



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

THE GENERAL COGNITIVE ABILITY OF MAORI
AND EUROPEAN SUBJECTS. A COMPARATIVE STUDY
OF QUEENSLAND TEST PERFORMANCES AND
PSYCHOMETRIC PROPERTIES

A thesis
submitted in partial fulfilment
of the requirements for the Degree
Doctor of Philosophy in Psychology
at the
University of Waikato
by
ROSS ST. GEORGE

University of Waikato
1977

ABSTRACT

Research was undertaken into the performances of New Zealand Maori and European children on the Queensland Test of Cognitive Abilities. In addition to studying the comparative performances of the two ethnic groups at six age levels, major psychometric properties of the instrument were investigated. Such data have an important bearing upon the issue of psychological equivalence across ethnic or culture group samples.

Seven hundred children (303 Maori and 397 European) aged between eight and a half and fourteen and a half years were tested with the Queensland Test. A small statistically significant ethnic group difference was found favouring the European children. Subsequent analyses indicated that the small difference was of little practical significance. Queensland Test subtest data were also analysed. This indicated that subtest performances were likely to be influenced by subject variables in a manner not necessarily evident in the overall test results. Gender differences in performance on the Pattern Matching Subtest were found and are consistent with other data.

Queensland Test reliability, validity and item characteristics were researched for the two ethnic groups.

To obtain test-retest reliability estimates and data on criterion-related validity longitudinal studies were undertaken with four ethnically mixed class groups (Standard 3 to Form 2).

The Queensland Test was found to be equally reliable with both ethnic groups although the levels of reliability obtained were not considered entirely satisfactory.

Validity data indicated that the same general intelligence construct could be employed in the interpretation of Queensland Test results with

both European and Maori subjects. Content validity issues were considered and it was argued that in terms of both test content and measurement procedures the Queensland Test should provide an equally valid measure of general intelligence with both groups. The criterion-related validity coefficients were of approximately the same order for each ethnic group although doubts were expressed about the adequacy of the criteria. Construct validity was investigated by considering the relationship of test performance to age and through factor analytic evidence. The analyses supported the view that the Queensland Test was primarily a measure of the same underlying construct within each ethnic group. In terms of prevailing psychological theories on human intelligence this construct was interpreted as being of a general intelligence nature.

Comparative data on item characteristics did not indicate any marked ethnic group differences. More apparent was the need to restructure particular subtests where items failed to discriminate adequately within both ethnic groups. It was concluded that the weight of evidence was in favour of the psychological equivalence of this ability measure with Maori and European children.

Data for combined ethnic group samples have been reported. Pooled data were employed in some factor analyses and these supported a general intelligence interpretation of Queensland Test results. Preliminary norms were also derived from combined ethnic group data.

Findings are discussed in relation to the concept of intelligence and its cross-cultural measurement, other Queensland Test research and recent cross-cultural studies on cognitive abilities in New Zealand.

ACKNOWLEDGEMENTS

No single sweep of the pen can give adequate recognition or thanks to the many people who have assisted with this study. There are hundreds of young New Zealanders who served as subjects. There are the teachers, administrators, professors, lecturers and students - young and not so young - who all helped in many and different ways. Then again there is all the assistance borne of love and friendship.

The Maori Education Foundation generously made financial support available for this study over a number of years. More recently material support has come to me through Massey University and I am indebted to Professor Clem Hill for easing some of the pains of completion.

This thesis belongs, however, to the University of Waikato and to people connected with the study of psychology there. The origins of this study lay with Professor Jim Ritchie. Throughout, Jim has given freely of his time and wisdom. Many thanks Jim.

The assistance of Dr Geoff Ord during his time at the University of Waikato is also gratefully acknowledged. Geoff's work in Papua-New Guinea laid the foundations to the Queensland Test and his intimate knowledge of its development has been invaluable.

On a number of occasions assistance was sought from Dr George Kearney, University of Queensland, Drs Colin Boswell and Richard Harker, Massey University. My thanks are expressed for their help with both theoretical and statistical questions.

Good fortune brought Professor William McDougall, Washington State University, to these shores last year and in an act of friendship Mac critically read a draft of this thesis. His efforts were important in shaping this thesis.

Finally, there is my wife Alison and my thanks for all her help will continue to be expressed as the years go by.

TABLE OF CONTENTS

	Page
ABSTRACT	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	xviii
LIST OF ABBREVIATIONS	xx
INTRODUCTION	xxi
<u>CHAPTER 1. THE STUDY OF HUMAN INTELLIGENCE</u>	1
Introduction	1
Some Early Views on Intelligence and its Measurement	1
Intelligence: Comments on Definitions	13
Psychological Theories of Intelligence	19
The Cross-Cultural Study of Intelligence	36
Theoretical Viewpoints	37
Cross-Cultural Psychometric Studies: Notes on Developments and Issues	50
Race and Intelligence	71
Summary	74
<u>CHAPTER 2. MAORI EDUCATIONAL UNDERACHIEVEMENT</u>	77
Introduction	77
The Evidence on Underachievement	78
Cultural Deficit and Cultural Difference Views	83
The Involvement of Intelligence	86
Ethnicity or Social Class?	89
Summary	98
<u>CHAPTER 3. MAORI INTELLECTUAL ABILITIES</u>	99
Anecdote and Opinion	99
Comparative Studies on Levels of Intelligence	105
Comparative Studies on Intellectual Qualities	119
Comments on New Zealand Test Standardizations	123
Conclusions, Issues and Questions	125

