent the extreme left of the scale, and one who would represent the extreme right....

Then think of one for each of the intermediate positions described.

This should help you to clarify and fix in your mind the nature of this scale and the particular characteristic it measures. Now go ahead (on the next page) and put a small tick in one of the boxes for each child.
Keep the descriptions of the 5 positions on the scale (see previous page) beside you as you work through this list.

Although the scale below is much smaller than that given as an illustration, the relative positions and the number of boxes are exactly the same. The scale has been reduced in size merely for convenience.

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| B.G | ☑ |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| B.M |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| B.B |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| B.P |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| C.W |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| D.B |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| E.S |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| H.E |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| I.D |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| K.R |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| K.W |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| M.T |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| M.V |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| E.B |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| O.K |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
APPENDIX G.

COMPARISONS OF SCORES ON OTIS I AND II, P.M.A. AND INTERMEDIATE D.

The following means, standard deviations and Pearson Product-Moment correlation coefficients are based on the scores of 200 children who completed the four tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>M</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otis I</td>
<td>104.8</td>
<td>12.8</td>
</tr>
<tr>
<td>II</td>
<td>108.2</td>
<td>13.4</td>
</tr>
<tr>
<td>Int. D</td>
<td>102.0</td>
<td>12.8</td>
</tr>
<tr>
<td>P.M.A (Q)</td>
<td>102.6</td>
<td>14.2</td>
</tr>
</tbody>
</table>

r. Otis I and Otis II  .90 ± .01
r. Otis II and Int. D  .88 ± .01
r. Otis II and P.M.A(Q) .70 ± .02
r. Int. D and P.M.A (Q) .70 ± .02

According to the Manual for the A.C.E.R. Intermediate D Test, the correlation between the Otis B and the Intermediate D, based on 809 Grade IV metropolitan children in Queensland (a younger age group than this sample) was .83, with a standard deviation of 13.8 (p.13).

Two correlations between the Otis and the P.M.A. (Q) score, have been reported in New Zealand. Ballantine reported a correlation of .696 on 200 cases, and Sexton a correlation of .73 on 100 cases. The correlation of .70 given above, is consistent with these two studies.

---


In view of the fact that Ballantine included a comparison of the scores of boys and girls in the P.M.A. subtests and Q, the following figures drawn from his table are compared with the scores of the 271 children who completed the P.M.A. test in this study.

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>101</td>
<td>99</td>
</tr>
<tr>
<td>Md. Percentile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>62</td>
<td>56</td>
</tr>
<tr>
<td>S</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>R</td>
<td>32</td>
<td>65</td>
</tr>
<tr>
<td>N</td>
<td>54</td>
<td>59</td>
</tr>
<tr>
<td>W</td>
<td>36</td>
<td>40</td>
</tr>
<tr>
<td>Q</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>163</td>
<td>108</td>
</tr>
</tbody>
</table>

There are some serious discrepancies in these results, although the sex differences are in the same direction in both studies. It is possible that the way in which the test is administered—in particular the checking of the practice examples in order to make sure that all children are quite clear what to do—may have a considerable influence on the scores. Coppell noted that boys were superior to girls on V and S while girls were superior to boys on N—thus there is some general consistency in the direction of the sex difference reported in these studies.

---


APPENDIX H.

CLASSIFICATION AND ILLUSTRATION OF ERRORS ON ALL ITEMS, BY CATEGORIES.

For each item, the distribution of errors for the total group on Otis I has been given \((N = 248\), i.e. those who completed both Otis tests\). \(N\) = number of errors on that item for the total group and the figures for each distractor or wrong response represent percentages, unless otherwise stated.

Exact frequencies for the types of errors illustrated have not been given, because it was not possible to classify all errors accurately, furthermore such frequencies would be distorted by the overlap of two or more types of errors on the one item. Where possible, however, a general indication of frequency has been given along with any sex or I.Q. group differences.

The following records may present an understatement of the prevalence of some types of errors, for these children were working on the same items for the third time within a year, and thus they may have corrected some of their earlier errors and developed a certain amount of test sophistication, especially in view of the fact that they had also worked the five subtests of the P.M.A. test.
(A) Category (1) - Word Meaning.

Item 5.
If we feel sorry for the suffering of another, we have a feeling of (?).
(1) hate, (2) pity, (3) contempt, (4) disdain, (5) eagerness.

N=20. (1) 25, (3) 10, (4) 15, (5) 50.

Although only two children repeated the error in the "interviews", the two errors made are of interest:-

B. 103/99 (3) (i.e. Boy; I.Q. 103 on Otis I, 99 on Otis II; distractor chosen - 3;). When questioned and asked to describe what the sentence meant, it was clear that he understood it, and his reason for choosing (3) was that he thought that the answer was likely to be one of the unfamiliar words, which he did not know - (3) and (4) - so he chose (3). His error was at least partially due to an assumption made about the nature of the distractors offered - perhaps also due to a lack of confidence.

G. 72/76 (4) and (1). Although no reason could be elicited for the choice of (4), which was spontaneously changed to (1), it appeared that (1) was chosen because it was the only one of the five words which was known. The sentence was only partially understood, the particular error arising from the satisfaction of choosing a familiar answer, even although the whole task was unclear. (Cf. Hamid - termination of emotional tension in dull subjects - see above p98).

Item 16.
When a new kind of machine is thought of, it is usually called (?).
(1) a discovery, (2) an adoption, (3) a creation, (4) a novelty, (5) an invention.

N=35. (1) 74, (2) 6, (3) 11, (4) 9.

Most children gave the correct answer on reworking, although three group 7 children showed a similar tendency on repeating their original errors.

G. 86/88 (1). When asked to give the meaning of each of the five words, she was not very confident, but when she had described (5) she spontaneously changed her answer to (5), and
explained that she had not really looked past (1) originally, because it seemed to be quite good as an answer.

B. 77/79 (1). A similar situation to that above - although he did not know (2), (3) and (4), he defined (5) correctly, spontaneously changed his answer to (5) and volunteered the information that he had not read past (1).

G. 72/76 (2). Gave a poor but essentially correct description of (1), (2) and (5), and changed to (5) spontaneously.

Item 23.

A boy who often tells big stories about what he can do is said to (?).
(1) lie, (2) fake, (3) cheat, (4) joke, (5) brag.

N=169. (1) 73, (2) 17, (3) 4, (4) 6.

Children in all 7 groups made many errors on this item, and from questioning it was clear that the majority did not know the meaning of "brag" - e.g.:-

G. 122/127 (1). Did not know meaning of (5) so decided that the only suitable answer was (1).

B. 99/105 (1). Gave correct definitions for all except (5), which he had never even heard before.

There were however, other errors worth reporting:--

(a) Not reading past the first distractor - (Cf. item 16).
(b) Not reading the whole sentence, but only the first half - e.g.:-

B. 122/116 (2). When asked his reason for choosing (2) he confidently pointed out that fake means "not the truth" and "big stories" means the same, hence (2) is correct. When asked to read the whole sentence aloud he saw the phrase "what he can do" and explained that when answering the item he had not read past "big stories". A similar situation arose with B 124/123 who had chosen (3), i.e. cheat = not truthful = "big stories" - but in fact he had not read past "big stories" and having read the whole sentence aloud, chose (5). The same condition led G. 101/100 to choose (1) although she understood (5).
(c) Rejection of 'brag' as inadequate e.g.:-

B. 126/128 (1). When explaining his choice he argued that to brag is to "skite", and this is not necessarily telling big stories, because a "big story" means a "lie". This explanation seems reasonable in view of the fact that colloquially "big story" is often used synonymously with "lie" especially with young children. Similar arguments were put forward by G. 114/117 (1) - brag = boast, so cannot be a big story or lie, and G. 105/101 (1), who also equated brag with boast.

(d) Misreading e.g.:-

B. 110/113 (2). When explaining his answer he read the sentence as "is said to be a (?)", and the only two words which fitted this were (2) and (3), and of these, (2) seemed the better. Such a misreading may easily happen in reading silently under pressure of time, but this boy still included the words "be a" when he read slowly and aloud. Other misreadings were B. 90/92 (2) "...because 'brat' has nothing to do with it," - and a misreading of "bag" for "brag" -(G.72/76).

Item 40.

An event which is sure to happen is said to be (?).

(1) probable, (2) certain, (3) doubtful, (4) possible, (5) delayed.

N=72. (1) 33, (3) 17, (4) 40, (5) 10.

All groups made errors on (1) and (4), although the errors on (3) and (5) were mainly confined to groups 4 to 7. There were four types of errors noted:-

(a) Failure to understand the implications of all the words in the sentence - i.e. only part of the task functioning.

G. 110/123 (1). She was asked to give the meanings of the five words and in defining "certain" said - "something that is sure to happen...oh!...so it must be (2)". Her attention had been focussed on "happen" and she was quite satisfied to choose a related word until the task was clarified by the introduction of a definition of "certain."

B. 100/105 (1). After he had been asked to read the sentence aloud and to point to the important part of the sentence, he said "sure to happen...oh!...so it must be certain". Even
the act of reading aloud was enough to clarify the requirements of the task for B. 100/112 (1) who immediately changed to "certain" although no suggestion was made that his first response was wrong.

In these cases it was clear that the prepotent word was "happen", although as soon as attention was focussed on the sentence or the alternatives again, the true nature of the task was seen.

(b) A variation of this type of error was the prepotency of "probable-possible" as a pair, e.g.:--

G. 106/104 (4) explained that as soon as she read the sentence and saw that it had something to do with "happen", she decided that the item was included in order to make her choose between two similar words, and although hesitant she finally chose (4). When asked the meanings of the other alternatives she immediately changed to (2) and claimed that she had not even seen it before.

(c) Going beyond the data, e.g.:--

G. 105/105 (5), when questioned about her choice said, "because delay usually means doing something without delay, which means it is sure to happen." Asked to look over the other words she changed straight away to (2). In this case the word "delay" suggested a familiar phrase and this took the place of a careful analysis of all the words given, although when her attention was directed to them she immediately picked the correct one.

B. 97/104 (3), - gave as his reason, "Well, it might happen and it might not happen," i.e. it seemed that he was trying to verify whether or not one could say an event was sure to happen, and he decided one couldn't be sure. When asked to read the question and see what it really asked for, he gave (2). A very similar reason was given by G. 100/109 who explained her choice of (3), "...because the event mightn't happen at all." When the item was read for her she chose (2) - her original choice being due to a vaguely formulated idea that she had to comment on the truth of such a statement as that given. Although doubtless due to a less sophisticated assumption, the choice of (3) by G. 83/83, was explained in much the same fashion - "Well you can't know
that it will happen."

(d) Not reading all the words given, e.g.:-

B. 126/116 (1). - "I took the first one that looked something like it."

G. 105/101 (1). - "Saw 'probable' so did not look over the others."

Item 44.
A quantity which grows smaller is said to (?).
(1) fade, (2) decrease, (3) dry up, (4) die, (5) sink.

N=91. (1) 30, (3) 31, (4) 14, (5) 25.

Errors were more evenly distributed over the four distractors in this item, due to the fact that where "decrease" was not known, a reasonable argument could be put forward for each of the other words. There was a distinction between those who confused "decrease" with "increase" (mostly in the higher groups), and those who did not know the word at all (mostly in the lower groups). Examples of particular errors made show some of the processes operating in individual cases.

(a) Concrete image of the item favouring certain distractors: e.g. -

G. 90/96 (3) - "Because as things get smaller they dry up and then you have to dig them out - they all shrivel up" (i.e. thinking of vegetables etc.). When asked directly the meaning of "decrease", she defined it as "to get smaller" and changed to (2) straight away. A similar answer was given by B. 77/79 who pointed out that when things die they shrink and get smaller - however he did not know the meaning of "decrease."

G. 83/83 (3) thought of quantity as "a lot of ingredients mixed up together...if you leave them they dry up" (i.e. cooking experience). Although she did not know the meaning of "decrease", her error was far from being random or due to a superficial association, but within the limits of her knowledge, quite reasonable.
G. 86/88 (3) thought of "quantity" as "...well, a quantity of ink or milk" - i.e. something which dries up as it gets smaller. When asked the meaning of "decrease" she said - "means...oh...get smaller, because increase is to get larger...so it will be (2) after all...", i.e. the visual image of "quantity" displaced the analysis of the correct word although she knew it all the time. A similar failure was noted with B. 117/118 who could not decide between (2) and (3) because he thought "quantity" referred to liquids.

(b) Misplaced emphasis, e.g:-

G. 93/100 (5) "...because a quantity is sometimes huge and sometimes smaller, and the smaller one would sink down." When asked to describe the other four words she gave a correct definition for "decrease" and on reading the question aloud, immediately changed to (2), i.e. she had concentrated on the properties of "quantity" instead of on the act of growing smaller.

(c) Some did not get as far as even attempting to answer, because they were held up by the word "quantity", while B. 82/83 (1) when asked to read the item, read "fade" as "fail" and could not pronounce "decrease". When the five words were read for him he chose the correct one and showed that he understood its meaning. G. 92/84 (4) gave correct definition of "decrease" and explained that she had not bothered to think about it the first time.

**Item 54.**

A person who wishes very much to succeed but fears he will fail is said to be (?)

(1) earnest, (2) anxious, (3) industrious, (4) energetic (5) cowardly.

N=106. (1) 14, (3) 14, (4) 14, (5) 59.

The general pattern for all groups was similar on (1), (3) and (5), but on (4) groups 1 and 2 made proportionately more errors. Throughout all groups an inadequate definition
of "anxious" was evident, although this definition is in common usage - e.g.:

G. 117/121 (1) - defined (2) as "keen to...wants very much to", but rejected it because it did not fit in with "fears he will fail". Other examples are: B. 100/105 (1) - defined (2) as "dying for something to come .... e.g. the boy was anxious for his birthday to come"; B. 113/118 (4) rejected (2) because it means "hoping for something very much," while he chose (4) "because it means got your mind set on one thing". This choice of (4), largely confined to the two top groups (i.e. 8 of the 15 choices of (4) were in these groups) may be due to the unfamiliarity of the word in the lower groups and to the assumption that a person needs to be energetic if he is to avoid this failure. From the examples given, it was clear that many of the children were in the habit of hearing "anxious" used in the colloquial way illustrated above.

Apart from this inadequacy, the remainder of the errors fell into two broad groups, i.e.:

(a) Those who ignored the second part of the sentence and found an appropriate word for wishing to succeed, e.g.:

G. 130/125 (1) - "...he has got to be determined so will be earnest." When asked to define (2) she gave - "...anxious doesn't mean 'wish to succeed', but worried", i.e. she was ignoring the second half and concentrating on the first.

B. 124/123 (3) explained that if a person wishes to succeed he must be industrious, and did not even see the relevance of the second part until his attention was specifically drawn to it.

G. 93/100 (3) but changed to (2) "...because it means 'really wants to' and this is the same as 'wishes very much'". This inadequate definition of (2) along with the ignoring of the second half, gives the correct answer; thus there were probably a number of children who gave the correct answer for this item, on the wrong grounds.

(b) Those who ignored the first half of the sentence and found an appropriate word relevant to fear of failing, e.g.:---
G. 117/118 (5) "...because he is scared to fail - he is cowardly". This girl even after rereading aloud could not see that the first half of the sentence was relevant. Similar cases of prepotency of the second half occurred at all I.Q. levels, even with some who could give a correct definition of "anxious".

Item 61.
An object or institution that will not perish or cease is said to be (?).
(1) permanent, (2) stable, (3) stationary, (4) solid, (5) sound.

N=66. (2) 2, (3) 26, (4) 62, (5) 11.

As with item 44, there was a distinction between those who did not know the meaning of "permanent" at all, and those who confused it with "temporary", however the sources of error were usually more complex.

(a) Some argued that although "permanent" implied something lasting, it did not eliminate the possibility of perishing at some time in the future, e.g.:-

B. 102/100 (4) gave the meaning of (1) as "to stay there" and when asked if he considered (1) before choosing (4), he said "...solid means won't perish, but something permanent might fall down, there might be the ruins of something that was permanent."

G. 130/125 (4) "permanent' means it could remain there but doesn't mean it couldn't perish."

G. 116/120 (4) explained that "an object which is solid does not cease," but although she gave a reasonable definition of (1) she was not prepared to concede that something permanent could not perish or cease at some future time.

B. 116/116 (5) argued that "sound" (i.e. "sound clothing... very strong, good wearing") was better than (1), although he defined (1) as "something that is there all the time," as opposed to temporary. His example of something permanent was "bicycle tube", and when asked if this would perish sometime he agreed that it would, and still held to (5).
G. 115/116 (5) gave as an example of (1) "permanent resident" - as contrasted with a casual guest - "someone who lives there and is not just staying." On these grounds she rejected (1) because it did not fit in with perish or cease to be.

All these examples draw upon common usages of the word "permanent", and it could be argued that some of these children have made errors because they have seen the more subtle overtones of the word as commonly used, whereas the definition offered in the sentence does not allow for them.

(b) "Solid" was chosen in the majority of cases, because the word is often used to indicate stability, hardness, lasting quality. There is something "long lasting" about "solid" even though it is not really a necessary characteristic, e.g.-

B. 124/123 (4) - "Well, if there is a stone in the ground, and stuck, it won't perish or cease".

G. 117/118 (4) chose "solid" because it represents hardness. When asked to describe the other words she defined (1) as "stay there and can't shift" and spontaneously changed to (1), i.e. she was attracted to the quality of "solid" so did not stop to analyse the other words. A similar choice was made by B. 95/96 because of the long lasting nature of "solid", although his definition of (1) was "stay there all the time."

(c) A good reason for rejecting (4) is the inclusion of "institution" in the definition. All the children who had originally made errors on this item were questioned on the meaning of "institution". Some of those who gave the correct response had no idea what "institution" meant, and like the majority, who still made an error on the item, they completely ignored the word. Of those who gave (4) as their answer, those who took "institution" into account had defined it as "place" or "thing". In this item, therefore, one of the real
difficulties was caused by the use of the word "institution" which was more unfamiliar than "permanent" and once "institution" was ignored the chances of choosing (4) instead of (1) became greatly enhanced.

(d) As in the preceding items, some children passed over the correct answer although they had good understanding of it, e.g:-

G. 117/118 (4) as above; also B. 115/110 (2), who when describing the other words, defined (1) as "stay there all the time", and then on rereading the question said "...oh...permanent...I just wasn't thinking"; and B. 110/118 (4) who when asked to describe the other words, started with (1) - "stays there for ever" and changed spontaneously to (1) instead of (4).

(e) Other children misread words or could not pronounce them, e.g:-

G. 118/115 (5) knew the meaning of (1) but would not accept it because she did not think a "permanent" thing would "perish"! This interpretation resulted from a misreading - i.e. when asked to read the sentence aloud she hesitated over "will not perish" because in her previous silent reading of the statement she had read "which will perish". B. 98/96 (4) read "instruction" for "institution" - and as it sounded irrelevant he just ignored it; G. 100/109 (4) misread "stables" for "stable", and could not pronounce "permanent", but when it was pronounced for her she gave its meaning as "for ever" and changed to (1); G 96/96 (4) was also unable to pronounce "permanent" but when it was pronounced for her she explained it as - "...can mean you are permanently put into something - there for good"; B. 100/112 (4) gave the following meaning when the word was pronounced - "there for...kind of resting there...like, not on a holiday", i.e. although he made an error, he was by no means ignorant of the general context within which "permanent" would be appropriate.

Although these examples illustrate the fact that answering an apparently straightforward question is not an "either-or" process, they go even further than this in show-
ing that the difficulty of the question arises from different sources for different children.

Item 75
An agreement reached in which both sides yield somewhat in their demands is called (?)

(1) a promise, (2) a compromise, (3) an understanding, (4) a deadlock, (5) an armistice.

N=34, (1) 44, (3) 24, (4) 3, (5) 29.

The word "compromise" was not known, and in all cases was pronounced as "comprouli se". Although most of the children who had attempted the item were in the upper groups (1-4), it was clear from their explanations that they had chosen (1) because it fitted in with "agreement", while the rest of the sentence was ignored.

One boy, 117/118 (1) or (3) found the most difficult word in the sentence to be "somewhat" and his attention was directed towards this word and away from the general meaning of the sentence.

Those who chose (3) did so for the same reason as those who chose (1), e.g.:-

B. 116/116 (3) "...because agreement has most to do with understanding". When questioned about the rest of the sentence he said that after a quick glance, he did not bother to read it again to see what it meant.