<u>CHAPTER 4.</u>	<u>RESEARCH PROPOSALS AND METHOD</u>	135
	Development of Proposals	135
	The Proposals	141
	Subjects	142
	Ethnic Identification	142
	Additional Variables	149
	- Gender	149
	- Age	150
	- Urban/Rural Location	150
	- Socio-Economic Status	150
	Sample Selection	153
	Representativeness	155
	Procedures	157
	Supplementary Data	159
<u>CHAPTER 5.</u>	<u>THE QUEENSLAND TEST</u>	163
	Development	163
	Subtest Descriptions	173
	(a) Knox Cube Imitation Subtest	176
	(b) Beads Subtest	179
	(c) Passalong Subtest	182
	(d) Form Assembly Subtest	182
	(e) Pattern Matching Subtest	190
	Some General Notes	196
	Comments on Interpretation	199
	Notes on Research Applications	203
<u>CHAPTER 6.</u>	<u>RESULTS</u>	212
	PART 1 Q.T. SUMMARY STATISTICS, SCORE DISTRIBUTIONS AND INTERCORREL- ATION ANALYSIS	213
	(a) QT Total Score	213
	(b) QT Subtests	221
	(c) Summary Figures and Comments	230
	(d) Score Distributions	237
	(e) QT Subtest and QT Total Score Intercorrelations	254
	 PART 2 DATA ON QT RELIABILITY	261
	(a) Internal Consistency Estimates of Reliability	261
	(b) Test-Retest Estimates of Reliability	264
	(c) Estimates of Standard Error of Measurement	272

PART 3	DATA ON QT VALIDITY	276
	(a) Content Validity of the QT	276
	(b) Criterion-Related Validity of the QT	280
	(c) Construct Validity of the QT	306
PART 4	DATA ON QT ITEMS	330
	(a) Item Difficulty	330
	(b) Item Discrimination Indices	336
	(c) Item Correlations with Subtest Score and QT Total Score	338
	(d) Factor Analysis of Item Intercorrelations	343
PART 5	PRELIMINARY DECILE RANK NORMS	361
<u>CHAPTER 7.</u>	<u>DISCUSSION OF RESULTS AND CONCLUSIONS</u>	367
PART 1	SUMMARY AND DISCUSSION OF MAJOR FINDINGS	368
PART 2	THE RESULTS IN RELATION TO OTHER QT STUDIES	375
PART 3	GENERAL CONSIDERATIONS	387
<u>APPENDICES</u>		
APPENDIX A	Queensland Test Record Form.	401
APPENDIX B	QT Summary statistics and intercorrelation analyses for class group samples in test-retest reliability studies.	404
APPENDIX C	QT and criteria data summary statistics for class group samples in criterion-related validity studies.	428
APPENDIX D	Tables on the proportion of age group sample passing each item of the five QT subtests: Data presented for Total Age Group Samples and European and Maori Age Group samples.	435
<u>REFERENCES</u>		440

LIST OF TABLES

	Page
CHAPTER 3	
TABLE 3.1 - Summary of Empirical Studies Comparing European and Maori Performances on Measures of Intellectual Ability	126
CHAPTER 4	
TABLE 4.1 - Distribution of Total Sample by Ethnic Group	149
4.2 - Distribution by Gender of the Separate and Combined Ethnic Group Samples	149
4.3 - Distribution by Age and Gender of the Separate and Combined Ethnic Group Samples	151
4.4 - Percentage Distribution by Urban/Rural Location of the Separate and Combined Ethnic Group Samples	151
4.5 - Percentage Distribution of Socio-economic Status Levels of the Separate and Combined Ethnic Group Samples and Comparative Data from a National Sample	152
4.6 - Details on Numbers, Ethnic Composition, Class Level and Age of Class Groups Employed in the Queensland Test Reliability and Validity Studies	154
CHAPTER 5	
TABLE 5.1 - The Block Tapping Sequences Employed in the Knox Cube Imitation Subtest of the QT	178
5.2 - The Bead Sequences Employed in the Beads Subtest of the QT	181
5.3 - Order of Practice and Test Items of the Form Assembly Subtest of the QT	193
5.4 - The Number of Tiles Required and Time Limits for Each Item of the Pattern Matching Subtest of the QT	197

LIST OF TABLES (Contd.)

CHAPTER 6	Page
PART 1	
TABLE 6.1.1 - Means and Standard Deviations for QT Total Score by Age, Gender and Ethnic Group	214
6.1.2 - Summary of ANOVA for QT Total Score	214
6.1.3 - Means, Standard Deviations and Mean Score Differences for QT Total Score by Age and Ethnic Group	214
6.1.4 - Linear Regression Equations and Correlation Coefficients for QT Total Score on Age for European and Maori Samples	217
6.1.5 - Correlations Between QT Total Score and the Five Independent Variables Entered in the Multiple Regression Analysis (N=673)	219
6.1.6 - Multiple Regression Analysis - QT Total Score	219
6.1.7 - Means and Standard Deviations for Knox Subtest by Age, Gender and Ethnic Group	222
6.1.8 - Summary of ANOVA for Knox Subtest	222
6.1.9 - Means and Standard Deviations for Knox Subtest by Age for Separate and Combined Ethnic Group Samples	223
6.1.10 - Means and Standard Deviations for Beads Subtest by Age, Gender and Ethnic Group	223
6.1.11 - Summary of ANOVA for Beads Subtest	224
6.1.12 - Means and Standard Deviations for Beads Subtest by Age for Separate and Combined Ethnic Group Samples	224
6.1.13 - Means and Standard Deviations for Passalong Subtest by Age, Gender and Ethnic Group	225
6.1.14 - Summary of ANOVA for Passalong Subtest	226
6.1.15 - Means and Standard Deviations for Passalong Subtest by Age for Separate and Combined Ethnic Group Samples	226
6.1.16 - Means and Standard Deviations for Passalong Subtest by Gender and Ethnic Group	226

LIST OF TABLES (Contd.)