Among those who did attempt to understand the whole sentence, the choice of (5) was well supported, e.g.:-

B. 108/115 (5) "...because it seemed to fit the armistice made by the Americans and Koreans."
(B) Category (2) - Opposites.

Item 3.
Which one of the five words below means the opposite of east?
(1) north, (2) pole, (3) west, (4) equator, (5) south.
N=34. (1) 41, (2) 9, (4) 15, (5) 35.

There were only two cases in which the error was repeated, the first of these probably representing the type of error which would account for the high percentages on (1) and (5) e.g.:-
G. 108/114 (5) - did not know the four points of the compass and opposed N-W, E-S.

The second error resulted from a more devious interpretation - e.g.:-
B. 90/92 (4) - reasoned thus - "...there is a north pole, a south pole - there is no east pole, so that must be the equator"; i.e. this boy was trying to visualize where east stands in relation to the poles, and in fact if the compass rose is superimposed on a globe, east and west fall on the equator. When asked to read the question aloud, and to look at "opposite of" he answered correctly straight away, in other words he had not clarified the task but constructed his own problem out of the five words given.

Item 8.
Which word means the opposite of succeed?
(1) win, (2) decline, (3) fail, (4) accede, (5) try.
N=64, (1) 28, (2) 6, (4) 52, (5) 14.

Four sources of error could be distinguished:-
(a) Giving a word which is the "same", instead of "opposite" e.g.:-
B. 95/103 (1) and (5) "...because 'win' and 'try' mean about the same as 'succeed'" When asked to read the question aloud he gave the correct answer immediately. A similar situation arose with G. 80/86 (1).
(b) Misreading a word, e.g.:-

B. 90/92 (4) - read "succeed" as "suicide". and as (1), (2), (3) and (5) did not seem to fit, he chose "accede" because he did not know it, and therefore it could have something to do with suicide. When "succeed" was pronounced for him he immediately gave the correct answer.

B. 89/93 (2) - read "succeed" as "ceased" and explained his choice of (2) - "...because it's like a train going along the line and it jumps off" - when asked what this has to do with "decline", his answer made it clear that he had read it as "disline" - i.e. the train goes off the line, thus it is "dis-lined", hence it has "ceased" to move. When the question was read to him he gave the correct answer - his error had resulted from failure to understand the task and a misreading of words, not from a lack of knowledge of the relation between "succeed" and "fail".

(c) Attraction of the unfamiliar "accede", e.g.:-

G. 107/112 (4). When questioned, it was clear that she knew the meanings of all the words except (4), and on re-reading the question gave the correct answer. She explained that she saw "accede", did not know it, so thought it must be the right one. A similar error was made by G. 111/106, and B. 97/98.

(d) Complete lack of task formulation, e.g.:-

B. 76/80 (4), when asked the opposite of "succeed", said "fail", then when directed to look at the instructions again, he gave (5) as his answer. Not until the question was read for him with emphasis did it occur to him that "fail" was the actual answer.

G. 72/76 (4) - gave her answer very quickly - i.e. superficial association of the two words, and had no idea what the task really required. When put in a familiar context of school work, she gave 'fail' correctly.

Item 24.
Which one of the five things below means the opposite of difficult? (1) hard, (2) quick, (3) soft, (4) easy, (5) common.

N=41. (1) 66, (2) 7, (3) 5, (5) 22.
Few errors were repeated, but those that were could be divided into two types:–

(a) Giving "same" instead of "opposite", e.g.:–

G. 97/98 (1) "...because it's a word about the same as difficult." When asked to read the sentence again, aloud, she still gave (1), having completely overlooked the crucial word "opposite", and did not get the correct answer until asked the opposite of "difficult" in an informal sentence. This was not merely a case of misreading due to haste, but rather of a particular assumption persistently distorting perception of the task. Similar errors were noted for B. 89/93 and B. 93/95.

(b) Failing to look carefully at the choices offered, e.g.:–

G. 93/89 (3) - "... 'difficult' might be something hard to do, so the opposite word would be something easy to do, so choose 'soft'" – when her attention was drawn to (3) and (4) she chose (4). This girl gave the correct answer in her own explanation but had not looked at all the words given, and therefore failed to see that "easy" was in fact one of the choices.

B. 91/96 (3) - this was also his original error, but before he could be asked to explain, he changed to (4) and remarked that he decided to look over the words again, as he had not really looked at them carefully before, and immediately noticed (4).

Although in the total group, proportionately more errors were made by boys than girls on this item, the reworking of errors did not throw any light on this.

Item 63.
Which word means the opposite of pride?
(1) sorrow, (2) humility, (3) miserable, (4) conceit, (5) proud.

N=81. (1) 41, (3) 35, (4) 13, (5) 11.

Only 3 children out of 30 gave the correct answer on reworking the question. The difficulty in most cases was not the understanding of the opposite of "pride" – for most
described this reasonably well, - but the unfamiliarity of the word "humility". Either the word was abandoned, as completely unknown, or it was defined as -"funny person; human; gentle; poor but well dressed; amusing; got a lot of humour in him; full of laughter; humiliation; joyful; made of fool of."

In the circumstances (1) or (3) was most suitable, although errors due to the giving of the "same" word occurred, e.g:-

G. 117/118 (5); G. 107/112 (4); G. 111/106 (5) - the wrong approach remaining unnoticed until specifically pointed out.

There were a few particular cases worth recording, e.g:-

G. 115/116 (3) - changed spontaneously to (2) "...because I looked again and saw (2)" - presumably her original error - (3) had been due to the same inadequate perusal of the choices.

B. 94/102 (2) - i.e. correct answer - "Humility means doesn't care about something, just doesn't take pride in books - is careless about them." Another example of a correct answer, although not really understood, is B. 83/95 (2) - "...just picked (2) because the others looked wrong" - he did not of course understand the meaning of "humility".

G. 113/115 (2) - i.e. correct answer; although she had no idea of the meaning of humility she could see that the others did not fit. Exactly the same process resulted in the choosing of (2) by B. 83/91.

(C) Category (3) - Word or Phrase Attribute or Definition.

Item 2.
Which one of the five words below tells best what a hammer is?
(1) thing, (2) tool, (3) furniture, (4) weapon (5) machine.

N=13. (1) 62, (3) 8, (4) 15, (5) 15.
There was only one repeated error on this item - i.e. - B. 108/110 (1) who took the first relevant word and did not consider the others - i.e. failure to make use of all the material.

Item 14.
Which statement tells best just what a gate is?
(1) A hole in a fence, (2) Something to swing on, (3) It has hinges, (4) A Door in a fence, (5) It opens and shuts.

N= 78. (1) 3, (2) 4, (3) 8, (5) 85.

From the reworking, it appeared that the main difficulty was in thinking of the function of a door, for example these objections to (4) were made:-

G. 114/117 (3) - "...it isn't big enough for a door and it hasn't a handle", B. 115/110 (5) - "...a door usually has a part of the wall over it at the top, but a gate hasn't", G. 106/104 (5) - "...you don't open it like a door", G. 97/98 (5) - "...a door is usually on a house, but a gate isn't", G. 93/89 (5) - "...you can't see through the door in the fence" (i.e. thinking of a glass door as in classroom), B. 80/86 (5) - "...couldn't be a door because there mightn't be anything to hold it, it might be just a springy one" (i.e. thinking of a door as something with a catch or lock).

Basically, all these objections arose from the fact that the children were thinking in terms of the salient characteristics of a gate as an object. When the question was reworded as - "If you wanted to tell someone who had not ever seen a gate what it was for and what it was like, which would be the best answer?" - almost all of them chose (4). This rewording does of course make the item easier, but it appears that the whole item hinges on the interpretation of "tells best", and this may be interpreted as "description", E.g:-
G. 112/114 (5), even after having her attention drawn to (4) insisted that (5) was better "because (4) could be anything, but (5) tells what a gate does", similarly B. 82/83 (5) "...because it tells what it does." In the case of G. 100/109 (5) the item was read for her with emphasis on "tells best", upon which she changed to (4) and commented that on her first reading she just "got the idea of something about a gate."

Item 18.
What is the most important reason that motor cars have displaced horses and carriages?
(1) Horses were getting scarce, (2) Horses often run away, (3) Motors save time and "time is money", (4) Motors are cheaper than carriages, (5) Motors cost less to repair than carriages.

N=71. (1) 25, (2) 41, (4) 4, (5) 30.

It was found difficult to elicit the reasons for the various choices, and hard to judge exactly what processes were involved. In each of the cases where the error was repeated after reworking, the question was asked "Have you checked over the five answers to make sure that you have chosen the most important reason?" As a result of rechecking, eight of the ten children changed to the correct answer, and further questioning established the fact that they had not really considered the question with the emphasis on "most important", but rather as "give a reason why". In some cases they were attempting to judge the question in terms of the first motivations which might prompt a man to change to a motor car, and considered the possibility that horses might have been scarce - a not unreasonable assumption for some city children. One girl - i.e. G. 93/89 (3) - gave the correct answer, but her reason for choosing this was - "If you go in a bus it
costs you money, if you went in a car it wouldn't cost you anything," - hence you save money - thus (3)!

**Item 27.**

**Which tells best just what a horse is?**

(1) It has a tail, (2) a live thing, (3) a thing that works and eats, (4) a large four legged animal, (5) something to pull a wagon.

N=94. (1) 0, (2) 14, (3) 17, (5) 69.

Apart from the common case of failure to read through all the choices offered (e.g.: G. 93/100 (3) "..thought (3) but really hadn't had a look at (4)") there were two types of errors:-

(a) Failure to understand the nature of the task - or more specifically the overlooking of the implications of the words "tells best". As in the previous item many children changed to the correct answer when they were asked to check over again with the emphasis on "tells best" - e.g.:-

B. 110/113 (2) - when asked to describe why he rejected the other four gave a correct justification for (1), (3) and (5), hesitated on (4) and then changed to (4) "...because it really does tell best."

G. 112/114 (3) - "...because it describes what it does and how it works," although when asked if that "tells best" what a horse is, she changed to (4).

B. 97/104 (5) - "...because it's useful knowledge to people in case they had a wagon," but when asked to reread the question aloud, he changed to (4) and explained that he had not been thinking of "best".

(b) Rejection of the correct answer on "reasonable" ground e.g.:-

G. 117/118 (5) - when asked to check over she still preferred (5) and argued that there are many four legged animals, while there are only a few things which pull wagons - horses in-
cluded - so that a person is more likely to get the wrong idea from (4) than (5). This argument was put forward by others - i.e. (4) could be an elephant, a cow, a giraffe etc. When it was pointed out to G. 90/99 (5) that perhaps a tractor could pull a wagon and that is very different from a horse, hence it may not be the answer, she replied - "...but in the olden days of wagons horses mostly pulled them."

Errors of this type may be due to a more acute analysis of the problem than that made by many of those giving the correct answer.

Item 31. Which tells best just what a lamb is?
(1) an animal with wool, (2) a creature with four legs and a tail, (3) a lively small animal, (4) a young sheep, (5) a young animal that eats grass.

N=59. (1) 42, (2) 12, (3) 32, (5) 14.

The errors were of the same type as those described for the previous item - i.e:-

(a) Failing to read all the choices: - e.g:-
B. 89/93 (1) - when asked to read the item aloud, he discovered that he had not noticed the second line of choices (4 and 5).

G. 72/76 (1) - on checking, it was found that she had not read past (1).

(b) Failure to understand the nature of the task, e.g:-
B. 133/132 (4) and (2) - could not decide between these two, because he regarded lively as an important characteristic of a lamb not covered by (4) - i.e. he was thinking in terms of description.

B. 102/113 (2) - chose (2) rather than (4) "...because (2) describes more about it - what it has."

B. 97/104 (1) - it was clear from the questioning that he was looking for an answer which contained the most 'distinctive' characteristic for purposes of description/
(c) Rejection of the correct answer on reasonable grounds, e.g.:

G. 117/121 (1) - when checked against (4), she retained (1) because "...people might not know what a sheep is." Similarly B. 122/127 (4) who gave the correct answer, qualified it by saying "...but only if you know already what a sheep is."

B. 99/105 (2) - when asked to comment on (4), rejected it "...because if you didn't know what a sheep was you wouldn't know anything"; similarly B. 101/109 (3) - "...because if they didn't know what a sheep was how would that tell them about a lamb,"- and G. 93/89 (2) who gave the same type of answer, but when asked whether it would make any difference if the person already knew what a sheep was, answered that her choice would then be (4).

Item 35.
Which tells best just what a lie is?
(1) a mistake, (2) a malicious false statement, (3) an accidental false statement, (4) an exaggeration, (5) a wrong answer.

N=104. (1) 8, (3) 9, (4) 26, (5) 57.

Errors on (1) and (3) were almost completely confined to groups 5 to 7, while girls (41% of total) made 51% of the errors, however there were no clear sex difference in the types of errors made.

(a) Giving the correct answer without fully understanding it - i.e. of the 25 children who reworked the item, 10 gave (2) as the answer, but had no idea of the meaning of "malicious". One boy - B. 91/96 (2) chose (2), because he did not know "malicious" and therefore thought it was likely to be the correct answer. These ten children ranged from B. 126/116 to B. 76/80, and chose the right answer in spite of the unknown word.
(b) On the other hand, another 10 (ranging from B. 117/118 – B. 83/91) rejected (2) and chose another, because of the unknown word "malicious", e.g.:

B. 115/110 (4) – chose (4) "...because usually when people tell a lie they exaggerate" and further questioning established the fact that he had considered (2), rejected it because he did not know "malicious" looked at the other choices and decided that (4) was the nearest in meaning.

G. 101/100 (4) – rejected (2) because of "malicious" although she saw that "false" was correct, and decided on (4) "...because an exaggeration makes something sound much better than it really is". When asked to describe "lie" in her own words, she said "...a person says something which is not the truth – something that isn't right – a fib....false means not the truth".

G. 104/108 (5) described "lie" as "...someone made it up, it's not the truth". When asked if there was a word for this she gave "false", and when asked about (2) said that the "false" was right, but she could not use it because of "malicious".

(c) Other errors may be illustrated by these two examples:–

B. 82/83 (3) – could not pronounce malicious, accidental, or false, but described lie as – "a person is not telling the truth".

G. 83/83 (3) – chose (3) "...because you don't mean to say it," and when asked to comment on (2) said – "well, telling a lie is not marvellous"!

These examples show how one word can completely dominate the structure of the item in such a way that the answers do not really reflect the underlying reasoning - e.g. it is reasonable to assume that the 10 who rejected (2) because of "malicious" were really being more discriminating that those who accepted (2) not knowing what it meant. It is probable that a large proportion of those who gave the correct answer in the total sample, did so by suppressing the unfamiliar
I1 "malicious".

**Item 52.**

What is the most important reason that bright lights are placed in front of theatres?

(1) so that people can see where they are, (2) to attract attention and look inviting, (3) so that people can see the advertisements better, (4) electricity is furnished to theatres cheaply, (5) to help light up the street.

N=99. (1) 34, (3) 42, (4) 3, (5) 31.

The distribution of errors showed that girls made proportionately more errors on (1) and (5), boys on (3), the errors being well distributed over all I.Q. groups. This difference was not reflected in the types of errors produced on reworking.

(a) A number of children (7 out of 20) were not clear whether "in front of" referred to the inside (i.e. around or on the screen) or the outside of the theatre (universally interpreted as "picture" theatre), e.g:-

G. 130/125 (2) - the correct answer, but when her reasons were checked it was found that she was thinking of the "inside" lights.

B. 121/125 (2) - gave as his reason for rejecting (3) -"You wouldn't see the advertisements on the screen if bright lights were placed around the front." Similarly B. 103/99 (2) - but when he was told that it referred to the outside, he changed his answer to (3)!

Some of those who confused inside and outside gave the wrong response because of this - e.g:-

G. 100/104 (1) "...because you need lights in the theatre to see where you are going"; also G. 93/100 (1) asked "does it mean inside or outside the theatre?" - when told "outside" she changed to (5) and then to (2), "...because it is the most important".
(b) A strong reason for rejecting (2) and (3) was provided by those who argued that people would normally look in the papers first and decide where to go, e.g.:

G. 116/120 (1) or (5) - when checking (2) said "...could be (2), but don't think it would be somehow - if people want to come to the pictures, they would look it up in the paper, they wouldn't go to a picture just because of the lights outside" - this is a reasonable assumption.

(c) Preference for (3) over (2) was argued reasonably - e.g.:

G. 105/105 (3) - when checking (2) she said - "...it must be (3) because if it attracts attention it attracts it to the advertisements," - and in fact many of the bright lights are connected with advertisements.

(d) Some thought of "bright lights" as lights in front of theatres on lamp posts. e.g.:

G. 113/125 (1) had this interpretation in mind when she said - "...people outside are interested in where they are going in the street".

B. 108/110 (1) also interpreted in the same way and said "...you might bang into someone when coming out."

G. 106/108 could not give an answer until she had asked - "...is it the street lights or the one on top of the theatre(entrance)?" - then gave (2) - her original error, (1), being due to the former interpretation.

(e) Other sources of error are illustrated in the following examples:-

G. 105/101 (4) - when questioned it was clear that she was just giving a reason, not the most important, which in fact she considered to be (3). A similar process was observed for B. 98/96 (5); B. 94/102 (5) - who gave (2) when his attention was drawn to "most important"; G. 93/100 (1) "I was just thinking of a reason, not the most important" - i.e. all cases of incomplete formulation of the task.

B. 102/100 (3) argued that it couldn't be (2) because the lights mightn't be bright enough to attract attention (he
was thinking of the small advertisements round the entrance) when he was asked to read the sentence again he saw "bright" and said he hadn't noticed it before. Also B. 87/94 (3) who, when taken over each reason, said that he had not read (2) at all before - similarly G. 83/83 (3) when checking over other reasons said that she hadn't really read (2) because she thought it was just part of (1), - i.e. all cases of failure to read all parts of the question.

This item has been treated at some length, because of the variety of interpretations which gave rise to wrong responses - many of these interpretations being quite reasonable.

**Item 66.**
Which tells best just what a foot is?
(1) to wear a shoe and stocking on, (2) the part of the body on which an animal stands, (3) it has five toes and a heel, (4) both feet are the same size, (5) men have larger feet than women.

\[ N=56. \ (1) \ 14, \ (3) \ 82, \ (4) \ 0, \ (5) \ 4. \]

The very high percentage of errors on (3) along with the wording of that particular distractor, indicates that most of the errors sprang from a definition by description being preferred to a definition by function. In other words the child must judge what "tells best" implies - and from the reports of the ways in which (3) was chosen, it is clear that most of the children actually reformulated the question into "which one describes a foot best?". Even so, there were other factors operating in the choice, e.g:-

(a) By far the most powerful factor was the confusion caused by the word "animal" in (2) - e.g:-
21 out of 28 children who reworked the question rejected (2) because "animal" was interpreted as "four legged animal" in contrast to "man". Typical comments were - "(2) would be alright if you were talking about an animal's foot... an animal's foot is a paw usually... an animal doesn't have a foot... animals and humans... an animal's foot is nothing like ours." A more sophisticated argument was put forward by B. 129/142 (3) who said that if it had been "mammal" instead of "animal" he would have picked (2). At the other extreme was the comment by G. 90/97 (3) who when asked to check (2) said "Well, an animal doesn't usually stand on your foot".

(b) Another reason for preferring (3) to (2) is illustrated in the following arguments, e.g.:

G. 115/116 (3) - "...it (i.e. 2) could refer to a 'leg' because that is what you really stand on"; B. 100/112 (3) "...well it (a foot) isn't part of the body, it is part of the leg" - i.e. he used body in the sense of "trunk"; G. 106/108 (3) "...a foot is for you to walk on, not just stand on"; B. 129/142 (3) also rejected (2) partly because of the possibility that it could refer to "leg".

(D) Category (4) - Forming a Concept From Given Examples.

In this category the five items are of the type in which a group of words must be classified in terms of underlying similarity, and in four of them the classification is tested by choosing another word which can be classified on the same basis.

**Item 1.**

Which one of the five things below does not belong with the others?

(1) peach, (2) banana, (3) orange, (4) baseball, (5) plum.

N=7. (1) 0 cases, (2) 2 cases, (3) 3 cases, (5) 2 cases.

There were no errors made in the cases where original errors were reworked, although of the seven original errors, three were made by children in groups 2 and 3 - presumably
"warming up" errors.

Item 9.

Which one of the five things below is most like these three: apple, peach, pear?
(1) seed, (2) tree, (3) plum, (4) juice, (5) peel.

N=34. (1) 18, (2) 29, (4) 38, (5) 15.

Errors arose from failure to understand what was required - i.e. the children looked for a word which described a characteristic common to all three given words, whereas the correct answer depended upon a further step - namely picking out another word which illustrated the common characteristic.

Typical examples were:

B. 102/100 (1) or (4) - when asked to explain his choice he changed to (2) and decided that that was correct "...because they all come off a tree". When the instructions were read to him - with emphasis on "most like" - he was still quite satisfied with (2) and could not appreciate the character of the actual question asked.

B. 98/96 (4) also appeared to be satisfied, although when the instructions were read with emphasis he saw the point, changed to (3) and said "I was thinking of something that these three have".

G. 86/85 (5) - chose (5) "...because all have skins...but this would go better with orange and there is no orange...perhaps seed". It was pointed out to her that one word could describe all three, and she immediately said "fruit...then tree goes with fruit". When asked if any of the words fitted in with "fruit" she said "plum...but that would not...yes it might...yes it is the best one". It was clear that she only gradually began to see the nature of the task, although she knew all the essentials quite well.

G. 72/76 (3) - i.e. the correct answer, but her reason for choosing this was "...because it begins with P"! Despite this she gave "fruit" as a word covering the three words given, and having said this chose "plum".

Although these children gave wrong answers, the errors
did not spring from an inability to see the various relationships involved, but from a failure to clarify the nature of the actual task.

**Item 25.**
Which one of the five things below is most like these three: snake, cow, sparrow?
(1) tree, (2) doll, (3) pig, (4) feather, (5) skin.

N=68. (1) 3, (2) 3, (4) 25, (5) 69.

Errors were of the same general type as those on item 9 - and although the children gave wrong answers, they understood the relationships between the three given words, animals and pig. Two examples illustrate this:-

B. 100/105 (4) "...because a sparrow has feathers." When asked to read aloud and to check whether "feather" was most like "these three" he said "...no - pig, because a pig has legs like a cow". The conversation then followed this course:- (E) - "What about snake?" - (B) "...a snake has a skin and a cow has a skin." (E) - "What is there the same about snake, cow and sparrow?" - (B) "The three have heads and all got tails" - (E) "Can you give one word which describes them?" - (B) "Tailed animals...so one word to describe them would be 'tailers'" (E) "Is that a good word to use for them" - (B) "no...animals...so pig"! It can be seen that this boy resisted strongly attempts to break down the "common characteristic" classification, in favour of an "essential similarity" classification.

B. 82/83 (5) showed a more spontaneous recovery from his initial "wrong direction" - i.e. he chose (5) "...because they all have skins" and when asked to check over the others answered,"...a snake is a long wriggly thing, a sparrow is a bird and a cow...and a pig is an animal, so "pig" because them three all living things." When asked why he chose (5) at first he said "...well, I though it was something to do with them three."

**Item 41**
Which one of the five things below is most like these three: president, admiral, general?
(1) ship, (2) army, (3) king, (4) republic, (5) soldier.

N=141. (1) 8, (2) 13, (4) 18, (5) 61.
This item was a confusing one for many children and illustrates three major types of error.-

(a) Failure to understand the requirements of the task, as in the two previous items, e.g:--

B. 124/123 (5) - "...all men in the army so (5)." When asked to check the question again he said "...oh...it's a president isn't it, and oh...an admiral is in a ship isn't he, so it must be(three) because they are all in high positions." His first reading had been quite inadequate, although he quickly corrected the defects on rereading.

B. 115/117 (4) - "...because just those three words made me think of it." When asked what there is about the words which makes them alike, he looked over the question again and chose king, for the correct reasons.

B. 102/100 (4) - "...because a republic is like a country and they might all belong to it." Asked what he was trying to find out, he answered - "Well...something they all belong to." He was then directed to look at the question, in particular at "most like" and changed his former approach to - "...they are all high up in something, all rule over people...so king is the answer."

B. 101/109 (5) explained his choice thus - "President man is fairly well up and would be a soldier, and admiral is in connection with army, airforce and navy, and general is in army, so the answer is soldier." When asked what he thought of (3) he said - "Well he is not allowed to go to war and a king is something like a president, but he doesn't have to be high in rank but is born a king." At this stage he was asked what there is about the three which makes them alike, and his answer was - "...they're all important men who rule...so king will be the right answer I suppose." In view of his "I suppose" he was asked whether he still preferred "soldier", and said "no, I think it would be king now if you look over it properly".

(b) Ignoring a disturbing element - in this case the meaning of "president" - e.g:--

G. 117/121 (5) - "...because I thought that in the army there would be admiral, general and soldiers so chose (5)" - from questioning it was clear that she knew that a president is not in the army, so just ignored "president".
B. 116/116 (5) — "...because all have something to do with commanding soldiers in war." A check on "president" showed that he did not associate president with commanding in war so just ignored the word.

G. 116/120 (5) — "...because admiral and general belong to the army, they couldn't belong to (1), (3) and (4), and the army is a lot of people, so only (5) left." When questioned about "president" she said that he didn't belong to the army (i.e. she had just dropped the word because it did not fit in), and then suddenly changed her answer to (3) "...because they are all in charge of something."

G. 118/115 (5) — "...because they are all in the army and a soldier is in the army too." When asked if a president was in the army she replied "no"!

B. 91/96 (5) — "...because admiral is a soldier, president is one and a general is one." Questioning showed that he knew a president was not a soldier, and that he knew they were all "high up". When asked to read the question again he gave (3) as his answer.

G. 86/88 (5) — "...because the three are all soldiers." When asked if there was anything else about them she said — "...all something to do with near the top", and when directed to read the question again, gave (3) as her answer.

(c) Giving an answer to a different task — i.e. classification in order — perhaps as a result of the four preceding items in which that was the task, e.g.:

G. 107/112 (5) explained that she decided they all went down in rank so "soldier" would be the next one. When asked to read the question again and work it aloud, she said — "...same job to do, looking after a committee, looking after a ship, looking after an army...so "king"...I'd just taken a quick look and thought they all belonged to the army."

G. 106/104 (5) — "...well first there is president, then admiral, then general, then soldier." — i.e. in order of rank. When asked to read the question again she hesitated for a minute then said "...oh...king, because he is high up like the others."

(d) There were a few isolated errors, such as not knowing the words at all, e.g. — G. 112/114 (1) who thought that all the
words had something to do with a ship; failing to find any significant relation e.g.- B. 76/80 (2), (4) and (5), who ignored the word "one" in the instructions, and when his attention was drawn to it could only find one basis of similarity - i.e. "all end in al" - despite the fact that he agreed that "president" did not; misreading a word e.g.-G. 100/104 (4) "...because they are all "public"." In two cases the right answer was chosen for the wrong reason - e.g.-

G. 90/97 (3) - who didn't know anything about admiral or president but "thought that king would go with general."

G. 93/99 (3) - "...because a king belongs to the country and tells them what to do, and these three belong to the king."

**Item 58**

Which one of the five things below is most like these three: cannon ball, wire, penny?
(1) pound note, (2) bone, (3) string, (4) pencil (5) Key.

N=66. (1) 52, (2) 18, (3) 27, (4) 3.

Three types of errors accounted for the majority of wrong responses:-

(a) Failure to clarify the requirements of the question, E.g.-

G. 110/103 (2) when asked her reason for this choice read the question again and gave (5) as her answer "...because they are all metal - I hadn't really looked properly before."

B. 101/109 (1) asked "...does it have to be one for the whole three or something separate for each one?" - i.e. he had not really read the question and when asked to read it again realized that his question was unnecessary.

(b) A related type was failure to use all the relevant parts of the data - e.g:-

G. 97/105 (3) - "...because string is like wire...long".
When asked how this fitted in with "cannon ball" she looked back at the question and said - "...no, a cannon ball is round, so can't be (3), but do they have the same shape...no...would it be that they are made of...yes it would be a "key" because they're kind of metally." This girl produced the right answer after a good analysis of the data, yet originally she had not even taken into account two of the three words given, but just ignored part of the data.

B. 95/96 (3) - "...because it's like wire..."-when asked to read the question again, he saw that there were two other words to be dealt with, and to the question "Is there anything the same about these three?" replied immediately "copper" - i.e. he had not answered the question fully but at least was capable of a more discriminating approach than his first partial analysis of the data.

G. 90/97 (3) - "...because it's like wire." When asked if that fitted in with the other two she realized that she had ignored them.

(C) The most interesting type was the result of a "false direction" - i.e. once the children had read the question and thought of a possible basis for similarity, they struggled to fit the data into this interpretation and had difficulty in changing to any other basis - e.g:-

G. 130/125 (3) - "...because a cannon ball is round, a penny is round, and wire looking at it from the end is round." When asked if there was anything else similar about them, she said - "...they are all made of copper...so could be a "key" - but there are all shapes of key." Although she decided to choose (5) because all were made of metal, she was still trying to fit in the concept of "shape" with "key".

B. 110/113 (2) - "...because they are all solid things, for instance a pound note isn't solid and snaps easily, but a bone needs force behind it to break it." At this point he became tangled with the possibility that a key was also solid and difficult to break, but decided that a bone would be bigger and harder. He was then directed to start again, forget about his "bone" and see if there was anything else similar about the three. Almost immediately after rerading the question he chose key because they were all metal, i.e. when the "false direction" was broken for him he went straight to a new, correct "direction".
G. 91/96 (3) - "...well you could put those all on a string."
To the question - "What about a cannon ball?" she replied,
"...no, couldn't put a cannon ball on...so choose (5)...because
you can put wire on a key and put penny on to a wire..."
i.e. she ignored the "cannon ball" difficulty and persisted
with her concept of "putting things on" with wire etc.

(d) Other types of response are illustrated in examples of
"giving up" without attempting, and giving the right answer
with inadequate reasons - e.g.:

G. 105/105 - "...they aren't alike..." - however, when question-
ed she produced several possible (although not adequate)
bases for comparison.

B. 98/96 - "I can't do that," but when asked to try made a
fair attempt to find a basis for comparison.

G. 119/119 (5) - i.e. correct answer - "...a key is round and
long, a piece of wire is round and long, and a penny is round."
When her attention was drawn to "cannon ball" - which she had
completely ignored - she added with satisfaction - "and that
is round too - so key is right."

G. 109/112 (5) - "...they are all made of wire."

B. 110/118 (5) - "...because key has sometimes copper in it
and sometimes wire in it."

G. 105/101 (2) - "...no (5), because a key is harder than a
bone."

G. 111/106 (5) "...they all look the same brownish silvery
colour."

(E) Category (5) - Mixed Sentences.

As was pointed out in chapter (IV) there were more
omissions in this category than on any other (42% of the total
omissions in Otis I were on category (5)), so that many child-
ren in the group of 84, were working on these items for the
first time.
Item 12.
If the words below were rearranged to make a good sentence, with what letter would the last word of the sentence begin? (Make the letter like a printed capital).

usually are of made tables wood  

(Ans. = W)

N=60. (D) 50, (T) 18, (S) 8, (U) 7, (W) 5
(Wrote sentence) 2, (Others) 10, Omissions 75.

Many children grimaced or said that they could not do this sort of question, however, when asked to try, most of them were able to rearrange the words correctly.

(a) Correct rearrangement of the sentence, but wrong letter in the brackets. Out of the 26 children who reworked the item, and gave a wrong answer, 21 rearranged the sentence correctly, although there were differences in the time taken and the number of attempts. In these 21 cases the error arose from a misreading or confusion of the actual instructions, and in many cases this confusion persisted despite opportunities to correct it; e.g.:-

B. 116/116 (T) "...because it asks for the first letter of the first word."

G. 118/115 (D) - she hesitated over (W) and (D), finally chose (D) and was quite satisfied until asked if she was quite sure that it was (D) - similarly B. 115/117.

G. 111/106 - wrote the rearranged sentence along the dotted line, and when directed to read the question again could not see anything wrong with her answer. She did not get (W) until the instructions were read for her, slowly and with emphasis.

G. 100/109 (Wood), was asked to read aloud - still (Wood), so directed to read carefully and changed to (T). Only when the instructions were read for her did she give the correct answer. Similarly B. 83/91, G. 97/98.
B. 95/103 (T) - after rereading instructions changed to (D) - but when instructions were read for him, gave correct answer. Similarly, G. 93/100, G. 83/83, and G. 104/108 who did not give (W) until asked - "Are you sure that you have followed the instructions?"

B. 91/96 (Wooden tables), was asked to make sure what he should write in the brackets, so changed to writing the whole sentence. When the instructions were read for him he gave (W).

B. 90/92 provides a good example of the hazards involved in completing this item. The record of his reworking runs:- "Unusually made tables are of wood;" (- attention drawn to "unusually"), "usually made tables are of wood...no..usually wood are of made tables;" (-directed to try starting with another word), "tables usually are made of wood...tables are made of usually wood...no...you can't say 'usually wood'... tables are usually made of wood." After this concentrated attempt he passed on to the final stage of writing in his answer and wrote the arranged sentence on the dotted line. When directed to read the instructions again he gave (Wood) as his answer, and finally gave (W) after a third reading of the instructions!

G. 93/89 (u) - she gave an incorrect response despite the fact that she had given the correct order of rearranged words. After rereading the instructions she gave (Y) - i.e. the last letter of the original first word. It appeared that she dissociated the instructions from the completed rearrangement, took the first (instead of last) word of the original sentence, gave the first letter in it, changed to the last letter, and finally when the instructions were read to her, with emphasis, saw that they referred to the rearranged sentence, and gave (D)!

It might be thought that the persistent rereadings and misinterpretations arose because the children became excited or worried - but in fact there was no evidence for this, nor was there any suggestion of pressure or hurry, and for the most part the answers appeared to be given with confidence.

(b) Several other types of response worth recording are illust-
rated below - e.g.:-

B. 89/93, did not seem to be progressing very well and when asked what he was trying to do, revealed that he was trying to join the disarranged sentence with the sentence in the brackets- (Make the letter ... etc.) in order to make one good sentence out of them both.

G. 100/95 gave as her answer - "(no)" and it transpired that she had read the instructions as: - "Would the last word of the sentence begin the sentence?"

G. 90/97 gave the correct answer, although her rearranged sentence was "Tables are usually wood" - and she resisted any suggestion that perhaps she had not used all the words.

B. 77/79 also gave the correct answer, because although he did not bother to rearrange the sentence, he did follow the other instructions, and it so happens that the first letter of the last word in the original disarranged sentence is (W)!

Item 28.
Do what this mixed up sentence tells you to do.
letter Write the the in A brackets
Ans. = A.

N=13. (the) 4 cases, (28) 2 cases, (W) 2 cases,
(others) 5 cases. Omissions 52 cases.

Originally there were only 13 errors on this item, while there were 52 omissions, although as in the previous item, children who had originally omitted this item were asked to attempt it during the interview. Whereas in the previous item the majority of the children had little difficulty in rearranging the words, in this item many failed to make sense of the words at all.

(a) Distortion of the task by a previous related task - i.e. some children apparently had in mind the instructions for item 12, two of the three examples given here are from children who had not worked item 12 prior to item 28, during the
interview.

G. 110/113 (W) "...because it begins the sentence." When asked to reread, she gave the correct answer.

B. 101/109 (Write) - "...because the sentence tells you to put what word is in front in the brackets."

G. 72/76 had previously worked item 12 correctly, and her answer for 28 was (B) "...because it is the first letter of the last word."

(b) Difficulty in "recentering" - i.e. a word was interpreted in a particular context, and this context ruled out the alternative and correct interpretation - e.g.:

B. 97/104 - struggled with the sentence for some time, changing "A" around into every possible position, because he could only think of it as a preposition and not as a letter. Exactly the same difficulty hindered B. 103/99 - although when it was suggested to him that "A" could be a letter, as in A.B.C, he gave the correct answer immediately.

B. 90/92 had difficulty because he interpreted "bracket" as a "group" - i.e. "write the letter in the brackets... As the first bracket...A.B.C.D," - and having interpreted bracket in this way, he tried various ways of getting a bracket of letters.

B. 89/93 - interpreted "letter" in the sense of a written communication, and could not make any sense of the sentence. When a hint was dropped that A.B.C.D etc. are letters, he rearranged the sentence and gave the correct answer.

(c) Confusion in following the instructions, E.g.:

B. 102/100 - arranged the sentence correctly but could not proceed any further because he was looking for further directions, having concentrated on the "mixed-up" in the instructions and ignored the "Do what...tells you to do."

B. 76/80 (the) - i.e. he had just done what the mixed up version of the sentence suggested, and had overlooked the fact that he must first "unmix" it.

(d) Failure to read or use all the data - e.g.:

G. 83/83 could not make sense of the sentence, because she was
only reading it from "Write..." and had not noticed the word "letter" in front of "Write". Several children in attempting to arrange the words dropped out one or more words and made no attempt to check up and see that they had used every word. This was the case with G. 97/105 (A) who gave the correct answer, but had rearranged the sentence as "Write in brackets the letter A" and was quite satisfied that this was the best rearrangement.

**Item 32.**

If the words below were rearranged to make a good sentence, with what letter would the third word of the sentence begin? (Make the letter like a printed capital).

- honey bees clover gather red from (Ans. = H).

N=116. (G) 47, (R) 14, (C) 9, (F) 8, (B) 7

Others) 15. Omissions 56 cases.

Types of errors noted were:-

(a) Correct rearrangement but incorrect response, i.e. as in item (2) the children arranged the words correctly and then misread or forgot the original instructions - e.g:-

B. 117/118 (C) - he had the correct order but then wrote in (C) "...because I was thinking of what I was saying before when I was still arranging the words "bees gather clover..."

G. 106/104 (R) - correct order, but then said that she gave (R) "...because red is the third word" - although she realized her mistake as she explained.

G. 110/103 (F) - arranged the sentence in an acceptable, although incorrect order - i.e. "From red clover bees gather honey" then explained (F) as "the first letter of the word beginning the sentence" - even after rereading the instructions, she still gave (F) as her answer.

B. 102/100 (honey) - correct order, and even after rereading, he still gave the whole word - did not change to (H) until the instructions were read aloud.

B. 83/91 (E) - correct order but chose the third letter of first word.

B. 83/98 (B) - correct order but chose first letter of the first word.
(b) Difficulty in recentering - i.e. the two words "honey bees" formed a very strong "gestalt," and many children puzzled over the sentence attempting to start it with "honey-bees," or to fit the words in elsewhere. Examples range from G. 114/117 to G. 84/88. In some cases the suggestion was made that they separate the two words, and this hint was usually enough to lead to a speedy solution - e.g.:-

B. 108/110 attempted the sentence as - "honey bees gather red clover from...honey bees gather from red clover...from red clover honey bees..." At this stage the suggestion was made that he break up the two words - and he gave the correct answer straight away.

(c) Failure due to satisfaction with an incomplete or poor solution - e.g.:-

B. 99/105 (G) - i.e. "honey bees gather red clover from..." - he realized that was not a "good sentence" but was content to leave it incomplete.

B 111/102 (B) - i.e. "from clover bees gather red honey" - this boy knew that the sentence was not a good one and said there was no such thing as "red honey" - yet was happy to leave it as it was.

G. 105/101 (G) i.e. "honey bees gather red clover" - although she realized that there was a "from" in the original words, she could not fit it in so dropped it.

(d) Dissatisfaction with the correct order - e.g.:-

G. 109/112 put the words in the correct order, but hesitated to give any answer, because "clover is usually green and 'red clover' doesn't sound very nice."

G. 107/112 was also unable to answer because although she had the words in the correct order, she argued that "bees don't gather honey - they gather pollen and then go back to the hive and make honey" - a difficult objection to refute!
(e) Other errors arose from complete failure to make any sense out of the words, but also from confusion about what was really required, or from the failure to restrict the sentence to the words given. Examples of these two latter types are:-

B. 97/104 (red) "...because it tells you what kind of clover ...perhaps it is really"bees" because it is the main part of the sentence." This boy arranged the words correctly but completely misinterpreted the task. When asked to read through the instructions he said "...oh - (H) - I hadn't thought what the instructions really meant."