	Page
TABLE 6.1.17 - Means and Standard Deviations for Form Assembly Subtest by Age, Gender and Ethnic Group	227
6.1.18 - Summary of ANOVA for Form Assembly Subtest	228
6.1.19 - Means and Standard Deviations for Form Assembly Subtest by Age and Ethnic Group	228
6.1.20 - Means and Standard Deviations for Pattern Matching Subtest by Age, Gender and Ethnic Group	229
6.1.21 - Summary of ANOVA for Pattern Matching Subtest	230
6.1.22 - Means and Standard Deviations for Pattern Matching Subtest by Age and Ethnic Group	230
6.1.23 - Frequency Distribution of QT scores for European Sample by Numbers and Proportions (N=397)	242
6.1.24 - Frequency Distribution of QT scores for Maori Sample by Numbers and Proportions (N=303)	243
6.1.25 - Frequency Distribution of QT scores for Total Sample by Numbers and Proportions (N=700)	244
6.1.26 - Frequency Distributions of Knox Subtest Scores for European, Maori and Total Sample by Numbers and Proportions	251
6.1.27 - Frequency Distributions of Beads Subtest Scores for European, Maori and Total Sample by Numbers and Proportions	251
6.1.28 - Frequency Distributions of Passalong Subtest Scores for European, Maori and Total Sample by Numbers and Proportions	252
6.1.29 - Frequency Distributions of Form Assembly Subtest Scores for European, Maori and Total Sample by Numbers and Proportions	252
6.1.30 - Frequency Distributions of Pattern Matching Subtest Scores for European, Maori and Total Sample by Numbers and Proportions	253
6.1.31 - Product-Moment Correlation Coefficients for European Age Group 8:6 - 9:5 (N=36) Among QT Subtests and QT Total Score	255
6.1.32 - Product-Moment Correlation Coefficients for Maori Age Group 8:6 - 9:5 (N=37) Among QT Subtests and QT Total Score	255

LIST OF TABLES (Contd.)

	Page
TABLE 6.1.33 - Product-Moment Correlation Coefficients for Age Group 8:6 - 9:5 (N=73) Among QT Subtests and QT Total Score	255
6.1.34 - Product-Moment Correlation Coefficients for European Age Group 9:6 - 10:5 (N=57) Among QT Subtests and QT Total Score	256
6.1.35 - Product-Moment Correlation Coefficients for Maori Age Group 9:6 - 10:5 (N=33) Among QT Subtests and QT Total Score	256
6.1.36 - Product-Moment Correlation Coefficients for Age Group 9:6 - 10:5 (N=90) Among QT Subtests and QT Total Score	256
6.1.37 - Product-Moment Correlation Coefficients for European Age Group 10:6 - 11:5 (N=125) Among QT Subtests and QT Total Score	257
6.1.38 - Product-Moment Correlation Coefficients for Maori Age Group 10:6 - 11:5 (N=69) Among QT Subtests and QT Total Score	257
6.1.39 - Product-Moment Correlation Coefficients for Age Group 10:6 - 11:5 (N=194) Among QT Subtests and QT Total Score	257
6.1.40 - Product-Moment Correlation Coefficients for European Age Group 11:6 - 12:5 (N=63) Among QT Subtests and QT Total Score	258
6.1.41 - Product-Moment Correlation Coefficients for Maori Age Group 11:6 - 12:5 (N=59) Among QT Subtests and QT Total Score	258
6.1.42 - Product-Moment Correlation Coefficients for Age Group 11:6 - 12:5 (N=122) Among QT Subtests and QT Total Score	258
6.1.43 - Product-Moment Correlation Coefficients for European Age Group 12:6 - 13:5 (N=78) Among QT Subtests and QT Total Score	259
6.1.44 - Product-Moment Correlation Coefficients for Maori Age Group 12:6 - 13:5 (N=60) Among QT Subtests and QT Total Score	259
6.1.45 - Product-Moment Correlation Coefficients for Age Group 12:6 - 13:5 (N=138) Among QT Subtests and QT Total Score	259
6.1.46 - Product-Moment Correlation Coefficients for European Age Group 13:6 - 14:5 (N=38) Among QT Subtests and QT Total Score	260

LIST OF TABLES (Contd.)

	Page
TABLE 6.1.47 - Product-Moment Correlation Coefficients for Maori Age Group 13:6 - 14:5 (N=45) Among QT Subtests and QT Total Score	260
6.1.48 - Product-Moment Correlation Coefficients for Age Group 13:6 - 14:5 (N=83) Among QT Subtests and QT Total Score	260
PART 2	
TABLE 6.2.1 - K-R 20 Reliability Coefficients for QT by Age Groups and Separate and Combined Ethnic Group Samples	262
6.2.2 - K-R 20 Reliability Coefficients for Knox Subtest by Age Groups and Separate and Combined Ethnic Group Samples	262
6.2.3 - K-R 20 Reliability Coefficients for Beads Subtest by Age Groups and Separate and Combined Ethnic Group Samples	262
6.2.4 - K-R 20 Reliability Coefficients for Passalong Subtest by Age Groups and Separate and Combined Ethnic Group Samples	263
6.2.5 - K-R 20 Reliability Coefficients for Form Assembly Subtest by Age Groups and Separate and Combined Ethnic Group Samples	263
6.2.6 - K-R 20 Reliability Coefficients for Pattern Matching Subtest by Age Groups and Separate and Combined Ethnic Group Samples	263
6.2.7 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 1 (N=27)	265
6.2.8 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 2 (N=35)	266
6.2.9 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 3 (N=34)	266
6.2.10 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 4 (N=31)	266
6.2.11 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Groups 1 to 4 combined (N=127)	267
6.2.12 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 1 European and Maori Subjects	268
6.2.13 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 2 European and Maori Subjects	269

LIST OF TABLES (Contd.)

	Page
TABLE 6.2.14 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 3 European and Maori Subjects	269
6.2.15 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 4 European and Maori Subjects	270
6.2.16 - QT Subtest and QT Total Score Test-Retest Correlation Coefficients for European and Maori Subjects from Class Groups 1 to 4 Combined	270
6.2.17 - Standard Error of Measurement Estimates for QT Total Score by Age Group and for Separate and Combined Ethnic Groups	273
6.2.18 - Standard Error of Measurement Estimates for Knox Subtest by Age Group and for Separate and Combined Ethnic Groups	273
6.2.19 - Standard Error of Measurement Estimates for Beads Subtest by Age Group and for Separate and Combined Ethnic Groups	273
6.2.20 - Standard Error of Measurement Estimates for Passalong Subtest by Age Group and for Separate and Combined Ethnic Groups	274
6.2.21 - Standard Error of Measurement Estimates for Form Assembly Subtest by Age Group and for Separate and Combined Ethnic Groups	274
6.2.22 - Standard Error of Measurement Estimates for Pattern Matching Subtest by Age Group and for Separate and Combined Ethnic Groups	274
 PART 3	
TABLE 6.3.1 - Product-Moment Correlation Coefficients for Class Group 1 (N=27) Between QT Subtests, QT Total Score and Criteria Data	285
6.3.2 - Product-Moment Correlation Coefficients for European Subjects of Class Group 1 (N=17) Between QT Subtests, QT Total Score and Criteria Data	286
6.3.3 - Product-Moment Correlation Coefficients for Maori Subjects of Class Group 1 (N=10) Between QT Subtests, QT Total Score and Criteria Data	287

LIST OF TABLES (Contd.)