G. 96/96 (F) - "honey is gathered from red clover...so the third word begins with (F)." When asked to look at the actual words given she gave (H) as her answer (i.e. correct answer) - and her explanation was "honey is gathered from red clover...so the first letter is (H)"! Another mis-reading was made by B. 82/83 who made a reasonable sentence by changing "from" to "for".

Item 49.
Do what this mixed up sentence tells you to do. Sentence the letter Write first this in  (Ans. = W)

N=15. (F) 3 cases, (l) 2 cases, (T) 2 cases, (D) 2 cases, (Others)6 cases, Omissions 83 cases.

Half the errors arose from inability to make a sensible sentence, but over and above these, certain types of errors were noted.

(a) Correct arrangement but confusion as to the requirements of the task, e.g:-

G. 112/114 - after some difficulty she got the correct order, but could not make any further move because she could not see that the sentence was telling her to do something - despite a rereading of the instructions.

B. 97/104 - got correct arrangement but then did not know what to do - "because it doesn't tell you what letter to
write" - i.e. he apparently expected a letter to be given in the sentence, as in item (28).

G. 72/76 (S) - right order, but then gave the first letter of the last word.

(b) Failure to use all the words, e.g.:-

B. 103/99 - "Write this letter in the sentence" - even when asked specifically to check up, he still could not see that he had missed out any words.

G. 97/98 - tried to make a sentence out of "Write first this in" because she completely overlooked the words before "Write" - even on a rereading of the question.

(c) Misinterpretation of "letter" (as in item (28)) - e.g.:-

B. 91/96 - attempted to arrange the sentence in various ways, and when asked what "letter" meant he said he was thinking of a letter to post.

G. 86/88 - arranged the word - "Write this sentence first in the letter." When asked what this meant she answered - "...well if someone is writing a letter and you tell them to write this sentence first in the letter...so the answer is to write the sentence down." A similar answer was given by G. 96/96.

(d) Ignoring the capital "W" - e.g.:-

G. 100/109 had great difficulty in starting, trying to open the sentence with 'this' and 'sentence' - she did not see the capital W until it was specifically pointed out.

G. 90/99 (I) - rearranged the sentence as "In this sentence write the first letter" - a reasonable arrangement if the capital W is ignored.

Item 53
If the words below were rearranged to make a good sentence, with what letter would the third word of the sentence begin? (Make the letter like a printed capital).
boys birch the a canoe made bark.                  (Ans. = M)

N=45.  (T) 31, (B) 29, (C) 16, (A) 13, (Others) 11, Omissions 51 cases.

The great stumbling block for most children who made
errors on this item, was the necessity to link "birch" and "bark", and many children could not achieve this. There were however other types of errors:-

(a) Objection to "birch-bark," e.g.:-

G. 117/121 had the correct arrangement but would not proceed because she thought "birch bark canoe" sounded wrong and awkward.

G. 118/115 - also objected to the combination as being "un-common" - similar objections were made by four other children.

(b) Correct arrangement but wrong response - e.g.:-

B. 124/123 (B) - right order but gave (B) because he was thinking of a stage earlier in his rearranging, when he was trying to begin the sentence with "The boys birch bark..."

G. 105/105 (A) - right order but gave (A) "...because it is the middle word of the sentence." When asked to read the instructions again she gave (D) as her answer - i.e. the third letter of the third word.

B. 98/96 - after much difficulty in arranging the sentence correctly, he finally gave (a) as his answer - "...because boys made a etc..." - i.e. he forgot to take account of "The", although he had included it in his rearranged sentence.

B. 83/91 - gave correct sentence after a struggle, then gave (D) - i.e. the third letter of the third word.

(c) Using extra words - i.e. going beyond the data given - e.g.:-

B. 111/102 (C) - because he arranged the sentence as - "the boys canoe was made of birch bark." When asked to check through the given words, he could find nothing wrong. Only on another rereading of the item did he see what was wrong.

G. 100/95 (C) - i.e. "the boys canoe was made of bark" - the addition "was" passed unnoticed until she was asked specifically to check the words.

(d) Correct answer - inadequate solution - e.g.:-

G. 108/114 (M) - had only rearranged part of the sentence - i.e. "The boys made..." and had gone no further.
G. 93/100 (M) - i.e. "The boys made a birch canoe" - she could not fit "bark" in, so ignored it.

G. 90/96 (Made) i.e. "The boys made a canoe" - she ignored "birch bark" completely - the same situation arose with B. 87/94 (M).

Item 60.

Do what this mixed up sentence tells you to do.

sum four Write three the one and of (Ans. = 8)

N=21. (7) 9 cases, (4) 5 cases, (5) 3 cases (1) 2 cases, (Others) 2 cases. Omissions 61 cases.

The same types of errors appeared on this item as on the other five items - i.e:-

(a) Correct arrangement but wrong response - e.g:-

B. 124/123 - no trouble in arranging the words, but then wanted to write the whole sentence out on the dotted line.

G. 117/125 - had some difficulty in rearranging the words, but when successful gave (3, 1 and 4) as her answer.

B. 91/96 - after finally rearranging the words he gave (W) as his answer "because it begins the sentence" (Cf. item 49).

B. 108/115 - after an initial mistake, which was pointed out to him, gave the correct arrangement, and the answer - (S) - i.e. the first letter of the third word (Cf. item 53).

(b) Misinterpretation of a crucial word - i.e. "sum". Of the 28 errors repeated on reworking, 18 were mainly due to failure to make sense of the words, because "sum" was interpreted as "a problem in arithmetic." The children who were unable to shift this word from this particular context ranged from B. 126/128 to B 83/91. They twisted the sentence in all sorts of ways - yet when a hint was dropped such as "Is there another meaning of "sum" in arithmetic?", many of them solved the problem immediately (with all the outward signs of sudden
insight). In some cases an answer was attempted on the basis of a rearrangement where "sum" was used in the "problem" sense - e.g.:

B. 110/118 (4) - "Write 3 and 1 of the sum 4, that is write down 3+1 makes up a sum, 4, so put four in the brackets" - i.e. he was trying to fit in the idea of "3+1=4 is a sum in arithmetic".

G. 91/96 - gave the correct arrangement and wrote (3, 4 and 1) as her answer - "...because it says to write the sum, so write it in the brackets."

(c) Ignoring elements in the problem - e.g.:

G. 115/117 (7) - ignored the "one" in the sentence.

G. 112/114 (4) - ignored the "four" and could not see anything wrong until asked to check each word.

B. 100/112 (12) - not only ignored the "one", but multiplied "three and four". When he realized what he had done he changed to (7), and could not see that he had missed out anything even after rechecking. Finally the words had to be counted and ticked off one by one before he saw what had happened.

B. 95/96 (7) missed out "one" and did not notice it until he had recounted and checked the words four times.

(d) Individual errors of various types occurred - e.g.:

B. 126/116 (one) - "...because there is no "two" to make the line of one, two, three, four - so write "one"" - he was attempting to apply some principle of serial classification, or perhaps the "analogies" pattern, and had not really clarified the problem.

G. 110/103 - could not arrange the sentence because the words "Write three" were accepted as a unit, and no matter how she tried to arrange the sentence, she could not break up the self imposed limitation.
(F) Category (6) - Classifying and Arranging Elements of a Concept in a Given Order.

Items in this category are of the "largest, first, arrange in order" variety.

Item 6.
Which one of the five things below is the largest?
(1) bird, (2) branch, (3) tree, (4) twig, (5) limb.

N=24. (1) 8, (2) 58, (4) 0, (5) 33.

Of the original 14 errors on (2), 5 were made by children in groups 2 and 3, however, as there was only one error on reworking (a group 7 boy), there is no indication of why these children made errors on what appears to be an elementary question. The one error, on reworking, shows a failure to follow the instructions, and a misreading:-

B. 89/93 (2, 1 and 4) - had the idea he had to find things which grow on a tree. When directed to read the instructions again he chose (5) - i.e. "lump - a lump of trees" (confusion with clump and misreading of "limb").

Item 10.
Which one of the ten numbers below is the largest? (Tell by letter).
A6456, B8968, C4265, D5061, E4108, F7549, G2335, H9472, J3286, K8970.

N=47. (K) 81, (B) 9, (G) 4, (10) 4, (A) 2. (Ans. = H)

(a) Failure to understand the requirements of the task - i.e. ignoring the numbers and presuming that the task was to give the letter highest in the alphabet (K). Of the 17 errors repeated, 8 were of this type, ranging from G. 122/127 to G. 90/97, e.g:-
G. 116/120 (K) "...because it says 'tell by letter', so I took the largest letter" - she had seen that H was the largest number, but presumed that for some reason this was to be ignored.

G. 112/114 (K) - looked for the largest letter because of "tell by letter" and when asked about the first part of the instructions, she said - "well some of the numbers are larger ...but what does 'tell by letter' mean?"

(b) Overlooking the highest number - e.g:-

G. 122/127 (K) thought that the task was to find the highest letter, but when this idea was corrected she still chose (K) "because it is the largest number anyway." When asked to check carefully, she still gave (K) - then when asked to read the item aloud she at last saw (H). Similarly G. 111/119.

G. 84/88 (8970) - when directed to the instructions gave (K), and could not see the highest number until asked to read the numbers slowly aloud.

B. 133/132 (K) when asked to read the item again said "oh... (H)....I didn't look properly" - a similar comment was made by G. 119/119 (K).

(c) Prepotency of two similar figures - e.g:-

B. 110/113 (K) - his attention was attracted to (B) as a likely answer, then to (K) as a similar number, and in deciding between (B) and (K) his attention to other numbers was weakened. When asked to read over again, he commented - "Oh - (H) - I was concentrating on comparing (K) and (B)."

G. 101/100 (K) - "I saw B89... then looked along and saw K larger, so decided to choose (K)." Even when she read the numbers through twice, she still could not see past (K). Similarly B. 100/112 and B. 91/96.

(d) Other types of errors noted were - correct choice, but wrong response; complete confusion as to the nature of the task; losing the task direction in concentrating on a detail - e.g:-
even after rereading, she still missed the point of "tell by letter".

G. 108/114 asked "Do you add the numbers together?" - When directed to read the item carefully she asked - "What does it mean 'tell by letter'?"

G. 93/100 (K) - "...because I went to the first figure first, saw that (H) started with 9, but then went on to check the second figure and saw that (K) is 89, which is a larger second figure than (H), so chose (K)."

Item 20.
Which one of the words below would come first in the dictionary.
(1) tramp, (2) saint, (3) razor, (4) quart, (5) grass
(6) night, (7) paint.

N=49. (1) 14, (2) 16, (3) 6, (4) 8, (6) 45, (7) 12.

There were only 8 errors repeated in the reworking - 5 of these being due to failure to notice "grass" and "night". As far as could be ascertained some children looked at the words as a whole, noticed the first few and the last, saw that the last was first alphabetically, and therefore chose it - e.g:-

B. 99/105 (7) described such a process and said that once he had chosen "paint" he did not bother to check up with all the words.

Others did not look over all the words, but took the first word (alphabetically) from only a group of the original 7 - e.g:-

G. 92/84 (3) said that she had not really looked properly past the first three words.

An interesting case of confusion of task direction was noted - e.g:-
G. 93/100 (7) "... because (p) comes nearer to (z), and (a) (i.e. paint) comes before the others, so it must be paint" - i.e. she was choosing between "grass" and "paint", but got the direction wrong for the first letter, (because she forgot momentarily that she had to find "what would come first"), and right for the second letter. As well as this she had ignored "night" completely.

Item 29.
Which one of the words below would come first in the dictionary?
(1) brave, (2) burst, (3) broke, (4) build, (5) breadth
(6) brown, (7) bunch, (8) bribe.

N=69. (2) 16, (3) 4, (4) 24, (5) 29, (6) 6, (7) 12,
(8) 9.

The 8 errors repeated were of three types:-

(a) Failure to examine each word - e.g:-

B. 106/118 (5) "... because 'e' is the first of the letters coming after 'br'" - when he was asked to check he saw 'brave' and explained that in his first reading he saw that there were 'br's' and 'bu's', came across (5), so picked that one, but had not even noticed 'brave'.

(b) Confusion over the position of 'r' and 'u' in the alphabet - e.g:-

G. 97/98 (4) saw that 'build' would come first of the 'Bu' words, but on checking realized that she had placed 'u' before 'r' although she knew the alphabet quite correctly.

(c) Losing sight of the whole task in concentrating on a part of it - e.g:-

G. 100/104 (5) - saw that she must pick between 'brave' and 'breadth' and argued "breadth has 'e' then 'a', while brave has 'a' then 'v' - so it must be (5)" - i.e. in deciding on the internal arrangement of these two words she had lost sight of the priority of 'a' because she was also trying to compare, at the same time, the fourth letters 'a' in breadth and 'v' in brave.
Item 43.

If the following words were arranged in order, with what letter would the middle word begin?

Eight Ten Six Nine Seven (Ans. = E)

N=76. (8) 39, (S) 13, (G) 9, (T) 7, (N) 5, (5) 5, (Others) 22, (Omissions) 24 cases.

Again, only 8 errors were repeated, three of these being due to a failure to follow the instructions - e.g.:-
B. 126/128 (8) - i.e. correct order, but ignored the direction to write the letter of the middle word - similarly G. 110/103, B. 91/96. All three immediately corrected themselves when asked to reread the instructions.

Two errors resulted from a correct arrangement but a misreading of the instructions - e.g.:-
G. 110/103 (G) - "I put down the middle letter of eight because they ask for the middle word of the sentence." Similarly B. 80/86.

Miscellaneous errors were:-
G. 112/114 (7) - i.e. arranged the words in alphabetical order, then gave the number of the middle word.
B. 82/83 (ivgie) - i.e. arranged the words correctly, then gave the middle letter of each word.
B. 77/79 (9) - put the words in order, miscounted the middle word, and gave its number.

Item 55.

If the following words were arranged in order, with what letter would the middle word begin?
Week Year Hour Second Day Month Minute (Ans. = D).

N=52. (H) 38, (W) 33, (S) 12, (M) 6, (I) 4, (Others) 8, (Omissions 29 cases).

Most of the errors arose as a result of confusion about the middle word, (in fact 19 of 27 errors). In all these cases the words were correctly arranged, but (H) or
(W) was chosen as the middle word (i.e. depending on whether the rearranged words started with second or year. This error came about in two ways:-

(a) Miscounting the total number of words as 6, and taking the third word along - e.g.:-

G. 118/115 (H) - counted the words as six and took the third word without checking.

(b) Counting the total number correctly, but confused about which would be the middle word - e.g.:-

B. 94/102 (W) - "...there are 7 words, and half 7 is $3\frac{1}{2}$ so would have to take the third word along - year, month, week." This confusion was evident over a wide range of I.Q. groups.

(c) Other types of errors noted in individual cases are illustrated in the following examples:-

B. 116/116 (W) - had great difficulty in arranging the words because he read "Second Day" as one unit in the series. A similar reading led G. 105/105 to attempt to arrange the words into a sentence!

G. 112/114 (M) arranged the words in alphabetical order and gave a correct middle word (using this principle of arrangement). She did the same thing in the previous item of this category.

B. 97/104 (Day) - overlooked the specific instructions, i.e. to put 'D' rather than 'Day' until he had reread it twice. Similarly G. 90/96 who did not change until she had read the instructions three times.

B. 98/96 (H) - arranged the words correctly and then gave (H) as his answer "because it says the third word."

B. 83/91 (W) - left out "hours" and had difficulty in seeing that he had not used all the words.

There were no cases where (S) was given, although from records on similar items in other categories it is probable that in the total group, children who gave (S) merely took
the middle word of the words as they appear in the item, and made no attempt to rearrange the words. The same was in all probability the case in the previous item (43) where (S) was also the middle word of the disarranged words.

**Item 64.**

If the following words were arranged in order, with what letter would the middle word begin?

General Lieutenant Private Colonel Sergeant (Ans. = L)

N=43. (S) 42, (P) 16, (G) 16, (C) 14, (5) 12.

(Omissions 25 cases)

More girls than boys omitted this item (14 to 11), and of those who answered, only 3 chose (S) as against 15 boys.

Many girls apparently had no confidence in their knowledge of army ranks, although the confidence shown by many boys was without foundation, for they were often completely confused.

The analysis of errors was not very rewarding, and pointed to the poor quality of the item rather than to the processes operating to produce the children's errors. Only 4 out of 30 of the children who reworked this item gave a fully correct answer, and most of those who gave an incorrect answer had very little conception of the relative ranks.

(a) Apart from the 4 mentioned above, there were 11 who gave the correct answer, but with a wrong or incomplete reasons - e.g.:-

G. 122/127 (L) - had arranged the words in alphabetical order, and (L) happens to remain as the first letter of the middle word! Similarly B. 111/102 (L) and G. 90/97 (L).

Eight others gave rearrangements which were incorrect in one
or more ways, yet in which "lieutenant" was the middle word -

**e.g.**:

**B. 128/128 (L)** - gave as his rearrangement, private, colonel, lieutenant, sergeant, general. Other arrangements were - 

(b) Other types of errors are illustrated in the following individual examples:-

**B. 116/116 (4)** - i.e. he arranged the words (incorrectly) with 'colonel' as the middle word, and as this is the fourth word along in the original order, he wrote (4) as his answer.

**G. 119/119 (E)** - "...because the order is private, colonel, general, and the middle letter of the middle word is (E)."

**G. 111/119** could not decide how to arrange the words because she thought that "private" went with one of the other words - i.e. in an adjectival sense.

**G. 97/105 (P)** - "...because private is the middle word in the group." - i.e. she took the words as arranged in the item.

**Item 69.**

If the following words were arranged in order, with what letter would the middle word begin?

Youth Infancy Manhood Childhood Birth (Ans. = C)

N=28. (I) 46, (Y) 36, (M) 7, (T) 4, (3) 4, (B) 4.

(Omissions)11 cases.

The word which caused most trouble was "infancy", although 'youth' was also misplaced frequently. When questioned it was clear that the children were able to distinguish between infants and children, and one of the reasons which may have led a number to choose (I) is illustrated in this example:-

**G. 96/96 (I)** - said that she chose infancy because she thought of children, and then the first stage of school for these children - i.e. the infant class; so first the child, and then the infant, hence birth, childhood, infancy...
Most of the errors were found in individual cases only - e.g.:—

B. 118/122 (D) - i.e. the middle letter of the middle word. (also G. 119/119 (D)).

B. 126/128 (Y) - could not pronounce 'infancy' and thought it might have something to do with age. When the word was pronounced for him he recognized it immediately and gave the correct answer.

B. 116/116 (Y) - when asked to work over the item again, it was found that he had completely overlooked "infancy".

B. 106/118 (Y) - gave the correct order and then miscounted to get the middle word.

Item 73.
Which one of the following words would come last in the dictionary?
(1) heart, (2) judge, (3) grass, (4) nerve, (5) horse, (6) north, (7) labour.

N=15. (2) 3 cases, (3) 2 cases, (4) 6 cases, (5) 1 case, (7) 2 cases, (6) 1 case. (Omissions) 3 cases.

There were only 6 errors repeated on reworking; one of these was caused by misinterpreting the instructions - i.e. "first" instead of "last", and the other 5 were all of the type described in the following example - e.g.:—

B. 126/116 (4) - picked out the two competing words - i.e. "nerve" and "north", compared the second letters, and said - "'e' and 'o' to choose between, well 'e' comes before 'o' so the answer must be 'nerve'." He had established a "directional" set - "which comes last" - but in concentrating on the details, this set was weakened and the "habitual" alphabetical direction of 'e' before 'o' displaced it.

(G)
Category (7) - Analogies.

Of the 13 items in this category, 10 were in the first half of the test, while the last was item 50, thus there were
few items which would extend the children in the higher
groups, nevertheless they made typical errors on most of the
items.

**Item 4.**
The peeling is to a banana and the husk to an ear of
corn the same as shell is to what?
(1) an apple, (2) an egg, (3) juice, (4) a peach,
(5) a pen.

N=22. (1) 14, (3) 23, (4) 5, (5) 59.

Six children - ranging from B. 126/116 to G. 72/76 -
gave the correct answer but it was clear from questioning
that they had not taken account of the first half of the
sentence but had merely given the obvious answer to the pro-
blem as they formulated it; - i.e. "What goes with a shell?"

One error illustrates a typical approach, e.g.:

B. 115/110 (2 and 5) - "...because you have to find two things
that go together" - i.e. he had read through the question,
formed a vague idea of what was required, and acted on that
idea. On rereading, he saw that he had not read the question
properly.

**Item 7.**
Wool is to sheep as feathers are to what?
(1) a pillow, (2) a rabbit, (3) a bird, (4) a goat,
(5) a bed.

N=11. (1) 4 cases, (2) 1 case, (4) 1 case, (5) 5 cases.

No errors were made by the few who reworked this item,
but two correct answers were given for the wrong reasons - e.g.:

B. 126/116 (3) - had just reread item (4) prior to attempting
this item, and gave as his reason for (3) "...because feathers
are round the bird" - i.e. just as a shell is round an egg.
He carried over the relationship from (4) to (7) and had again
failed to read the first part of the sentence - or at least
to see what it implied.
C. 72/76 (3) - gave her answer immediately, but had no idea why the "wool is to a sheep" was there; i.e. she gave a common association of "feathers". It is not known how many other children gave a correct answer as a result of this inadequate formulation of the problem.

**Item 11.**

Hat is to head as thimble is to what?

(1) finger, (2) needle, (3) thread, (4) hand, (5) sewing.

N=45. (2) 36, (3) 13, (4) 13, (5) 38.

Relatively fewer girls (31%) than boys made errors on this item - although this could be expected in the light of the illustrations given above, for "thimble-finger" would be a stronger common association for girls. Errors made illustrated the fact that the implication of "hat-head" was either ignored or only partially attended to, e.g:--

B. 133/132 (5) - "...because sewing has to do with thimble." When asked why the first part of the sentence was there, he said that it gave an idea how to do it, and then when directed to look at it more carefully he saw the relationship and gave the correct answer. His failure was not due to an inability to deduce a correlate, but to a failure to clarify the task.

B. 129/142 (1 or 5) - "...could be both, put it on your finger and use it for sewing - you use it more for sewing than on your finger." When asked why "hat is to head" was there he replied - "...well it gives a clue, wear a hat on your head, so a thimble on your finger, but you wear a hat to keep your head warm and a thimble for sewing...you don't go round wearing a thimble on your finger for sewing." His answer was the result of a rather complicated set of relationships dominated by the purpose of a hat on head, and the fact that a thimble is not always worn on a finger while sewing, although it has no function except in sewing.

B. 82/83 (2) - completely ignored the first part, while B.80/86 (5) read it and interpreted the relation as "is used for" - i.e. a hat is used for your head, a thimble is used for sewing.

One error was due to the overlooking of "finger" e.g:--
B. 83/91 (4) said that "hand" seemed to be the best, but on rereading, he noticed "finger", which he had not even seen before.

**Item 15.**
A hand is to an arm the same as a foot is to what?
(1) leg, (2) toe, (3) finger, (4) wrist, (5) knee.

N=42. (2) 64, (3) 5, (4) 10, (5) 21.

Only 2 errors were repeated - each illustrating a different approach - e.g.:-

B. 101/100 (5) - when asked to check gave 'leg', not having even noticed it in his first reading.

B. 82/83 (2) - "...because a foot hasn't got the others, but has got a toe" - i.e. an arm has got a hand, so a foot has got a toe. He not only failed to see the exact relationship, but reversed the direction of the original statement.

A correct response was given by B. 99/105 although he had not worked out the relationship, and said that "toe" would be just as good as "leg".

**Item 19.**
Coal is to a railway engine as what is to a motor car?
(1) motor cycle, (2) smoke, (3) wheels, (4) petrol, (5) horn.

N=15. (1) 5 cases, (2) 2 cases, (3) 5 cases, (5) 3 cases.

The only error which was repeated was that of G.84/88 (5) who formulated the question as "coal goes with railway engine, horn goes with car" - i.e. she was looking for some distinctive characteristic of a motor car.

**Item 22.**
A motor car is to a wagon as a motor cycle is to what?
(1) walking, (2) horse, (3) buggy, (4) train, (5) bicycle

N=35. (1) 9, (2) 20, (3) 53, (4) 9.

Two interesting points were illustrated in the reworking:-
(a) Some children chose (3) because they interpreted it as referring to the "side-chair" of a motor cycle - e.g.:-

G. 130/125 (3) - "...a wagon is what a motor car tows behind, and a motor bike takes a sort of buggy with it." (i.e. wagon = trailer).

B. 116/116 (3) - "...because a buggy fits on to a motor cycle." When told that a buggy is a small cart, he changed to (5).

G. 115/116 (3) - "...buggy goes with motor cycle." The first part of the sentence was just ignored.

(b) Six children gave the correct answer, but had not really formulated the question adequately - e.g.:-

G. 116/120 (5) - "...because you can sit in a motor car and in a wagon, and can sit on the seat of a motor cycle and on the seat of a bicycle."

G. 108/114 (5) - "...because motor car and wagon have wheels, and motor cycle and bicycle have wheels."

G. 106/108 (5) - "...the first part is to tell you that a wagon goes with a car, and so a bicycle goes with a motor cycle."

In this, and other similar cases, the words "is to a", which are crucial, were passed over very lightly as merely indicating some rather general association. Further questioning to clarify the original relation usually resulted in the seeing of the correct answer with all the signs of "insight".

Item 26.
A hospital is to the sick as what is to criminals?
(1) doctor, (2) asylum, (3) judge, (4) prison, (5) sentence.

N=19. (1) 10 cases, (2) 3 cases, (3) 4 cases, (5) 2 cases.

There were no errors made in the few cases in which this item was reworked. The original weighting of errors on (1) may have been due to "atmosphere effect" in conjunction
with a failure to clarify the requirements of the problem — i.e. hospital — sick — doctor. Such could be the case where the "hospital-sick" part of the sentence was prepotent — and the "criminals" thus ignored.

**Item 30.**

Better is to good as worse is to what?

(1) very good, (2) medium, (3) bad, (4) much worse, (5) best.

\[N=57, \text{ (1) 11, (2) 7, (4) 75, (5) 7.}\]

This proved to be one of the most difficult items in the test for the children to explain, and of the 19 intelligible explanations, 14 concerned the choice of (4), although for different reasons:-

(a) Failure to read the question fully — e.g:-

B. 133/132 (4) — "...we've been taught that you start off with better, best and so on, and this does the same — starts bad, worse, much worse." When asked to read the question again and to note what the first part started with, he said — "good — better, so "bad" — I didn't really read the question properly."

G. 105/101 (4) when asked to read the question again said that she hadn't bothered about the first part.

(b) Confusion over "bad" and "worse" — e.g:-

B. 113/118 (4) — "...bad means the same as worse practically, so can't take that, so take much worse."

B. 100/112 (4) — "...bad goes with badder so much worse goes with worse."

(c) Altering the direction of the analogy — e.g:-

B. 121/125 (4) — "...better is to good is on the good side of it and much worse is worse than worse" — i.e. is on the worse side of worse. In this and 5 other cases the analogy was read thus:- better down to good, worse down to much worse.
B. 111/106 (4) - illustrated this way of handling the problem in his comment - "If you are worse than worse, then you must be much worse."

G. 101/100 (4) argued - "Good is a bit worse than better, so much worse is a bit worse than worse."

B. 115/110 (4) illustrated his choice as - "...normal writing, and it's much improved so better; normal writing and it gets worse, so much worse;" - i.e. he interpolated 'good' and 'bad' without recognizing the fact - normal-good-better; normal-bad (i.e. worse)-much worse, and then equated "bad" with "worse than normal".

(d) Other types of errors are illustrated in the following examples:

G. 117/118 (2) - "...worse would be medium between worse and excellent and better is just about as good as good" - i.e. she was trying to judge the words "absolutely" instead of "relatively". Another choice of (2) was justified more reasonably by G. 97/105 (2) - "...good seems not to do so good when it's better, so worse must be to medium" - i.e. she was thinking of a classification such as, better-good-medium-worse.

G. 72/76 (4) chose "worse" because it occurred in both the instructions and the answers.

G. 76/80 (5) explained her reasoning as - "I said better is to good and I just looked at "best" and said "best"," i.e. ignored "worst" altogether.

(e) In two cases the correct answer was given for the wrong reason - e.g:-

B. 108/110 (3) - "...because it is the opposite of worse."

B. 106/118 (3) - "...because better is a bit more than good and much worse is worse than worse" - i.e. he should have chosen (4) on his own argument, and in fact could not justify his choice of (3).

Item 34.
Grass is to cattle as bread is to what?
(1) butter, (2) flour, (3) milk, (4) man (5) horses.

N=101. (1) 53, (2) 41, (3) 5, (5) 1.
All groups made errors on this item, mainly on (1) and (2), and these can be classified as:-

(a) Failure to read all the data, i.e. the first part of the sentence, or all the answers, e.g.:

G. 113/125 (1) - "If grass is to cattle, then bread you usually have butter with it" - in fact she had not looked along the line past "butter".

G. 108/114 (2) - "You make bread from flour." When asked what the first part of the sentence was there for, she reread the item and said - "Cattle eat grass and bread is made from flour... oh... perhaps (4) because cattle eat grass and men eat bread." She commented that previously she had only read the first two answers.

B. 100/112 (1) when asked to explain he reread the question and changed to (4) "...because I wasn't looking at grass and cattle."

B. 95/103 (2) when directed to read it again and to look at the part at the beginning of the sentence, he said - "Well, cattle eat grass... so (4)...I hadn't really read the first part."

(b) Educating a different relation - e.g.:-

G. 117/125 (2) "...because cattle eat grass and bread is made out of flour" - i.e. she educed the relation "grass is what makes cattle" and then reversed the direction of the analogy.

B. 110/113 (2) changed to (4), "...because at first I thought it might have something to do with ingredients."

B. 82/83 (2) - "I thought grass is to cattle (makes cattle) and flour makes bread."

B. 108/110 (1) - "...because you usually have butter with bread." When questioned about the significance of the first part, he could not see past the relation "goes with".

B. 93/100 (1) - "...well the nearest to bread is butter."

G. 96/96 (1) in her explanation of (1) summed up the attitude of those who chose "butter" because it is "nearest to bread" - you have to find what goes with it."
(c) Educing the relation but misapplying it - e.g.:

G. 93/89 (1) - "...because you can put butter on bread, and you can't put flour on it to eat it." When asked how she got the idea of "eating", she said - "Well, 'grass is to cattle' is eating, and if you put butter on bread you can eat that."

B. 76/80 (1) - "You eat bread and butter together, cattle eat grass and we eat bread and butter."

(d) A more sophisticated form of type (b), is illustrated in the following report:

B. 121/125 (3) - "Bread and milk go together; when bread is spoken of, then usually such necessities of life as milk usually go with it, for example you go to the shop to get bread and milk...but...oh...I hadn't seen "butter" it could be (1)." When asked why the first part of the sentence was given he said - "You can't have cattle without grass, so it can't be (3) but it's (1) because bread and butter always go together." A further question as to the sort of connection between grass and cattle resulted in this comment - "Grass is the main food, so it would be (4) if you look at it that way."

Item 38.
A king is to a kingdom as a president is to what?
(1) vice-president, (2) senate, (3) republic, (4) queen
(5) democrat.

N=93. (1) 48, (2) 20, (4) 19, (5) 12.

Girls made 55% of the errors on this item, mainly because of a heavy choice of (1). Almost all the errors resulted from ignorance or confusion about the meaning of president, republic, senate and democrat, hence the analysis of this item did not produce much of value from the point of view of error causing processes. The obvious choice of (1) was forced on many children, because it was the only word which had at least a modicum of relationship with president, as (2), (3) and (5) were unknown. One characteristic which appeared,
when children were asked to work the question aloud, was the
tendency for attention to be focussed on the five words, at
the expense of the sentence, so that when questioned about the
significance of "king to kingdom", these children showed that
they had not really considered this part of the problem.

In some cases the correct answer was given for the
wrong reason - e.g:--

G. 117/118 (1) changed to (3) when asked to reread, because
"the president rules the public of a club." Similarly
G. 116/120 who thought of a president as having to do with
a group of people.

G. 112/114 (3) could not explain her choice clearly, but
thought of republic as "something to do with a council."

B. 94/102 (3) chose his answer because - "When there was a
poll day dad was telling us about it and saying that Mr.
Hollandis the president of the republic in New Zealand."

B. 76/80 (3) described "republic" as "government or something
like that," and was not able to give any relation between
"king" and "kingdom" except to point out that one had "dom"
added on.

Item 42.
Large is to object as loud is to what?
(1) soft, (2) small, (3) heavy, (4) weight, (5) sound.

N=57. (1) 81, (2) 4, (3) 14, (4) 2.

Few errors were repeated, and those that were, result-
ed from a failure to understand the structure of the question,
or a misreading of a word.

(a) Failure to understand the requirements of the problem - e.g:--

B. 108/115 (1) - "...because "loud" and "soft" are opposites." When asked if that was what he was required to find, he re-
read the question and saw that 'large' described 'object' so
(5). A similar situation arose with G. 97/98 (1).
B. 95/96 (5) - (i.e. correct answer) thought that 'loud' would go with 'sound', although he had not really seen the structure of the question at all. A similar reason was given by G. 92/84 (5) for her choice.

(b) Misreading or wrong interpretation of a word - e.g:-

G. 108/114 - could not make any headway with the question. When she read it aloud it was found that she was reading 'object' as 'object' (i.e. protest) and could not think of the word as having any other meaning. When the correct pronunciation was given she saw that "large" was an adjective hence (5). B. 95/96 (5) also mispronounced 'object', while B. 82/83 (4) mispronounced 'object' and read 'loud' as 'load', hence the answer given was 'weight'.

Item 46.
A captain is to a ship as a mayor is to what?
(1) state, (2) council, (3) city, (4) boss, (5) lawyer.

N=110. (1) 15, (2) 68, (4) 5, (5) 13.

Errors were concentrated on (2) throughout all groups, although other errors are included in the types illustrated:-

(a) Rejection of the correct answer on reasonable grounds - e.g:-

G. 130/125 (2) - "A captain rules over a ship so mayor over a council." Her reason for rejecting (3) was - "well, he is not the highest person in the city...the council is made up of people responsible for different things that happen and the mayor is one of them."

G. 118/115 (2) - chose (2) rather than (3) because the mayor is head of the council - "a body of men who keep rule over a city - to keep order."

G. 110/123 (2) - "A ship has a captain so a council must have a mayor over it."

G. 106/108 (2) - "I used to live in Kaiapoi, that had a mayor and yet it is not a city, so it can't be a 'city', it must be 'council'."

G. 83/83 (2) - "The head of the council is called the mayor, just as the captain is head of the ship."

(b) Misreading or misinterpretation, e.g:-
B. 100/105 (1) - "There is a major in the army" - i.e. this is part of the state, so (1).

B. 82/83 - read 'major' for 'mayor'. When corrected he gave the correct answer.

G. 100/104 (5) - "A lawyer puts down laws and if he says it is to be the law everyone has to obey it." When asked what this had to do with "mayor", she was puzzled because she had read it as 'magis' and thought this must be 'magistrate' - hence (5).

B. 97/104 (1) - "Well a mayor is high up, and he should be to state." Questioning elicited a description of 'state' as - "a big high building that rules or something," - i.e. a captain rules in his ship, a mayor rules in his state.

(c) Failing to read all parts of the question - e.g.:

G. 109/112 (1) - "Mayor belongs to 'state'." When asked why she said "belongs to", she answered - "just a word that came to me." She had completely ignored the first part of the sentence.

B. 111/102 (1) - also completely ignored the first part, and on rereading changed to (2).

B. 90/92 (4) - "...because both bosses or leaders." Had not read the first part. Similarly G. 72/76.

(d) Unfamiliarity of "mayor".

G. 114/117 (2), when asked if Christchurch had a mayor, she said "no"!

B. 98/96 - had no idea what a 'mayor' is.

B. 90/92 (4) - said that there was only one mayor for New Zealand.

G. 90/96 (3) - thought that there was a mayor for each suburb of a city.

(e) Correct answer but inadequate explanation - e.g.:

B. 115/110 (3) took 'city' because "a council is where a mayor works" - i.e. he was thinking of "council" as a building - e.g. council chambers.
G. 108/114 (3) - had completely ignored the first part of the sentence - similarly G. 90/96 (3).

Item 50.
A revolver is to a man as what is to a bee?
(1) wings, (2) honey, (3) flying, (4) wax, (5) sting.

N=81. (1) 17, (2) 78, (3) 0, (4) 5.

(a) Educting a more general relation - e.g:-

G. 117/121 (4) - "A man uses a revolver and bees use wax."

B. 91/96 (2) - "A man might own a revolver just as bees own honey."

B. 90/92 (1) - "Well, a man needs a revolver, like a crook, he needs it like a bee needs wings."

G. 97/98 (1) - "...because they belong to a bee." When asked what the first part of the sentence was there for she said that it was just an example of "what goes with what."

These examples show how an error may result from the educing of a reasonable but insufficient relation. When children were asked the question - "What does a man use a revolver for?" in most cases they gave a correct answer and immediately changed to "sting".

(b) Failure to use all the data - e.g:-

G. 117/118 (1) - "...because a bee has to have wings to fly with." When her attention was drawn to the first part she gave the correct answer and said that she had just passed over it before.

G. 100/104 (1) - gave her answer straight away, completely ignoring the first part.

B. 87/94 (2) - "...because a bee gathers honey." When asked about the first half he read it and said - "Oh! - suppose you should say 'sting' is to bee because it is a weapon," - i.e. he had overlooked the first part.

G. 91/96 (2) - changed to (1) - "...because a bee has to have wings to be able to fly." She was looking for the most im-
important characteristic of bee and ignoring the first part.

B. 77/79 (1) - also looked for the most important characteristic - i.e. "If a bee didn't have wings it couldn't fly."

In some cases the whole sentence was considered, but some of the five words given were ignored - e.g:-

B. 83/91 (1) - when asked to read the question again, he changed to (5) because he had not seen it on his first reading.

G. 96/96 (2) - when questioned she said that when she came to "honey" she did not read past it.

(c) Misreading and unfamiliarity of a word, e.g:-

B. 102/100 (2) - when asked to read the sentence he had difficulty in pronouncing "revolver" but finally managed it and chose (2) as the best word. When asked to check over the other words he rejected (5) because "string" had nothing to do with the question.

G. 104/108 (2) - "...because honey belongs to bees." Questioning elicited the fact that she did not know what a "revolver" was - but when told it was a "gun" she changed to (5).

G. 90/96 (2) - passed over the first part because she was not at all sure what a "revolver" was - just "something to do with a man."

(d) Correct answer but inadequate reason, e.g:-

G. 90/97 (5) - "...because a bee stings." She did not have any idea what a "revolver" was.

G. 100/95 (5) - "All bees have stings, but not all bees have honey." When asked to explain how this was related to the first part she said that she did not know the word "revolver."

G. 84/88 (5) - "I just chose (5) because it looked right" - although she had no idea why the first part was there.
(H) Category (8) - Verbal Reasoning.

(A).- Multiple Choice Items.

Item 36.
The son of my father's sister is my ( )
(1) brother, (2) nephew, (3) cousin, (4) uncle,
(5) grandson.

N=95. (1) 16, (2) 51, (4) 12, (5) 22.

This proved to be a most confusing item for some children, as can be seen from the following examples:-

(a) Unfamiliarity with, or misinterpretation of the word "cousin", e.g.:-

B. 106/113 (2) - did not know the word, and has no cousins himself.

B. 83/91 (4) - when asked about the word "cousin" he said that he thought that cousins were only girls, - i.e. he has no boy cousins.

B. 110/113 (5) illustrates a related confusion which muddled some children - i.e. inability to distinguish between "cousin" and "nephew".

(b) Confused about the relationships in the item, yet can describe "cousin" correctly - e.g.:-

B. 124/123 (5) - hesitated between (2) and (5), but when asked if he had any cousins he was able to describe the cousin relationship accurately, and changed to (3).

B. 92/92 (5) - "...because the son of father's sister...well it was sort of a wild guess." When asked if he had a cousin he described the relationship accurately.

B. 82/83 (5) - illustrates very clearly the confusion caused by presenting an abstract relationship which has to be applied concretely, - "My father, I would have a sister, my sister has a son and if I was my father's brother so my sister's son would be my grandson." When asked if he had a cousin and if so what relation the cousin would be, he answer-
ed - "Yes - the daughter of my mother's sister."