	Page
TABLE 6.3.4 - Product-Moment Correlation Coefficients for Class Group 2 (N=35) Between QT Subtests, QT Total Score and Criteria Data	288
6.3.5 - Product-Moment Correlation Coefficients for European Subjects of Class Group 2 (N=27) Between QT Subtests, QT Total Score and Criteria Data	289
6.3.6 - Product-Moment Correlation Coefficients for Maori Subjects of Class Group 2 (N=8) Between QT Subtests, QT Total Score and Criteria Data	290
6.3.7 - Product-Moment Correlation Coefficients for Class Group 3 (N=34) Between QT Subtests, QT Total Score and Criteria Data	291
6.3.8 - Product-Moment Correlation Coefficients for European Subjects of Class Group 3 (N=18) Between QT Subtests, QT Total Score and Criteria Data	292
6.3.9 - Product-Moment Correlation Coefficients for Maori Subjects of Class Group 3 (N=16) Between QT Subtests, QT Total Score and Criteria Data	293
6.3.10 - Product-Moment Correlation Coefficients for Class Group 4 (N=31) Between QT Subtests, QT Total Score and Criteria Data	294
6.3.11 - Product-Moment Correlation Coefficients for European Subjects of Class Group 4 (N=18) Between QT Subtests, QT Total Score and Criteria Data	295
6.3.12 - Product-Moment Correlation Coefficients for Maori Subjects of Class Group 4 (N=13) Between QT Subtests, QT Total Score and Criteria Data	296
6.3.13 - Spearman-Rank Correlation Coefficients for Class Group 1 (N=27) Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'	297
6.3.14 - Spearman-Rank Correlation Coefficients for European (N=17) and Maori (N=10) Subjects of Class Group 1 Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'	297
6.3.15 - Spearman-Rank Correlation Coefficients for Class Group 2 (N=35) Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'	298

LIST OF TABLES (Contd.)

	Page
TABLE 6.3.16 - Spearman-Rank Correlation Coefficients for European (N=27) and Maori (N=8) Subjects of Class Group 2 Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'	298
6.3.17 - Spearman-Rank Correlation Coefficients for Class Group 3 (N=34) Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'	299
6.3.18 - Spearman-Rank Correlation Coefficients for European (N=18) and Maori (N=16) Subjects in Class Group 3 Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'	299
6.3.19 - Spearman-Rank Correlation Coefficients for Class Group 4 (N=31) Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'	300
6.3.20 - Spearman-Rank Correlation Coefficients for European (N=18) and Maori (N=13) Subjects in Class Group 4 Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'	300
6.3.21 - Summary of QT Criterion-Related Validity Coefficients - Class Groups 1 to 4	301
6.3.22 - Summary of QT Criterion-Related Validity Coefficients for European Subjects - Class Groups 1 to 4	301
6.3.23 - Summary of QT Criterion-Related Validity Coefficients for Maori Subjects - Class Groups 1 to 4	302
6.3.24 - Principal Factor Matrix for QT Subtests: Age Group 8:6 - 9:5 (N=73)	309
6.3.25 - Varimax Rotated Factor Matrix for QT Subtests: Age Group 8:6 - 9:5 (N=73)	309
6.3.26 - Principal Factor Matrix for QT Subtests: Age Group 9:6 - 10:5 (N=90)	310
6.3.27 - Varimax Rotated Factor Matrix for QT Subtests: Age Group 9:6 - 10:5 (N=90)	310
6.3.28 - Principal Factor Matrix for QT Subtests: Age Group 10:6 - 11:5 (N=194)	311
6.3.29 - Varimax Rotated Factor Matrix for QT Subtests: Age Group 10:6 - 11:5 (N=194)	311

LIST OF TABLES (Contd.)

	Page
TABLE 6.3.30 - Principal Factor Matrix for QT Subtests: Age Group 11:6 - 12:5 (N=122)	312
6.3.31 - Varimax Rotated Factor Matrix for QT Subtests: Age Group 11:6 - 12:5 (N=122)	312
6.3.32 - Principal Factor Matrix for QT Subtests: Age Group 12:6 - 13:5 (N=138)	313
6.3.33 - Varimax Rotated Factor Matrix for QT Subtests: Age Group 12:6 - 13:5 (N=138)	313
6.3.34 - Principal Factor Matrix for QT Subtests: Age Group 13:6 - 14:5 (N=38)	314
6.3.35 - Varimax Rotated Factor Matrix for QT Subtests: Age Group 13:6 - 14:5 (N=38)	314
6.3.36 - Frequencies of Factor Loadings $> .30$ on QT Subtests for the Six Age Groups (Varimax Rotated Analyses)	316
6.3.37 - Means and Standard Deviations for Age, QT Subtests and QT Total Score: European sample (N=397)	319
6.3.38 - Product-Moment Correlation Coefficients Between Age, QT Subtests and QT Total Score: European sample (N=397)	319
6.3.39 - Means and Standard Deviations for Age, QT Subtests and QT Total Score: Maori sample (N=303)	320
6.3.40 - Product-Moment Correlation Coefficients Between Age, QT Subtests and QT Total Score: Maori sample (N=303)	320
6.3.41 - Principal Factor Matrix for Age and QT Subtests: European sample (N=397)	321
6.3.42 - Varimax Rotated Factor Matrix for Age and QT Subtests: European sample (N=397)	321
6.3.43 - Principal Factor Matrix for Age and QT Subtests: Maori sample (N=303)	322
6.3.44 - Varimax Rotated Factor Matrix for Age and QT Subtests: Maori sample (N=303)	322
6.3.45 - Means and Standard Deviations for Age, QT Subtests and Four Additional Tests: Combined Factor Analysis Sample (N=127)	326

LIST OF TABLES (Contd.)