(c) Failure to read all the requirements of the question, e.g.:

G. 91/96 (4) - gave the correct relationships but had not seen that what was required was the relationship to herself - i.e. she had ignored the "my" as the end of the sentence.

G. 93/89 (5) - just saw "father" and "son" so gave (5). She had not stopped to consider "sister" at all.

(d) Correct answer with inadequate reasons:

B. 115/117 (3) - but could not clarify the distinction between "nephew" and "cousin" - similarly G. 106/108 (3).

G. 83/83 (3) - "...because (1), (2), (4) and (5) couldn't be girls, and a cousin can be a girl, so it must be (3)."

Item 37.

If George is taller than Frank, and Frank is taller than James, then George is (?) James.

(1) taller than, (2) shorter than, (3) just as tall as, (4) (cannot say which)

N=83. (2) 40, (3) 24, (4) 36.

(a) Lack of confidence, e.g.:

B. 126/116 (4) - "These always get me muddled." he had chosen (4) because he was muddled, not because he thought that there was enough evidence.

G. 113/125 - "I don't like these, I start reading then it goes all wonky." When asked to read the question carefully she gave the correct answer.

(b) Failure to note the specific requirements of the task, e.g.:

B. 126/128 (2) - "...because James is shorter than George." On rereading the question he gave the correct answer - i.e. he had worked out the relationship correctly, but overlooked the specific requirements of the question.

G. 106/108 (2) - put them in reverse order, i.e. read the words "taller than" but worked backwards - George smaller than Frank etc., so that George would be shorter than James.
B. 80/86 (2) - "...because James is shorter than George;" - i.e. he got the correct order but forgot or overlooked the particular requirements.

(c) Failure to see a crucial word - e.g:--

G. 93/99 (4) - on reading the question she stopped at the question mark, and had not seen "James" at the end of the line. This error occurred with 5 other children, who gave (1) as the answer (i.e. correct answer), because not having read "James" at the end, they thought that the question was concerned with a characteristic of George, not specifically related to Frank or James. These 5 ranged from B 102/113 to G. 86/88.

B. 126/116 (4) - when asked to work the question, was so intent on "James" that he kept on overlooking "George" in the sentence.

(d) Confusion in keeping in mind the requirements of the task - e.g:--

B. 91/96 (2) - in demonstrating his answer he drew lines to represent the names, i.e. a long line for George, shorter for Frank and a shorter one for James, then a fourth line shorter still to represent George again (because his name occurs again after James) - therefore George is the shortest, so (2).

B. 83/91 (3) - "Well George is taller than Frank, so Frank and James must be about the same height, so (3)." When asked what he had to find, he said that he had to compare Frank and James, although on being directed to reread the question he saw the requirements and gave (1). A similar process occurred with G. 97/98 (3).

B. 76/80 (2) had difficulty in reworking the item, but finally established the correct relationships of George as taller than James, and then put "J" in the brackets.

(e) Misreading of words - e.g:--

B. 103/107 (3) - when reading the question read "taller than" in each case as "smaller than".

(f) Correct answer with inadequate reasons (apart from the 5 cases described in (c) above) - e.g:--
G. 105/101 (1) - but she could not explain clearly, and it seemed that her answer was due to "atmosphere effect" - i.e. "tallness". The same situation arose with B. 97/105 (1).

Item 48.
If Harry is older than William and William is just as old as Charles, then Charles is (?) Harry.
(1) older than, (2) younger than, (3) just as old as, (4) (cannot say which).

N=107. (1) 34, (3) 57, (4) 9.

(a) The greatest number of errors resulted from the failure to see the word "Harry" after the question mark (8 out of 20 errors) - ranging from G. 115/116 (3) to B. 80/86 (2). A typical explanation was:

G. 86/88 (3) - "William is just as old as Charles, so they are just as old as each other." When she was asked what it was that she had to find, she said - "What Charles is."

(b) Another recurrent source of error (5 out of the 20) was the establishing of the correct relation, but the reversal of the specific requirement - i.e. Harry is older than Charles (1), instead of Charles is younger than Harry (2), e.g.:

B. 124/123 (1) - "Well, if Harry is older than William... (etc)...then Harry is the oldest so (1).

(c) Similar to type (b) was failure to maintain direction and the confusion of the task requirements - e.g.:

B. 110/118 (4) - "...because 'just as old as' doesn't help, you can't tell who is older." From questioning it appeared that although he read the first part - "Harry is older..." it was completely overshadowed by the prepotent phrase - "just as old as," which he presumed was the key statement.

B. 100/112 (3) and B. 108/115 (3) both became confused about what they were trying to find. On rereading they gave the correct answer.
G. 96/96 (2) and G. 92/84 - read "older than" throughout - i.e. an assumption that this was the same type of question as (37) where "taller than" occurred throughout.

B. 91/96 (3) and B. 94/102 (1) were completely lost in their explanations, because they could not reconcile the presence of two different Harrys in the same question!

Item 51.

If Paul is older than Herbert and Paul is younger than Robert, then Robert is (?) Herbert.

(1) older than, (2) younger than, (3) just as old as, (4) (cannot say which).

N=85. (2) 53, (3) 22, (4) 25.

The same general types of error appeared on this item - e.g.:-

(a) Reversal of the specific requirement - e.g.:-

B. 110/118 (2) established the correct relationship and placed Robert as oldest then said - "...so Herbert is younger than Robert...so (2)"

(b) Failure to see the word Herbert after the question mark, e.g.:-

B. 108/110 (1) - he argued that Robert is older, but did not know that he had to compare Robert and Herbert specifically.

(c) Confusion of direction within the task - e.g.:-

B. 111/102 (2) - "Paul is younger than Herbert, so Paul is youngest, so if Paul is younger than Herbert, then...oh...the other way round, Paul is older than Herbert, then Paul is younger than Robert, then...oh...Robert is younger than Herbert:" i.e. the necessity to think of Paul in two "directions" at once, as both older and younger, was the stumbling block. This proved to be the major difficulty of the item in several cases, as the children refused to think of Paul in two roles.

B. 101/109 (3) read - "Paul is younger than Robert, then Robert is younger than Paul, but Herbert is younger than Paul so Herbert is equal to Robert, so (3)." This bow reversed the direction of one statement - thereby ruining an otherwise well reasoned answer.
B. Open Answer Items.

Item 45
In a foreign language, boy = Kolo
good boy = Kolo Daak
The word that means good begins with what letter (Ans.=D).
N=61. (K) 80, (G) 15, (Others) 5.

The few errors made on reworking, added little to the obvious classification of errors suggested by the percentages above, i.e. -

(a) Partial understanding of the task resulting in the choice of (K) despite the fact that the first line rules it out - e.g.:
B. 95/103 (K) - "...because Kolo means good." When asked to make sure by reading again, he said, "I think Daak is, because Kolo means boy," i.e. he had not really stopped to look at the question carefully.

G. 90/96 (K) - When questioned, she said that she thought you couldn't change them around, i.e. Kolo Daak could not be reversed.

(b) Ignoring the main part of the item in carrying out the instructions of the last line - e.g.:
B. 91/96 (g) - When asked why he chose "g" he replied - "I don't know any foreign language."

B. 76/80 (g) - When asked what all of the first part of the question was about, he said that it was a foreign language, but did not help to give the answer. After the first part was read slowly and clearly to him, he immediately chose (D), and gave a correct explanation.

Item 57.
In a foreign language very hot = Soto Gran
very cold = Foss Gran
The word that means very begins with what letter? (Ans.=G)
N=49. (F) 41, (S) 39, (V) 16, (Others) 4.

Only 10 errors were repeated, and these can be divided
up into two broad groups - i.e.:

(a) Assumption that the words cannot be reversed (a type of "habitual direction"). This led to confused attempts to explain the choice of (S) or (F) or both - e.g.:

B. 102/100 (S and F) - "...because I thought you wouldn't be allowed to reverse them."

G. 101/100 (S and F) - When explaining her choice she got confused because she did not think the words could be reversed. When asked - "What if the words don't need to be in the same order?" - she answered "...might be (S) because ...no, hot and cold are both the same word...no, I think very begins with (G) because hot and cold can't be the same word - must be back to front." Similarly G. 91/96 (S) and B. 95/96 (S), both gave (G) when told that they could reverse the words if they wished.

(b) Failure to note the requirements of the task, ranging from the forgetting of the actual mode of recording the answer, to the ignoring of the crucial data, e.g.:

G. 111/106 (Gran) - i.e. did not give the letter as asked; similarly G. 105/105.

B. 98/96, could not give an answer, "...because they should be round the other way," i.e. he established the correct relationships but then overlooked the actual answer required. When asked to read the item again he gave (G).

G. 83/83 (V), ignored the first part, but when it was pointed out that she could not answer the question without it, and when it was read again to her, she gave the correct answer.

B. 83/95 (V), also ignored the first part, but when his attention was directed towards it, he decided on (F), "... because Foss sounds like hot," i.e. he completely missed the crucial point of the task.

B. 103/99 (S and F) - "...no, (V), because the top part might be just to fool you".
**Item 62.**

In a foreign language

some food = Beko Prac

some milk = Klup Prac

some food and milk = Beko Otoh Klup Prac

The word that means and begins with what letter? (Ans. = O)

N=49. (K) 32, (B) 6, (P) 4, (a) 4, (3) 4.

The main source of error was failure to note the requirements of the task - as described for item 57, e.g.:-

G. 112/114 (K) ignored the top half of the item completely and just counted along the third line to the third word. Similarly B. 110/118 (K), B. 108/115 (K) and B. 97/104 (K).

B. 103/99 (P) read only the first two lines, saw that "Prac" was repeated so gave (P) as his answer, i.e. a carry over from the procedure for item 57. A similar neglect of the third line of the data led B. 98/96 into a confused attempt to find where the word "and" fitted in.

G. 106/108 (Otoh) had overlooked the specific instructions to give the "letter", similarly B. 91/96.

There were 5 cases in which the correct answer was given for a wrong or inadequate reason, e.g.:-

B. 108/110 (C), merely read the four letter sentence backwards until he came to the third word - i.e. "Otoh". This procedure was due to the persistence of the method of reversal which had been necessary for the correct solution of item 57. Three other children gave the correct answer for this reason, while the fourth, G. 105/105 (C) explained that she had chosen O because it was the only new word in the last sentence, although when asked which words the other three "foreign language" words represented, she was unable to work these out correctly.

**Item 68.**

Write the letter that follows the letter that comes next after K in the alphabet. (Ans. = M).

N=40. (L) 35, (H) 8, (J) 2, (Others) 4.

Of the 27 errors repeated, 24 were (L) and 3 (N). It
is clear, without any individual analyses, that the error sprang from a superficial or inadequate reading of the item in the majority of cases. It is hard, however, to convey the surprise which many children expressed, when asked to read the item again and when noticing, for the first time, that there was more to it than just giving "the letter that follows K." Furthermore, in reworking the item it was clearly evident that some children had made up their minds what the item required and several readings were necessary before this "set" could be broken. Several examples illustrate the type of error involved:

B. 121/125 (L), when asked to read aloud, read through the question and gave an immediate reaction, "...oh...M." This was typical of many ranging from the top I.Q. group to the lowest group.

G. 115/116 (L) illustrates the stubbornness with which some cling to their original interpretation. When asked to read aloud she still gave (L) - again, after reading slowly she gave (L), and did not see her error until it was read for her twice, with emphasis.

In the case of the three who answered (N), the reasoning was very similar - e.g.:-

B. 108/115 (N) argued thus - "Write the letter, that's one; that follows the letter, that's two; that comes next after the letter, that's three; so it must be three letters on after K, so (N)."

Most of these errors illustrate a type of "difficulty reduction", i.e. the item sounds a little complicated, but the clearcut feature "write...letter...follows...K." becomes prepotent and obliterates the other terms of the item,
thus reducing it to a straightforward question, easily manageable.

(I) **Category (9) - Proverbs.**

The distribution of errors for the total group was more evenly spread over the distractors in this category.

**Item 33.**
There is a saying, "A stitch in time saves nine." This means (?)
(1) A little sewing may save nine shillings. (2) It pays to attend to troubles before they get worse. (3) Work hard and save as much as you can. (4) You can save time by sewing.

N=120. (1) 42, (3) 38, (4) 21.

Success or failure on this item appears to hinge to a large extent on the subjects' interpretation of the word "saying". In all 17 cases of error on this item, the children concerned did not know what a "saying" was, i.e. they did not understand that the statement had a general application and therefore required a general statement to explain it. Without this understanding the children chose an answer which was related to the statement either by way of "sewing" or by "saving". In many cases, ranging from B. 121/125 (1) to G. 84/88 the correct answer was given when it was explained that "a proverb tells a little story, or is an example, which tells us something about the way we behave or the way we should behave, in all sorts of situations."
Item 59

There is a saying, "A drowning man will grasp at straws."

This means (?

(1) A man will sink more easily than a straw.  (2) Everyone should learn to swim.  (3) Desperate people cling to absurd hopes.  (4) Those who cannot swim should stay on land.

N=69.  (1) 35, (2) 33, (4) 32.

(a) As in item 33 the word "saying" was not understood, e.g:-

G. 111/119, described a "saying" as a superstition; B. 83/95 as "just a bit of information".

(b) Some errors arose from a partial generalization of the "saying", e.g:-

B. 124/123 (2) 'Well, the saying is a sort of a...like, there's one about fox and grapes, a proverb, and means it's no use trying to swim when you can't so everyone should learn to swim." Although he had not really bothered about the "grasping at straws", when his attention was drawn to it he rationalized by adding "If you are drowning and can't swim, nothing can save you."

G. 115/116 (4) said that a saying is "something true, but cut short" and when her attention was drawn to "straws" she said - "Oh...then it must be (1)" - i.e. she had generalized a part of the statement, but when the other part was pointed out she went straight to the statement in which it was prepotent, although this statement was more specific than her first choice.

B. 102/100 (2), also ignored "straws" and justified his choice of (2) by saying - "A man mightn't drown if he could swim."

G. 100/109 (1) ignored "straws" and read absurd" as "absent".

(c) Right answers for wrong or inadequate reasons, e.g:-

B. 110/118 (3) when asked to explain (3) he interpreted "absurd hopes" as "excited hopes".

B. 108/110 (3) - "...because desperate people grasp at straws when they are drowning." He did not appreciate the generalization involved in choosing (3).
G. 97/105 (3) - "...because they kind of hope that nothing will happen to them."

B. 94/102 (3) chose (3) because of the similarity of "grasp" and "cling". "Absurd" he described as "a severe hold". Similarly G. 96/96 (3), i.e. "grasp = cling".

Item 65.
There is a saying, "Make hay while the sun shines." This means (?)
   (1) Hay made in cloudy weather is poor. (2) Haste makes waste. (3) Make the best of your opportunities. (4) Hay grows best in summer. (5) It is easier to work in the sun than in the shade.

N=48. (1) 23, (2) 8, (4) 42, (5) 27.

(a) As can be seen from the above percentages, the literal interpretation of the saying had a strong attraction, and this was borne out by the fact that of the 8 errors made in which the generality of the saying was not appreciated, 5 were on (4) ranging from B. 115/117 to B. 87/94, the explanations given being similar, and quite reasonable, apart from the failure to understand "saying".

(b) Two examples of well supported errors demonstrate the reasonable arguments that may lie behind a wrong response - e.g.:-

B. 121/125 (1) explained that, "a saying is a sort of 'joking tip' - the answer isn't (2), and I don't think (3) means the same because you have an opportunity to make hay all the time, but it's best in the sun. Not (4) because it's only harvested in summer, and (5) is not true."

G. 119/119 (2) - "...because if you make the hay before it is ready the cattle will not eat it."

(c) Change to correct answer - e.g.:-

B. 103/99 (2) when asked to check all the others changed to (3) "because it's a better one, I didn't really see it the
first time."

G. 100/109 (1) changed to (3) - "...because to 'make the best' sounds important."

(d) Misreading occurred in one case only - e.g.:

B. 100/105 (1) had difficulty in explaining his choice, until it was found that he was reading - "Make hay while the sun rises."

Item 70
There is a saying, "All is not gold that glitters."
this means (?)
(1) Some gold has a dull finish. (2) Appearances are often deceptive. (3) Diamonds sparkle more than gold. (4) Don't wear cheap jewellery. (5) Some people like to make a show of wealth.

N=36. (1) 31, (3) 25, (4) 14, (5) 31.

It was clear from the reworking of this item, that these children did not grasp the significance of the saying fully, nor did they understand the statement - "appearances are often deceptive". Although the most popular choice was (5) - i.e. 12 of the 20 errors were on (5) - this choice was not made confidently, and when asked to check all the other statements, many of the children changed to one of the other distractors.

In several cases statement (2) was considered, and either accepted or rejected - e.g.:

B. 124/123 (2) - his original choice was thus correct, but when asked to check through the others he changed to (5) "...because wealthy people have more money and like to show it off;" i.e. he had not really understood the saying.

B. 110/118 (5) when asked to check through the statements again, changed to (2) with a correct explanation and said that he had not really looked at (2) before. Similarly B. 100/112.
G. 111/106 (2) changed to (1) "...because appearances
doesn't just mean jewellery, but could refer to clothes,
and the saying refers to jewellery."

B. 110/113 (4), rejected (2) "...because a deceptive appear-
ance is only a shiny artificial appearance, so it must be
(4)."

G. 90/97 (2) did not know the meaning of "deceptive" and ex-
plained her choice thus - "Even though gold glitters you
don't have to take any notice of the appearance of it."

(J) Category (10) - Number Series.

Item 17.
One number is wrong in the following series. What should
that number be? (Just write the correct number in the
brackets).
5 10 15 20 25 30 35 39 45 50 (Ans. = 40).
N=62 (39) 97, (Others) 3.

Although it is clear from the percentage of 97, and
the structure of the item, that practically all the errors
were due to a failure to meet all the requirements of the
problem, the errors which were repeated (ranging from B.115/
110 to B. 76/80) gave further insight into the strength of
the set which obliterated the second half of the instructions.
In 7 cases the direction to reread the instructions carefully,
resulted in the correcting of the response to (40), but in
the other 10 cases two or more readings were required before
the second half of the instructions was noticed - e.g:-

B. 108/115 (39) when directed to read the instructions aloud
still could see nothing wrong with (39).

G. 110/103 (39) reread the instructions, but still gave (39).
When the instructions were read aloud to her, with emphasis,
she still could see nothing wrong (although she had demonstrat-
ed that the correct number was 40).

G. 91/96 (39) read instructions aloud and gave (39). When asked if 39 was the correct number she answered - "Yes - because 35-40." Not until she was asked directly whether she should write (39) in the brackets, did she change to 40. Similarly with 7 other children.

Item 21.
One number is wrong in the following series. What should that number be?

1 7 2 7 3 7 4 7 5 7 6 7 8 7 (Ans.=7).

N=54. (8) 46, (77) 22, (9) 17, (6) 7, (Others) 7,
Omissions 29 cases

There was an unusual number of omissions on this item, indicating that some children were put off by the format of the problem or by their experiences with item 17 (there were no omissions on 17). Although 77 is listed as an error above, it is a very slight error, of the same nature as (8). As in the previous item the few errors repeated resulted from failure to meet the specific requirements of the problem.

Item 47.
One number is wrong in the following series. What should that number be?

2 3 4 3 2 3 4 3 2 4 (Ans.=3)

N=68. (4) 44, (2) 25, (1) 13, (5) 12, (Others) 6,
Omissions 30 cases

(a) Failure to meet the specific requirements of the problem - e.g.:

B. 106/118 (4) when asked to reread immediately changed to (3).

(b) Inability to break away from a grouping of three, e.g.:
G. 107/112 (1) attempted to group the numbers 2 3 4, 3 2 1, i.e. reversing the direction of the second group.

In this and the 8 other cases of this type, the strong set to group in 3s was coupled with the ignoring of disturbing figures, e.g.:

G. 111/106 (3) - explained that her answer referred to the second 3 in the series, thus the series should read 2 3 4, 2 3 4. Although she was satisfied to ignore the remainder of the series, when asked whether her arrangement fitted in with all the figures she said that it did not, and began another rearrangement in 3s. Finally after many unsuccessful attempts, it was suggested that she try pairs, and almost straight away she gave the correct answer.

This "threeness" of the series, suggested by the first 3 figures - 2 3 4, is a good example of what Simmel (see p. 89 above) described as "divisibility" in his experiment - i.e. the attraction of an overlearned property of numbers.

(c) In a few cases the principle chosen involved the ignoring of other figures, although not the adoption of a triplet arrangement - e.g.:

B. 82/83 (3) - i.e. the 6th figure - 3 - should not be included, i.e. 2 3 4 3 2, 4 3 2 4.

G. 72/76 (4) - i.e. the 5th figure should be (4) - "...because there are the other 4s; - 2 3, 4 3 4 3 4 3, 2 4."

B. 76/80 (5) - i.e. the 4th figure should be (5) so that the series could read 2 3 4 5.

In the first two of these examples, the wrong response arose from an inadequate but quite complex attempt to structure the problem.
One number is wrong in the following series. What should that number be?

\[1 \ 2 \ 4 \ 8 \ 12 \ 32 \ 64\]  
(Ans. = 16).

\[N=28. \ (12) \ 4 \text{ cases}, \ (1) \ 4 \text{ cases}, \ (0) \ 4 \text{ cases}, \ (32) \ 3 \text{ cases}, \ (3) \ 3 \text{ cases} \ (24) \ 2 \text{ cases} \ (2) \ 2 \text{ cases} \ (6) \ 2 \text{ cases}, \ (\text{others}) \ 4 \text{ cases} \]

Omissions 12 cases.

Errors were well spread out on this item, although most of the repeated errors fell into the first three of the types described.

(a) Partial structuring of the problem, i.e. a principle which applied to part of the data was accepted without any check to see whether or not it fitted in with the rest of the problem - e.g:-

B. 118/122 (24) - "12-32 should be 12-24 because 1-2, 4-8, 32-64, so 12-24, and as 12 isn't half 32, change 32 to 24." This boy found a principle which required the insertion of 24 into the series, but in explaining this he was also conscious of the fact that he must change one of the numbers, so changed 32 to 24, and overlooked the fact that he now had no place for 64. When asked to read the problem again he changed to 16 with a correct explanation.

B. 100/105 (24) "...instead of 32 - you double each time." When asked to work the question aloud he saw this his original reading was incomplete, did not fit all the data, and changed to 16.

G. 117/118 (32) - "...it should be 60, because 4 + 4 = 8, 8 + 4 = 12, so the one before 64 must be 60, 60 + 4 = 64." She had found a principle (+4) which fitted some of the data without having checked against the total structure of the series.

G. 118/115 (24) - "...1, 2, 4, 8, 12, 24 - I saw 2, 4, 8 and that made me think of doubling, so \(2 \times 12 = 24\)." Until directed to read the series slowly aloud she did not see the juxtaposition of 8 and 12.

G. 108/114 (24) - "... 1 + 1 = 2, 4 + 4 = 8, 12 + 12 = 24."
She ignored the 64 and in subsequent attempts could not break away from the pattern she had chosen.

B. 86/96 (24) also said that he saw that each one was doubled, but had great difficulty in seeing the inconsistency between 8 and 12.

(b) Satisfaction with a justifiable, but inferior solution - e.g:-

G. 107/112 (1) - "...all the rest are even, so (1) must be the wrong number; should be (0)." Her satisfaction at finding a solution blocked further exploration of the situation - similarly with five others ranging from G.110/103 to G. 94/102.

(c) Failure to meet the specific requirements of the task was evident in this item as in others of the same type. In some of the examples given above, the children committed the double error of following a wrong or incomplete principle, and then expressing this answer in the wrong way. There were 4 cases in which the correct reasoning was followed, but (12) instead of (16) was entered in the brackets (ranging from B. 121/125 (12) to B. 108/115 (12) ).

(d) Loss of direction or inconsistency within the task - e.g:-

B. 113/118 (32) - "...is the wrong number because 1 + 4 = 4, 4 + 4 = 8, 8 + 4 = 12, but 8 + 12 does not equal 32, so 32 is wrong." This boy not only used a false statement (1+4=4) but after having established the principle of (+4), lost sight of it in attempting to relate 8, 12 and 32.

G. 96/96 (1), changed to (64) as the wrong number, ";...because it should be 52; 2 + 2 = 4, 4 + 4 = 8, 8 + 4 = 12, and 12 + 20 = 32, 32 + 20 = 52;" i.e. she established two inconsistent principles, did not attempt to relate them, and then gave her answer on the basis of one of them.

(e) There was one case of complete lack of confidence, which may have been paralleled by others who omitted the item - e.g:-
G. 106/108, would not even attempt the item after one brief glance and dismissed it with the comment - "I can't see any reason." When asked to read it aloud, however, and look for some connections, she worked through the problem thus - "Well 2 x 32 = 64, and 1 x 1 = 2, and 2 x 2 = 4, so 12 should be 16, answer 16!"

**Item 74**

One number is wrong in the following series. What should that number be?

```
1 2 5 6 9 10 13 14 16 18
```

(N = 18. (13) 4 cases, (5) 3 cases, (3) 3 cases, (12) 2 cases, (19) 1 case, (8) 1 case, (2) 1 case, (7) 1 case, (9) 1 case, (4) 1 case. Omissions 7 cases.

Note the spread of errors, as for item 67. Of the 23 cases in which there was either an error, or an inadequately supported correct response, 11 were of the latter type - e.g:-

G. 122/127 (17), saw pairs of adjacent numbers, 1-2, 5-6, etc., so concluded that the last pair should be 16-17, hence the 18 should be changed to 17, and 17 given as the answer.

This justification was inadequate, as it ignored the constant difference (of 3) between the pairs and made use of only one of the two principles involved. As there is no way of distinguishing between a 16 corrected to 17 and an 18 corrected to 17, it is not possible to judge how many of those giving 17 (in the total sample) had really worked the item correctly. The 11 who gave the correct answer for this inadequate reason ranged from B. 133/132 to B. 101/107.

The main type of error noted was similar to that described in the previous item - i.e:-

(a) Partial structuring of the problem, e.g:-

G. 126/124 (7) explained that "...6,7, +3 = 10 so the 9 should be changed to 7, because 1,2, +3 = 5; 6,7, +3 = 10."
She had found a principle which could account for part of the series but which did not fit in with the other numbers. When this was pointed out to her, she reread the problem and changed to (17) - "...because it goes in pairs, 16-17."

B. 110/113 (5), i.e. 5 should be 4 so that the series would go 1 + 1 = 2, 2 + 2 = 4, 4 + 2 = 6, 6 + 3 = 9, 9 + 1 = 10; the progression was thus +1+2+2+3+1, then at 13 the same progression started over again, i.e. 13 + 1 = 14, 14 + 2 = 16, 16 + 2 = 18. This rather complex principle accounted for most of the data, but was based on the ignoring of the gap between 10 and 13. When attempting to rework the item, this boy could not break away from the schema he had constructed.

B. 115/117 (5) - "...because it goes by 2s, i.e. 2 4 6." When asked to read the item aloud he saw that he had ignored most of the data.

B. 108/115 (7) - "...7 instead of 6 because 1 + 2 = 3. +2 = 5, +2 = 7." He ignored the remainder of the series.

G. 110/103 (12) - "...13 should be 12, because it goes 10, 12,14,16;" - i.e. he ignored the first part of the series.

G. 108/114 (3) provided a good illustration of the "piecemeal" approach to finding a principle - "...2 should be 3 because 2 is an even number and it should be an odd number because 1,3, 5,...oh...no, that isn't right...16 is wrong because it should be 15, it goes odd even, odd even..." When asked about the gap 15-18, she said - "Oh! - should be 17, because 16-17."

(b) Correct principle with incorrect application, e.g: -

G. 114/117 (15) - "...16 should be 15 because +3,+1,+3,+1." When she applied this principle and came to 16 she changed to 17 - i.e. she had made a slip in getting the right order of +3,+1.

B. 102/113 (13) - "...13 should be 12, because 1-2,...oh gee...thought it was 1-2, 2 figures, then 5-6, 2 figures, then 9-10, 2 figures, then 13-14, 2 figures, so should be 16 changed to 17," i.e. merely a lapse in applying the correct principle. Similarly B. 100/112 (12) and G. 106/108.

B. 110/118 (16) gave the correct explanation but overlooked the specific requirements - i.e. the correct number as the answer.
(K) **Category (11) - Arithmetical Reasoning.**

**Item 13.**

At sixpence each, how many pencils can be bought for 48 pence?  (Ans. = 8).

N=30.  (24) 43, (6) 23, (28) 10, (48) 10, (Others) 14,  
Omissions 15 cases.

Of the 15 omissions, 9 were concentrated in group 7; presumably the item looked very similar to problems in arithmetic which many group 7 children would still be struggling with in schoolwork.

There were only 6 errors in the reworking, 2 due to confusion as to how to tackle the problem (B. 77/79, G. 84/88), 1 due to wrong division (G. 72/76 (9)), 1 due to irrelevant procedure (G. 86/88 (28) - "2 sixpences in one shilling, 2 into 48 goes 28 times, answer 28), and 2 due to misreading, e.g.:

B. 102/113 (£1-4-0) - i.e. 48 x 6d - he read the question as "at 6d each how much would 48 pencils cost?"

G. 97/105 (248) - i.e. 48 x 6d = 248d, a similar misreading.

Although there are no examples here of 24 as an answer (43% of the total errors) it is likely that this answer was due to a misreading of 48/- for 48 pence.

**Item 39.**

Count each 5 below that has a 7 next after it. Tell how many 5s you count.

7 5 3 5 7 2 3 7 5 6 7 7 2 5 7 3 4 7 7 5 2 0 7 5 7 3 3 7  
2 5 1 7 9 6 5 7  
(Ans. = 4)

N=108.  (8) 41, (3) 23, (7) 15, (5) 6, (2) 6, (Others) 6.  
Omissions 5 cases

This item was included in category (11) because of its
numerical format, although it is different from the other four, and has something in common with item 68 of category (8). Errors resulted from:-

(a) Miscounting. - Only 3 of the 22 errors were directly due to miscounting - (B. 118/122 (2), G. 90/97 (5), G. 83/83 (3)) - i.e. establishing the correct procedure but then misapplying it.

(b) Misreading. - The remaining errors were accounted for by various types of misreading, as illustrated in the following examples:-

B. 129/142, interpreted "next after" as "next but one after," although he corrected himself after some hesitation.

G. 106/108 (7) - i.e. she counted each 5 with a 7 either side of it; similarly 9 others ranging from G. 100/109 to B. 72/79.

G. 100/109 (7), not only misread (as above) but when asked to read the item aloud, read "... that has a 7 next to it."

B. 101/109 (5), read the instructions as "7 before it." and not only misread, but miscounted; similarly 3 others ranging from B. 101/109 (5) to G. 72/76 (3).

B. 95/103 (8), merely counted the 5s and ignored part of the instructions. When asked to reread the item, he read "... a 7 before it." Similarly G. 100/95 (8).

B. 91/96 (1), looked for a 5 with a 7 each side of it.

G. 93/89 (7), in explaining her choice pointed to 1, i.e. the 6th figure from the end, and said that should be a 7. She was attempting to find the correct number of a series - Cf. item 21 (number series) where 7 is recurrent as in this item.

(c) In two cases the correct answer was given for the wrong reason - e.g:-
B. 82/83 (4), counted every 7 with a 5 after it, and there happen to be 4 of these, just as there are 4 in the correct order of 5-7; similarly G. 84/88 (4).

Item 56.

If a man has walked east from his home 7 miles and then walked west 4 miles, how many miles is he from home? (Ans. = 3).

N=74. (11) 73, (7) 3, (23) 5, (Others) 14.
Omissions 5 cases.

(a) Failure to clarify the requirements of the problem - i.e. the item was attempted as if it read, "...east from his home 7 miles and west from his home 4 miles...". Coupled with this was the ignoring of the last part of the problem, "...how many miles is he from home?" and the substitution of the question - "how many miles has he walked?".

Of the 27 errors, 20 were of this type, with variations in the answers depending on the extent to which the false direction was followed up, e.g:-

G. 112/114 (11), in her explanation she drew a line representing 7 miles to the E, came right back across her starting point and drew another line from that point to the W, then explained that he had gone 11 miles.

G. 108/114 (22), i.e. 7 miles to the east, 7 miles back home; 4 miles to the west, 4 miles back home; 22 miles in all.

B. 115/117 (11), said that he had just added the figures, but when asked if he had noticed the words "from home" he worked it out again and changed to 18 - i.e. 7 miles E, 7 miles back to home, 4 miles W, so 18 miles altogether. B. 110/118 who went through exactly the same procedure verbalized his working and said - "...then how far has he walked?"

G. 100/104 (15) - "...11 miles altogether, no...15, because it is 4 miles back again to home."

(b) Wrong answer with correct explanation, i.e. 6 children when asked to explain their answer corrected it spontaneously
when working it aloud - e.g:-

B. 97/104 (11) "...well, he walks 7 miles E from home, then 4 miles W from the 7 miles...then he will be 3 miles from his home."

B. 87/94 (7), when asked to read again, changed to 3, with a correct explanation. In his first reading he had seen the words - "from his home 7 miles," and so just put down 7 without working the problem any further.

B. 102/100 (7), drew a diagram to explain his working, with E and W at right angles, but as soon as he began to explain this he saw his mistake and corrected it.

(c) Wrong directions for E and W, e.g:-

G. 119/119 (7), drew a diagram with E and W at right angles and then said that the end of the 4 miles W line would be about 7 miles from home.

(d) No attempt to analyse the question - e.g:-

G. 101/100 (28), multiplied 7 x 4 "...because I thought they were a long way away."

G. 83/83 (28), also multiplied 7x4 and when questioned showed that she had little idea what the question was about.

(e) Misreading, and lack of confidence, e.g:-

B. 118/122, asked, "Does it mean distance from home by the way he came, or straight across?" From his diagram it appeared that he had mistaken the direction of W, but when asked to reread he gave the correct answer and explained that the class was doing a problem in school like this one in which the direction was S, and he had read W as S.

G. 113/125, said she could not do this sort of problem, and wanted to leave it, yet when urged to try, she gave the correct answer with a correct explanation.

(f) Right answer for the wrong reason, e.g:-

B. 108/110 (3), "I just said 4 from 7 = 3," but when asked to explain this he drew the wrong directions and changed his answer to 4.

G. 93/89 (3) drew a diagram showing a line to the E, and one
to the S.W. and estimated that it would be about 3 miles across to home.

**Item 71.**

If I have a large box with 2 small boxes in it and 5 very small boxes in each small box, how many boxes are there in all? (Ans. = 13).


There were several typical procedures used in arriving at the wrong answer, e.g.: forgetting to add in the large box (Ans. 12); forgetting to add in the 2 small boxes (Ans. 11); failing to add in the 5 boxes twice (Ans. 8); multiplying the 5 boxes by another 5, or 4 (Ans. 28 or 23); adding an extra 5 boxes for the large box and then subtracting the large box (Ans. 14); multiplying 5 by 2, then again by 5 (Ans. 53).

Almost all of these errors could be classified in one group - i.e.:

(a) Failure to clarify the requirements of the problem due to a hasty reading, or to the prepotency of some elements at the expense of others, (20 out of 24 errors) e.g.:

G. 117/121 (11) - "1 large and 2 small and 5 in each of the 2 small boxes," i.e. in concentrating on the "5 in each of the 2 small boxes" she overlooked the counting in of the 2 small boxes.

B. 126/128, merely gained a vague idea of the requirements of the problem, multiplied 5 x 2 - a correct step, but then multiplied this again by 5 and added in the 3 boxes. On rereading he gave the correct answer.

G. 115/116 (14), "2+1 = 3, 5 very small boxes in each, so 3x15 =15, now take away the large box, so 14." When asked to reread the question she gave the correct answer.
B. 115/110 (8) explained - "1 large and 2 small = 3, and 5 very small boxes in each small box = 8." When asked to reread the question he said - "Gosh! I didn't read it through, I thought 5 small boxes, but it's 5 in each small box."

B. 110/113 (15) illustrates the difficulty involved in keeping all the elements of the problem in mind - "Well, 1 box, 2 boxes = 3 with 5 very small boxes in each box, so 15." When asked to reread the item he said - "It says that there's a large and two small and you have five boxes in the small boxes and the big box." When asked to check whether there were 5 in the big box, he went on - "No, I didn't notice the 'small' so its 11: 2 sets of 5, that's 10, then...13, I see now...I wasn't thinking about the 2 small boxes."

G. 108/114 (8) - "1 large and 2 small and 5 very small...oh! it's 13."

G. 109/112 (8) - "1 large and 2 small = 3, and inside the 2 small boxes there are 5 very small boxes, so it's 8." After rereading she said - "Oh!...I see, there are really 5 in each box, so 13."

B. 102/113 (11) "1 and 2 small boxes = 3 and 5 in each small box so that's 5 in each = 10...oh! - change my views, it's 13, I didn't add in the 2 little boxes.

All these examples are typical of this type of error, where a rereading or an attempt to explain led in most cases to the production of the correct answer.

(b) Lapse in applying the correctly formulated procedure - e.g:-

G. 117/118 (14), gave the correct explanation but had made a slip in her adding up. Similarly B. 95/96 (12) and G. 90/97 (12).

(c) Misinterpretation of the wording - one case only - i.e:-

B. 93/96 (12) asked "Do you count the big one?...I didn't count the big one because it says 'in all' and the large box is not in one."

Item 72.

If a boy can run 250 ft. in 10 seconds how many feet can he run in one-fifth of a second? (Ans. = 5).

N=24. (50) 42, (125) 8, (10) 8, (2) 8, (150) 8, (others)20 (Omissions 14 cases)
There were relatively more omissions on this item than on the others in the category - possibly due to a genuine inability to make sense of the problem, or to lack of confidence as a result of experiences in school arithmetic.

It was clear that for many children this item was not within their range of experience, and the first two types of errors described below, illustrate two ways of reacting to the situation - i.e.:-

(a) Failure to structure the problem at all, e.g.:

G.130/125 could not see where to start, and the problem had to be broken up into parts for her before she could see what to do. Five other children, ranging from G. 117/121 to G. 90/97 were also unable to tackle the problem.

(b) Partial structuring through "difficulty reduction", i.e. the problem was reshaped into the simple form "250 divided by 5" by ignoring "10 sec" and treating "one-fifth of a second" as "five seconds" - e.g.:

G. 108/114 (50) was satisfied with her procedure, until she admitted its inadequacy when questioned. Similarly 7 others ranging from B. 108/115 to G. 96/96.

(c) Correct procedure but incorrect application - e.g.:

G. 122/127 (3) gave the correct explanation but made a simple error in dividing - which she corrected spontaneously (i.e. 250 divided by 50 = 3). Similarly 4 others ranging from B. 128/128 (20; i.e. 250 divided by 50 = 20) to B. 115/110 (10; i.e. 250 divided by 50 = 10).

(d) Attempted solution with confused structure, e.g.:

B. 113/118 (50) - "Well 1/5 of a second = 5 into 250 = 50, because 5/5ths in...no!...can't be right because there are more than 5/5ths in 10; 20 into 250 = 12 1/2 because there are 20 1/5ths of a second in 10 seconds." This boy was quite confused about the properties of 1/5th.
G. 110/123 (125) - "5 into 25 = 2; 250 divided by 2 = 125, and this is the answer because a fifth of 10 seconds is 2 seconds," i.e. she took the 1/5th out of context.

B. 100/112 (25) - "Cross off the 0 and that will give you a fifth because when you multiply by 10 you just add an 0." When asked if he was sure about this he changed his ground - "1/5th of 10 seconds is 25 so divide 25 into 250 = 10." This boy was also confused about "1/5th".

G. 111/119 (125) - "250 divided by 2 because 1/5th is 5 seconds = \(\frac{1}{5}\) of 10." When asked to read the question again she said - "Well... divide by 50 because 1/5th x 10 = 50 fifths, so the answer is 3." When asked to check this she changed to 5.

B. 83/91 (2ft. 10 in.) argued - "Divide 250 by 10 = 25, then you want 12 into 25, and it's 2 ft. 10 inches" - i.e. he was taking 1/5th, as 1/5th of a minute = 12 seconds, so divide by 12.

These examples have been given in some detail because they demonstrate the attempt to extract meaning from a partially understood task.