	Page
TABLE 6.3.46 - Product-Moment Correlation Coefficients Among Age, QT Subtests and Four Additional Tests: Combined Factor Analysis Sample (N=127)	326
6.3.47 - Principal Factor Matrix for Age, QT Subtests and Four Additional Tests: Combined Factor Analysis Sample (N=127)	327
6.3.48 - Varimax Rotated Factor Matrix for Age, QT Subtests and Four Additional Tests: Combined Factor Analysis Sample (N=127)	327
 PART 4	
TABLE 6.4.1 - Proportion of Total Sample, European and Maori Samples Passing Each Item of Knox Subtest	331
6.4.2 - Proportion of Total Sample, European and Maori Samples Passing Each Item of Beads Subtest	331
6.4.3 - Proportion of Total Sample, European and Maori Samples Passing Each Item of Passalong Subtest	331
6.4.4 - Proportion of Total Sample, European and Maori Samples Passing Each Item of Form Assembly Subtest	331
6.4.5 - Proportion of Total Sample, European and Maori Samples Passing Each Item of Pattern Matching Subtest	332
6.4.6 - QT Item Discrimination Indices -D, Based on Upper and Lower 27% of Total Sample	337
6.4.7 - Point Biserial Correlations of Knox Subtest Items with Knox Subtest Score	339
6.4.8 - Point Biserial Correlations of Knox Subtest Items with QT Total Score	339
6.4.9 - Point Biserial Correlations of Beads Subtest Items with Beads Subtest Score	340
6.4.10 - Point Biserial Correlations of Beads Subtest Items with QT Total Score	340
6.4.11 - Point Biserial Correlations of Passalong Subtest Items with Passalong Subtest Score	340

LIST OF TABLES (Contd.)

	Page
TABLE 6.4.12 - Point Biserial Correlations of Passalong Subtest Items with QT Total Score	340
6.4.13 - Point Biserial Correlations of Form Assembly Subtest Items with Form Assembly Subtest Score	341
6.4.14 - Point Biserial Correlations of Form Assembly Subtest Items with QT Total Score	341
6.4.15 - Point Biserial Correlations of Pattern Matching Subtest Items with Pattern Matching Subtest Score	342
6.4.16 - Point Biserial Correlations of Pattern Matching Subtest Items with QT Total Score	342
6.4.17 - Matrix of Product-Moment Correlation Coefficients Amongst QT Items (N=700)	346
6.4.18 - Principal Factor Matrix for QT Items Based on Product-Moment Correlation Coefficients Among Items: The Six Factors Extracted After Applying Cattell's Scree Test (N=700)	354
6.4.19 - Varimax Rotated Factor Matrix for QT Items Based on Product-Moment Correlation Coefficients Among Items: The First Four Factors From Rotated Principal Factor Matrix After Applying Cattell's Scree Test (N=700)	357
PART 5	
TABLE 6.5.1 - QT Decile Rank Norms for Five New Zealand Age Groups	364

LIST OF FIGURES

CHAPTER 5	Page
FIGURE 5.1 - Knox Cube Imitation Subtest Materials	177
5.2 - Beads Subtest Materials	180
5.3 - Item 6 of the Beads Subtest as Presented to Subject	180
5.4 - Passalong Subtest - Practice Item	183
5.5 - Passalong Subtest - Item 1	184
5.6 - Passalong Subtest - Item 2	184
5.7 - Passalong Subtest - Item 3	185
5.8 - Passalong Subtest - Item 4	185
5.9 - Passalong Subtest - Item 5	186
5.10 - Passalong Subtest - Item 6	186
5.11 - Passalong Subtest - Item 7A	187
5.12 - Passalong Subtest - Item 7B	187
5.13 - Item 4 of Passalong Subtest and Stimulus Card as Presented to Subject	188
5.14 - Form Assembly Subtest Formboard as viewed by Subject	191
5.15 - Item 5 of the Form Assembly Subtest as viewed by Subject	191
5.16 - Form Assembly Subtest Practice Items (4) and Test Items (13) as viewed by Tester	192
5.17 - Pattern Matching Subtest Practice Item (1) and Test Items (15) as viewed by Tester	194
5.18 - Item 12 of the Pattern Matching Subtest as Presented to Subject	195
 CHAPTER 6	
FIGURE 6.1 - Regression of QT Scores on Age for European (N=397) and Maori (N=303) Samples	216
6.2 - Mean Scores for Each Total Sample, European and Maori Age Group on the QT	231
6.3 - Mean Scores for Each Total Sample, European and Maori Age Group on the Knox Subtest	232

LIST OF FIGURES (Contd.)

	Page
FIGURE 6.4 - Mean Scores for Each Total Sample, European and Maori Age Group on the Beads Subtest	233
6.5 - Mean Scores for Each Total Sample, European and Maori Age Group on the Passalong Subtest	234
6.6 - Mean Scores for Each Total Sample, European and Maori Age Group on the Form Assembly Subtest	235
6.7 - Mean Scores for Each Total Sample, European and Maori Age Group on the Pattern Matching Subtest	236
6.8 - Frequency Distribution of QT Scores for European Sample (N=397)	238
6.9 - Frequency Distribution of QT Scores for Maori Sample (N=303)	239
6.10 - Frequency Distribution of QT Scores for Total Sample (N=700)	240
6.11 - Frequency Distribution of Knox Subtest Scores for Total Sample, European and Maori Samples	245
6.12 - Frequency Distribution of Beads Subtest scores for Total Sample, European and Maori Samples	246
6.13 - Frequency Distribution of Passalong Subtest scores for Total Sample, European and Maori Samples	247
6.14 - Frequency Distribution of Form Assembly Subtest scores for Total Sample, European and Maori Samples	248
6.15 - Frequency Distribution of Pattern Matching Subtest scores for Total Sample, European and Maori Samples	249
6.16 - Proportion of Total Sample Passing Each Item (N=700)	333
6.17 - Proportion of European Sample Passing Each Item (N=397)	334
6.18 - Proportion of Maori Sample Passing Each Item (N=303)	335

LIST OF ABBREVIATIONS

A.C.E.R.	Australian Council for Educational Research
AJHR	Appendices to the Journals of the House of Representatives
ANOVA	Analysis of Variance
A.P.A.	American Psychological Association
NGPS	New Guinea Performance Scale
N.Z.C.E.R.	New Zealand Council for Educational Research
N.Z.O.Y.B.	New Zealand Official Year Book
PIPS	Pacific Infants Performance Scale
P.I.R.	Pacific Islands Regiment
PPVT	Peabody Picture Vocabulary Test
PRST	Pacific Reasoning Series Test
QT	Queensland Test
WAIS	Wechsler Adult Intelligence Scale
WISC	Wechsler Intelligence Scale for Children
WISC-R	Wechsler Intelligence Scale for Children - Revised

INTRODUCTION

This thesis is concerned with the measurement of general intelligence in two ethnic groups. It is primarily the study of one particular measure of ability that has been developed for cross-cultural use. The thesis presents in considerable detail the extensive but necessary psychometric data that should accompany such a test. This data is essential in determining the applicability and equivalence of particular psychological measures for subjects from different cultural and ethnic group backgrounds. It is only through the generation of such detailed psychometric data that the general adequacy of psychological measures can be determined.

The data reported bears upon many strictly empirical questions. Yet the study is not independent of a number of broader issues within psychology and society itself. It has been necessary for discussion and debate to range across both the narrow empirical concerns and the more encompassing questions that lie behind the study and the data.

A brief outline of each chapter is given to assist the reader in perceiving the links between these domains in the development and execution of the study.

Chapter 1 reviews in some detail developments in the psychological study of human intelligence. Practical steps in the development of intelligence tests are discussed along with the major theories that have arisen from the study of intelligence test data. The cross-cultural study of intelligence is then considered. Attention is given to both the theoretical and practical issues that have arisen in this increasingly complex domain.

It is argued that the concept of general intelligence and the tests of general intelligence continue to be important in psychology. The cross-cultural application of such tests depends, however, upon their

adequate empirical investigation within the cultures concerned.

Chapter 2 focuses upon Maori and European educational attainment differences in New Zealand. The nature of the explanations for these differences are commented upon and it is noted that the issue of intelligence and its measurement has received more than a passing mention over the years. The basis of an ethnic group differentiation in New Zealand is also considered in this chapter.

The empirical evidence from the many comparative studies on European and Maori levels of performance on tests of intelligence are reviewed in Chapter 3. Study of this data led to a number of conclusions on the adequacy of much previous research in this field and in turn to the formulation of a set of research proposals. Few studies have investigated the comparative performances of Maori and European subjects on any individual scale of general ability in the years covering late childhood to early adolescence. Virtually no data have been reported on the psychometric properties of such instruments with these groups. Consequently research directed at these issues was undertaken with the Queensland Test, a recently developed measure of general ability for use in cross-cultural contexts.

The research proposals and methods are described in Chapter 4 and the origins, form and previous research applications of the Queensland Test are described in Chapter 5.

Chapter 6 presents the results. These include summary and descriptive statistics on Queensland Test performances for various sample groups, data on the reliability, validity and item characteristics and a set of preliminary norms.

The results are summarized in Chapter 7. The data from these studies are considered in their own terms and also in relation to other Queensland Test research. The findings are related to the general study of intelligence, to other recent studies on the cross-cultural measurement of abilities in New Zealand and to concerns in the educational domain.

CHAPTER 1

THE STUDY OF HUMAN INTELLIGENCE

Introduction

The study of human intelligence has been a central theme of modern psychology. If the volume of literature and research is any guide to the importance of a concept then it is apparent that the nature of human intelligence and its assessment have been dominant concerns of psychology.

An exhaustive survey of the literature on intelligence and intelligence testing is well beyond the scope of this chapter. Nevertheless, it is important to comment upon a number of developments that have taken place in the study of human intelligence within psychology.

The first section of this chapter recalls some early views on intelligence and traces the development of the modern intelligence test. The second section focuses on the continuing problem of defining intelligence and the third on the somewhat conflicting psychological theories of intelligence that have evolved.

The research presented in this thesis concerns the cross-cultural assessment of intellectual ability. Special issues relating to the definition and nature of intelligence and its assessment in cross-cultural contexts are considered in the fourth and final section of this chapter.

Some Early Views on Intelligence and its Measurement

As McNemar (1964) has pointed out, concern over the reasoning and thinking powers of people has been with us for a long time. From the Greeks we have inherited the term nous which referred to mental operations in the form of some inner "spontaneous power acting according to ends giving movement, unity and system to what had previously been a jumble of inert elements" (Peterson, 1926, p.11). Nous is still in popular use today as an expression meaning common sense.

The Romans also had a word for this. They called it intellectus.

This term was used in translating nous from Greek to Latin, and in the main, took its meaning from the Greek. To the Romans intellectus meant both the conduct of thought and to reason. The English language derivatives of intellect and intelligence have come to mean knowing, reasoning and the power of understanding (see Fowler and Fowler, 1964, p.633).

A long and very involved philosophically orientated discourse on the capacity of people to know reason and understand has been pursued over the centuries. In Herrnstein and Boring's (1966) A Source Book in the History of Psychology material is brought together from Aristotle (c.350 B.C.) on the one hand, to Binet (1895) the father (one might say) of the modern intelligence test, on the other. General texts such as Thomson's (1968) The Pelican History of Psychology also provides interesting material on early views on the mind and the concept of intelligence in the developing discipline of psychology along with notes on the methods instituted for its measurement.