(e) Misreading - one case only, i.e.:-

B. 133/132 (5) when asked to reread said "Oh! I read 1 second for 10 seconds."

(f) Correct answer for wrong reason - one case only, i.e.:-

G. 107/112 (5) "...because 5 into 25 is 5, and 1 into 5 is 5." When questioned it was apparent that she had not really grasped the correct structure of the task but was reading 250 as 25.
APPENDIX I.

A SUMMARY OF SPECIFIC WEAKNESSES OF OTIS ITEMS, WITH SOME SUGGESTIONS FOR IMPROVEMENT.

Since the completion of the investigation, revised forms of the Otis intermediate test have become available. Although some items have been dropped (e.g. the "foreign language" items) and others have been amended, the revised forms are of the same general nature as the original forms, and contain many of the same items. Improvements in the lay-out of the items have been made, e.g. better spacing, and the use of capitals and heavy type for emphasis or clarity. This should result in the reduction of some types of errors, yet in the light of the preceding records, it is clear that further improvements could be made. The following comments are based on one of the revised forms (CD), and are intended as illustrations of the ways in which individual error analyses may provide evidence, which can be used in the improvement of item construction.

(a) Analogies:— Such items as "CLOTHES are to a MAN as what are to a BIRD" (item 1) could be made more specific by some such wording as "CLOTHES are related to (or are useful to, or have something to do with) a MAN in just the same way as ___ are related to (or useful etc...) a BIRD." This type of wording brings out more clearly the essential structure of the item, a structure often overlooked by children who are capable of handling the item, when they understand what is really required.

(b) Definitions:— The wording "which tells best" may be misleading, as children are sometimes uncertain of what point
of view to take, e.g. item 32.

Which tells best what a WHEEL is?
(1) Something that turns.
(2) It goes around.
(3) A circular rim and hub connected by spokes.
(4) A round thing to put on a motor car.
(5) A bicycle always has two of them.

A child may argue that not all wheels have spokes, so that (3) must be rejected; or that "rim" and "hub" may be unfamiliar to a person, whereas a motor car with its "round things", or a bicycle, is more likely to be familiar. This sort of reasoning depends on an interpretation of "tells best" as "would give a person the best idea about."

Another example of confusion which could result from such an interpretation, is provided by item 2:—

Which one of the five words below tells best what a GUN is?
(1) shoot, (2) a weapon, (3) a tool, (4) kills, (5) a thing.

Distractors (1) and (4) are ruled out as grammatically incorrect (a weak use of distractors), yet a child may be prepared to overlook this and choose (1) because "a weapon" may refer to a wide range of objects, whereas "shoot" does at least refer to the most important characteristic of a gun, and this could be a reasonable interpretation of "tells best".

That some children are likely to make such interpretations as those illustrated, is borne out by the responses given to similar items in the original from B.

(c) Number series:— In order to eliminate errors due to the writing of the wrong, instead of the correct number, the questions could be reworded—"One number is wrong in the following series. What should the CORRECT number be?"
Item 67 (i.e. 3 4 6 7 9 10 12 14) should be reformulated in order to avoid the possibility of giving "13" as the correct answer, based on a pairing 9-10, 13-14, instead of 12-13.

(d) Emphasis on certain parts of the item:— Instead of having heavy type for "one"—e.g. "Which one of the words... etc...", the emphasis should be on the crucial words, e.g. "tells best", "most important reasons", "first", "last", "most like", as these are the parts of the instructions often overlooked.

On item 41—"If a man has walked north from his home 11 miles...
and then walked south 6 miles, how many miles is he from home?" - the words "is he from home?" should be emphasised; also the words "and then turned and walked south etc." could be used instead of "and then walked", as many children interpret the question as referring to two separate journeys, each starting from home.

(e) Question marks: In items of the type "If Harry is taller...etc." the question mark should be made more distinctive, and should not come near the end of a line, as many children assume that they have to read no further, e.g. item 36 has the line ending in (?) with "Harry" (to which the question mark refers) at the beginning of the next line. Instead of (?), the form ? could be used, not only in these items, but also in "analogs," where the question mark occurs in the middle of a sentence.

(f) Disarranged sentences: In order to reduce the confusion resulting from the unfamiliarity of disarranged sentences, and the lengthy instructions (so often partially overlooked), items of this type could be reworded - "Rearrange the following words to make a good sentence, and then write down the letter with which the second word of the sentence begins." An even clearer arrangement of the second half of the sentence would be - "...write down the beginning letter of the second word of the sentence." (In order to avoid confusion of "first" and "second").

Item 57 reads - "Do what this mixed-up sentence tells you to do: alphabet the letter Write twenty-third the of..." There would seem to be little justification for introducing the hazard of a simple miscount, into a problem which may, in all other respects, be correctly answered.

(g) Unfamiliar words: As indicated for several items in the Otis B records, it is difficult to predict what will happen when children meet unfamiliar words. Unless the item is included as a test of the particular word given, then a word which could be unfamiliar to children, merely through a specific lack of experience, may alter or distort the whole point of the item. Examples in form CD are "institution" (item 43), "communication" (item 27), "domestic" (item 30) - although of these, the only evidence available from form B refers to "Institution".

Two general comments arise from experience with form B; namely:-
(a) The examples on the front page should be more representative, and much more time should be given to explaining the types of items, and what is required.

In form CD there are four sample questions:- Question (1) uses a form of wording not used in the test, i.e. "Which one of the five words below tells what an APPLE is?" Although this is doubtless meant as a practice item in order to ensure familiarity with multiple choice questions, it may predispose children to read "tells best" as merely "tells" in subsequent items in the test proper (e.g.: the second item is of this type). Question (2) is a type of item not used in the test. Question (3) is an analogy type item with a slightly different lay-out from those included in the test. Question (4), although very similar to a question given early in the test (item 7), uses a type of answer not used elsewhere - e.g. "...( ) pence." In view of this example, the words of the very similar type of question given in item 7 should be emphasised, in order to avoid confusion - i.e. "At sixpence each, how many notebooks ..... etc."

It cannot be assumed that such a brief introduction is adequate, or that the giving of examples slightly different from the standard form of the items in the test, will not lead to at least some confusion. It is suggested that:-

i. There should be examples of the items most likely to be misinterpreted - e.g.: definitions, disarranged sentences, analogies, number series. As there are six "proverbs" items in the test, and as the records from form B indicate that many children were unfamiliar with proverbs, and did not appreciate the significance of "saying", an illustration of a "saying" could be included in the examples.

ii. There should be specific warnings about the necessity for reading all the instructions, looking at all the choices, checking the answer to make sure that it is given in the exact form required.

(b) It may be fruitful to experiment with a form of the test in which plenty of room is left, around or after each item, or on the opposite page, for rough working. There is no
evidence to show that tests of the "mental arithmetic" type, are better than those where children are expected to use rough working as they go along. Children could, in fact, be encouraged to jot down any working or draw diagrams, if they wished to. This might result in a reduction of some of the superficial errors, which spoil an otherwise well reasoned answer. Apart from the fact that this is a common procedure in written examinations of the "problem" type, it could also be of value in providing useful clues to the child's methods of work.
OTIS SELF-ADMINISTERING TEST

Adapted by
NEW ZEALAND COUNCIL FOR EDUCATIONAL RESEARCH
INTERMEDIATE EXAMINATION : FORM B.
For Standard 3 to Form 2

Read this page. Do what it tells you to do.

Do not open this paper, or turn it over, until you are told to do so. Fill these blanks, giving your name, age, etc. Write plainly.

Name.............................................................................................................. Age__________________ years________________ months

Boy or Girl........................................ Father’s Occupation...........................................

Class or Standard......................... School................................................................. Date.........................19

This is a test to see how well you can think. It contains questions of different kinds. Here is a sample question already answered correctly. Notice how the question is answered:

Sample: Which one of the five words below tells what an apple is?
     1 flower, 2 tree, 3 vegetable, 4 fruit, 5 animal

The right answer, of course, is “fruit”; so the word “fruit” is underlined. And the word “fruit” is No. 4; so a figure 4 is placed in the brackets at the end of the dotted line. This is the way you are to answer the questions.

Try this sample question yourself. Do not write the answer; just draw a line under it and then put its number in the brackets:

Sample: Which one of the five things below is round?
     1 a book, 2 a brick, 3 a ball, 4 a house, 5 a box

The answer, of course, is “a ball”; so you should have drawn a line under the words “a ball” and put a figure 3 in the brackets. Try this one:

Sample: A foot is to a man and a paw is to a cat the same as a hoof is to a—what?
     1 dog, 2 horse, 3 shoe, 4 blacksmith, 5 saddle

The answer, of course, is “horse”; so you should have drawn a line under the word “horse” and put a figure 2 in the brackets. Try this one:

Sample: At four pence each, how many pence will 6 pencils cost? 

The answer, of course, is 24, and there is nothing to underline; so just put the 24 in the brackets.

If the answer to any question is a number or a letter, put the number or letter in the brackets without underlining anything. Make all letters like printed capitals.

The test contains 75 questions. You are not expected to be able to answer all of them, but do the best you can. You will be allowed half an hour after the examiner tells you to begin. Try to get as many right as possible. Be careful not to go so fast that you make mistakes. Do not spend too much time on any one question. No questions about the test will be answered by the examiner after the test begins. Lay your pencil down.

Do not turn this page until you are told to begin.

By arrangement with Australasian Publishing Co., Agents in Australia for Geo. Harrap & Co. (Copyrighted.)
1. Which one of the five things below does not belong with the others?
   1 peach, 2 banana, 3 orange, 4 baseball, 5 plum

2. Which one of the five words below tells best what a hammer is?
   1 thing, 2 tool, 3 furniture, 4 weapon, 5 machine

3. Which one of the five words below means the opposite of east?
   1 north, 2 pole, 3 west, 4 equator, 5 south

4. The peeling is to a banana and the husk to an ear of corn the same as a shell is to what?
   1 an apple, 2 an egg, 3 juice, 4 a peach, 5 a hen

5. If we feel sorry for the suffering of another, we have a feeling of (?)
   1 hate, 2 pity, 3 contempt, 4 disdain, 5 eagerness

6. Which one of the five things below is the largest?
   1 bud, 2 branch, 3 tree, 4 twig, 5 limb

7. Wool is to a sheep as feathers are to what?
   1 a pillow, 2 a rabbit, 3 a bird, 4 a goat, 5 a bed

8. Which word means the opposite of succeed?
   1 win, 2 decline, 3 fail, 4 accede, 5 try

9. Which one of the five things below is most like these three: apple, peach, pear?
   1 seed, 2 tree, 3 plum, 4 juice, 5 peel

10. Which one of the ten numbers below is the largest? (Tell by letter.)
    A 6456, B 8968, C 4265, D 5061, E 4108, F 7549, G 2335, H 9472, J 3286, K 8970

11. Hat is to head as thimble is to what?
    1 finger, 2 needle, 3 thread, 4 hand, 5 sewing

12. If the words below were rearranged to make a good sentence, with what letter would the last word of the sentence begin? (Make the letter like a printed capital.)
    Usually are of made tables wood

13. At sixpence each, how many pencils can be bought for 48 pence?

14. Which statement tells best just what a gate is?
    1 A hole in a fence, 2 Something to swing on, 3 It has hinges, 4 A door in a fence, 5 It opens and shuts

15. A hand is to an arm the same as a foot is to what?
    1 leg, 2 toe, 3 finger, 4 wrist, 5 knee

16. When a new kind of machine is thought of, it is usually called (?)
    1 a discovery, 2 an adoption, 3 a creation, 4 a novelty, 5 an invention

17. One number is wrong in the following series. What should that number be? (Just write the correct number in the brackets.)
    5 10 15 20 25 30 35 39 45 50

18. What is the most important reason that motor cars have displaced horses and carriages?
    1 Horses were getting scarce, 2 Horses often run away, 3 Motors save time and “time is money,” 4 Motors are cheaper than carriages, 5 Motors cost less to repair than carriages

19. Coal is to a railway engine as what is to a motor car?
    1 motor cycle, 2 smoke, 3 wheels, 4 petrol, 5 horn

20. Which one of the words below would come first in the dictionary?
    1 tramp, 2 saint, 3 razor, 4 quart, 5 grass

21. One number is wrong in the following series. What should that number be?
    1 2 3 4 5 6 7 8 9 7

22. A motor car is to a wagon as a motor cycle is to what?
    1 walking, 2 horse, 3 buggy, 4 train, 5 bicycle

23. A boy who often tells big stories about what he can do is said to ( ?)
    1 lie, 2 fake, 3 cheat, 4 joke, 5 brag

24. Which one of the five words below means the opposite of difficult?
    1 hard, 2 quick, 3 soft, 4 easy, 5 common

25. Which one of the five things below is most like these three: snake, cow, sparrow?
    1 tree, 2 doll, 3 pig, 4 feather, 5 skin

26. A hospital is to the sick as what is to criminals?
    1 doctor, 2 asylum, 3 judge, 4 prison, 5 sentence

27. Which tells best just what a horse is?
    1 It has a tail, 2 a live thing, 3 a thing that works and eats, 4 a large, four-legged animal, 5 something to pull a wagon

Do not stop. Go on with the next page.
28. Do what this mixed-up sentence tells you to do. 
letter Write the the in A brackets ......................................................... ( )
29. Which one of the words below would come first in the dictionary? 
1 brave, 2 burst, 3 broke, 4 build, 5 breadth, 6 brown, 7 bunch, 8 bribe ( )
30. Better is to good as worse is to what? 
1 very good, 2 medium, 3 bad, 4 much worse, 5 best ...........................................
31. Which tells best just what a lamb is? 
1 an animal with wool, 2 a creature with four legs and a tail, 3 a lively small animal, 
4 a young sheep, 5 a young animal that eats grass ...........................................
32. If the words below were rearranged to make a good sentence, with what letter would the third word of the sentence begin? (Make the letter like a printed capital.) 
honey bees clover gather red from ............................................................
33. There is a saying, "A stitch in time saves nine." This means (?)  
1 A little sewing may save nine shillings. 2 It pays to attend to troubles before they get worse. 
3 Work hard and save as much as you can. 4 You can save time by sewing .... ( )
34. Grass is to cattle as bread is to what? 
1 butter, 2 flour, 3 milk, 4 man, 5 horses ...........................................................
35. Which tells best just what a lie is? 
1 a mistake, 2 a malicious false statement, 3 an accidental false statement, 4 an exaggeration, 
5 a wrong answer .................................................................................
36. The son of my father's sister is my (?)  
1 brother, 2 nephew, 3 cousin, 4 uncle, 5 grandson ...........................................
37. If George is taller than Frank, and Frank is taller than James, then George is (?) James. 
1 taller than, 2 shorter than, 3 just as tall as, 4 (cannot say which) ...........
38. A king is to a kingdom as a president is to what? 
1 vice-president, 2 senate, 3 republic, 4 queen, 5 democrat ...................................
39. Count each 5 below that has a 7 next after it. Tell how many 5's you count. 
7 5 3 5 7 2 3 7 5 6 7 7 2 5 7 3 4 7 7 5 2 0 7 5 7 8 3 7 2 5 1 7 9 6 5 7 .................
40. An event which is sure to happen is said to be (?)  
1 probable, 2 certain, 3 doubtful, 4 possible, 5 delayed .................................
41. Which one of the five things below is most like these three: president, admiral, general? 
1 ship, 2 army, 3 king, 4 republic, 5 soldier .......................................
42. Large is to object as loud is to what? 
1 soft, 2 small, 3 heavy, 4 weight, 5 sound .............................................
43. If the following words were arranged in order, with what letter would the middle word begin? 
Eight Ten Six Nine Seven ........................................................................
44. A quantity which grows smaller is said to (?)  
1 fade, 2 decrease, 3 dry up, 4 die, 5 sink ...........................................
45. In a foreign language, boy = Kolo Daak  
good boy = Kolo Daak  
The word that means good begins with what letter .................................
46. A captain is to a ship as a mayor is to what? 
1 state, 2 council, 3 city, 4 boss, 5 lawyer ................................................
47. One number is wrong in the following series. What should that number be? 
2 3 4 3 2 3 4 3 2 4 ........................................................................
48. If Harry is older than William and William is just as old as Charles, then Charles is (?) Harry. 
1 older than, 2 younger than, 3 just as old as, 4 (cannot say which) ...........
49. Do what this mixed-up sentence tells you to do. 
sentence the letter Write first this in .................................................
50. A revolver is to a man as what is to a bee? 
1 wings, 2 honey, 3 flying, 4 wax, 5 sting .............................................
51. If Paul is older than Herbert and Paul is younger than Robert, then Robert is (?) Herbert. 
1 older than, 2 younger than, 3 just as old as, 4 (cannot say which) ...........
52. What is the most important reason that bright lights are placed in front of theatres? 
1 so that people can see where they are, 2 to attract attention and look inviting, 
3 so that people can see the advertisements better, 4 electricity is furnished to theatres cheaply, 
5 to help light up the street ........................................................................

Do not stop. Go on with the next page.
53. If the words below were rearranged to make a good sentence, with what letter would the third word of the sentence begin? (Make the letter like a printed capital.)

   boys, birch, the a canoe made bark

   ( )

54. A person who wishes very much to succeed but fears he will fail is said to be (?)

   1 earnest, 2 anxious, 3 industrious, 4 energetic, 5 cowardly

   ( )

55. If the following words were arranged in order, with what letter would the middle word begin?

   Week, Year, Hour, Second, Day, Month, Minute

   ( )

56. If a man has walked east from his home 7 miles and then walked west 4 miles, how many miles is he from home?

   ( )

57. In a foreign language very hot = Soto Gran

   very cold = Foss Gran

   The word that means very begins with what letter?

   ( )

58. Which one of the five things below is most like these three: cannon ball, wire, penny?

   1 pound note, 2 bone, 3 string, 4 pencil, 5 key

   ( )

59. There is a saying, “A drowning man will grasp at straws.” This means (?)

   1. A man will sink more easily than a straw. 2. Everyone should learn to swim. 3. Desperate people cling to absurd hopes. 4. Those who cannot swim should stay on land

   ( )

60. Do what this mixed-up sentence tells you to do.

   sum four Write three the one and of

   ( )

61. An object or institution that will not perish or cease is said to be (?)

   1 permanent, 2 stable, 3 stationary, 4 solid, 5 sound

   ( )

62. In a foreign language some food = Beko Prac

   some milk = Klup Prac

   some food and milk = Beko Otoh Klup Prac

   The word that means and begins with what letter?

   ( )

63. Which word means the opposite of pride?

   1 sorrow, 2 humility, 3 miserable, 4 conceit, 5 proud

   ( )

64. If the following words were arranged in order, with what letter would the middle word begin?

   General, Lieutenant, Private, Colonel, Sergeant

   ( )

65. There is a saying, “Make hay while the sun shines.” This means (?)

   1 Hay made in cloudy weather is poor. 2. Haste makes waste. 3. Make the best of your opportunities. 4. Hay grows best in summer. 5. It is easier to work in the sun than in the shade

   ( )

66. Which tells best just what a foot is?

   1 to wear a shoe and stocking on, 2. the part of the body on which an animal stands, 3. it has five toes and a heel, 4. both feet are the same size, 5. men have larger feet than women

   ( )

67. One number is wrong in the following series. What should that number be?

   1 2 4 8 12 32 64

   ( )

68. Write the letter that follows the letter that comes next after K in the alphabet

   ( )

69. If the following words were arranged in order, with what letter would the middle word begin?

   Youth, Infancy, Manhood, Childhood, Birth

   ( )

70. There is a saying, “All is not gold that glitters.” This means (?)

   1. Some gold has a dull finish. 2. Appearances are often deceptive. 3. Diamonds sparkle more than gold. 4. Don't wear cheap jewellery. 5. Some people like to make a show of wealth

   ( )

71. If I have a large box with 2 small boxes in it and 5 very small boxes in each small box, how many boxes are there in all?

   ( )

72. If a boy can run 250 feet in 10 seconds how many feet can he run in one-fifth of a second?

   ( )

73. Which one of the following words would come last in the dictionary?

   1 heart, 2 judge, 3 grass, 4 nerve, 5 horse, 6 north, 7 labour

   ( )

74. One number is wrong in the following series. What should that number be?

   1 2 5 6 9 10 13 14 16 18

   ( )

75. An agreement reached in which both sides yield somewhat in their demands is called (?)

   1 a promise, 2 a compromise, 3 an understanding, 4 a deadlock, 5 an armistice

   ( )

If you finish before the time is up, go back and make sure that every answer is right.