Peterson's (1926) detailed study of the evolution of the concept of intelligence and the construction of the Binet and Simon scales probably still provides however, one of the most authoritative accounts. Apart from considering the views of the ancients, Peterson has recorded in considerable detail, the philosophical contributions from Europe.

According to Peterson it was Decartes with his "mathematical and rationalistic turn of mind (who) opened up the way for an independent consideration of the mental processes" (p.41). Decartes was to leave however, the philosophical difficulties of the mind - matter distinction and the problems of their interaction by adopting an extreme dualist position.

Hobbes, who is placed by Peterson in the forefront of what he has called the "British Empirical Movement" (p.45), rejected however, the view that the mind was a separate and different substance to the body. He sought to explain human behaviour in terms of natural or mechanical processes. To Hobbes"Man's nature is the sum of his natural faculties

and powers, as the faculties of nutrition, motion, generation, sense, reason, etc." (quoted in Peterson, 1926, p.46, Peterson's italics). Using the term faculties to mean functions, Hobbes divided the faculties in two areas. These were "..... those of the body - nutritive, motive and generative - and those of the mind - cognitive and motive"(Peterson, 1926, p.46). The cognitive function was that concerned with all".....Knowledge and imagination as we have of things in the outer world, things whose images or conceptions would remain even if they themselves were destroyed" (Peterson, 1926, p.46). The origin of these images or conceptions lay in both the action of things in the outer world that produced sensations and with the inner bodily stimulations that recalled images by association.

Hobbes cut through the dualist dilemma by conceiving of body and mind as an indivisible sense system. He suggested that there existed associations between sensation and cognitive events and in advancing the view that the world was only known through the senses, challenged the notions of pure reason and innate ideas held by Spinoza and others (see Peterson, 1926, p.44).

In Peterson's view the foundation for the empirical study of intellectual or cognitive functions was laid by Hobbes. But, it was the writings of Locke which were to greatly strengthen this position. Locke also rejected the innate ideas proposition, holding that "..... the mind begins a tabula rasa and must gain all its knowledge through the senses" (Peterson, 1926, p.48).

Hume, Hartley and James Mill were to follow the lead given by Locke in the emphasis given to sensation in the creation and execution of intellectual skills or capacity. Around these men arose the associationist school of thought which asserted that all perceptions, memories and thoughts were the result of processing sensory and ideational data, with the mechanism for this processing being the power of association as evoked through the contiguity of sensations and ideas. As Thomson points out however, Mill had to employ a concept of synchronous association which led to a reductio ad absurdum position since complex experiences such as

thinking were taken to be the synchronous association of all simpler relevant ideas and sensations. "The principle of association had been pressed beyond the limits of creditability as an empirical construct" (Thomson, 1968, p.25).

James Mill's son, John Stuart Mill, was however, to salvage a little from the wreckage of associationist theory. To J.S. Mill the mind was active and therefore able to synthesize sensory and ideational data creating new forms. Sensations and ideas remained the ingredients of experience and associations between these ingredients continued, but with the new analogy of 'mental chemistry' ideas could -

"..... coalesce by forming rapid associations in which some ideas become so peripheral as to drop out. Total perceptions may be telescoped and economised" (Thomson, 1968, p.26).

The views of J.S. Mill were held in high regard by the philosophers and the new psychologists in both Britain and Europe throughout the nineteenth century.

German philosophy under Kant had a strong metaphysical flavour, nonetheless, Herbart who succeeded Kant as Professor of Philosophy at Königsberg argued for a separate and empirical psychology to study 'mental dynamics'. He argued that psychology must use mathematics, that "attributes of mental states must be measured and their relationships must be treated statistically" (Thomson, 1968, p.31).

While the British empiricist philosophy dominated much thought and speculation on the nature and organisation of the mind, including the quality of intelligence, it was principally the work of the German scientists Fechner, von Helmholtz, Weber and Wundt who took psychology from the realms of philosophy with their psychophysical methods.

They sought to study the physical and measurable properties of sensation as distinct from the introspective recalling of sensations or events in the 'stream of consciousness'. An experimental methodology was established along with a set of measuring devices. In their efforts they were also

struggling to establish psychology as a scientific discipline by attempting to meet Kant's conclusion to his 1786 essay Metaphysical Foundations of Natural Science "..... that sciences must be mathematized" (Levine, 1976, p.229).

By the late nineteenth century a number of strands of thought and technical development in European psychology were beginning to come together in a way that would fashion much of intelligence test theory and intelligence test technology as we know it today.

The psychophysicists of Germany were studying numerous approaches to the measurement of sensation in the attempt to understand mental abilities and in the search, Weber had noted that weight discrimination depended upon relative not absolute differences - a principle that still underlies much psychological measurement.

In France, Pinel and Itard were laying the foundations of scientific psychiatry with systematic studies into mental deficiency. To assist in classifying mental defectives for instructional purposes, Itard's pupil, Sequin, developed a formboard test. The work of the German psychophysicists was brought to France by Ribot who assisted in establishing a psychological laboratory at the Sorbonne. This laboratory was placed under the direction of Binet.

A third strand of considerable importance originated in Britain. In 1859 Charles Darwin published The Origin of Species by Natural Selection propounding an evolutionary thesis which was to strongly incline British psychology towards a biological inheritance orientation. Francis Galton, a half-cousin of Charles Darwin, undertook studies into the inheritance of mental characteristics and produced theories on the mechanisms of inheritance of ability.

Galton promoted the idea that there existed a general intellectual power which was distributed in some continuous manner from the intellectually backward to the brilliant. This contrasted sharply with the then popular view that behaviour was governed by various mental faculties. The latter view had led to phrenology and 'mental gymnastics' for exercising

and improving the faculties.

In the course of his studies into individual differences in the physical, behavioural and mental domain, Galton was responsible for the development of a wide range of novel measuring devices (see DuBois, 1970, pp.11-14).

At the International Health Exhibition held in London in 1884, Galton opened an Anthropometric Laboratory. For a small fee a series of measures were made of both static characteristics (e.g. height, weight) and various behaviours (e.g. strength of pull, memory of form, colour discrimination). The laboratory was moved to the Science Museum at South Kensington when the exhibition closed. Data on a wide range of variables was collected from over 9,000 subjects from which Galton was able to produce tables of norms for males and females on a number of physical and behavioural characteristics, the norms for the latter requiring expression in percentile terms since absolute measures could not be devised.

Galton also developed both the basic concept of correlation and made the first attempts to establish mathematical procedures for measuring it. Given the mass of unreduced data at his disposal, Galton was prompted to seek an expression for understanding the imperfect relationships that existed between the variables he had studied. He was seeking to discover the degree of 'concomitant variation' between certain variables - a task which J.S. Mill had seen as fundamental to any science. The development and elaboration of a number of statistical concepts and procedures for studying the correlation amongst variables was the work of Galton's close friend, Karl Pearson.

By the end of the nineteenth century Galton had laid the groundwork for a theory of general intelligence. In France, Binet had become involved with the question of diagnosing mental deficiency and beginning with the German psychophysicists the new science of psychology had an extensive library of measures for all manner of sensory functions.

The introduction of these European developments to the United States was undertaken by James McKeen Cattell. Cattell had studied under Wundt in Germany and was subsequently an assistant in Galton's Anthropometric Laboratory. On his return to the United States he inaugurated systematic psychological testing in laboratories at both the University of Pennsylvania and Columbia University. His tests were of both physical and mental phenomena, the latter being predominantly sensory in nature. Cattell's tests are described in some detail by DuBois (1970).

Cattell administered the battery of tests to university students for a number of years and reported statistical summaries of the results (Cattell and Farrand, 1896). More importantly however, Wissler (1901) made an elaborate statistical analysis of the relationship amongst the tests themselves and their correlations with class grades. The results were most disappointing. As DuBois points out -

"The laboratory mental tests had low intercorrelations. While class grades showed appreciable correlations with grades in other courses, no dependable relationships were discovered with class grades and laboratory tests. A promising start in mental measurement apparently had failed" (DuBois, 1970, p.23).

New and perhaps radically different approaches were needed if both the theory and practice of mental measurement were to move forward. Two essentially unrelated publications were to swing the course of the mental measurement movement in a new direction. These were Spearman's (1904) paper "General Intelligence", objectively determined and measured' and Binet and Simon's (1905 a,b) papers on their systematically organised tests for diagnosing mental retardation by what they called the psychological method.

Spearman's paper presented the core of what has become known as the two-factor theory of intelligence. He suggested the existence of a general ability factor that contributed in varying degrees to all tasks of an intellectual nature and an array of specific abilities that represented more unique talents or skills. Furthermore, he set down statistical methods for locating this general ability factor that was seen to underlie

performances on a group of psychological tests.

Flügel (1933) looking back on this period when the work of Galton, Pearson, Spearman and others was to come together in a new synthesis said -

"The need for some such method of correlation had been realised by Galton, and the mathematical foundations of the method had been laid by Karl Pearson but it was C.E. Spearman who first realised the full importance of correlation for psychology, who derived new and simpler methods of calculation, who elaborated methods for correction of errors inherent in the 'crude' correlation co-efficient, and who by much further work, both mathematical and experimental, showed how mental tests could be used to attack an all-important problem of general psychology - no less a problem indeed than that involved in the ancient doctrine of faculties. He thus united the psychology of individual differences with the psychology of laws or principles, the apparently divergent schools of Galton and of Wundt, and proved that the former could fructify the latter. It is to Spearman, far more than to anyone else, that we owe our present knowledge as to the structure or 'make up' of the human mind, together with the necessary methods for increasing this knowledge by further research" (Flügel, 1933, p.310).

Spearman had contributed an important theoretical perspective to psychology and the foundations to the statistical methodology of factor analysis, a technique fundamental to a number of current psychological theories of intelligence. Spearman was however, still working with measures of sensory capacities such as the discrimination of weight, pitch and light which Wissler had found did not correlate with the external criterion of university grades. It was the work of Binet and Simon that was to change the face of mental measurement.

Binet was commissioned in 1904 by the French Ministry of Public Instruction "to devise a standardized test for detecting defective children in primary grades so they could be placed in special schools" (Thomson, 1968, p.190). In collaboration with Simon he produced a series of thirty tests whose content included amongst other things, the naming of objects in a picture, digit span, word definition, comparisons of length of lines and copying of designs. With items such as these Binet and Simon -

"produced a standard scale consisting of a series of tests of increasing difficulty, each of which matched a specific developmental level - all tests at a given level were capable of being solved by any normal child in a specific

age group" (Thomson, 1968, p.191).

Again, in Thomson's words -

"This was a turning point in psychology. A new type of test had been produced in which the average level of performance was the criterion, and in which test contents were correlated carefully with other criteria (results of school examinations, assessments of teachers, etc.)" (Thomson, 1968, p.191).

To DuBois, what had been achieved was -

"a new combination of existing elements. Binet and Simon had developed the first modern psychological test, consisting of separate items, chosen systematically in relation to difficulty level and outside criteria (age and judged intelligence), and published with careful instructions for administration and interpretation" (DuBois, 1970, p.36).

The refinement and development of the 1905 scale by Binet and others in Europe, the English language publication of the Stanford Revision of the Binet-Simon Intelligence Scale by Terman (1916) and its international acceptance as a model for the individually administered intelligence test is well documented in basic texts such as Anastasi (1968) and Cronbach (1970).

The type of tests developed by Binet and Simon did in many respects represent a radical departure from the attempts to measure mental processes via tests of a sensory nature. In so far as Binet and Simon's work did not appear to arise out of the prevailing psychophysical psychology of the day, it has been said that "The test developed by Binet was very largely atheoretical" (Kamin, 1974, p.5). However, DuBois' account of the development of the Binet-Simon Intelligence Scale would seem to suggest otherwise. DuBois has shown how a number of themes in the work of Binet played an important part in the theory and rationale behind the 1905 scale. Binet's interest and studies in the area of individual differences lead to an appreciation of developmental processes in such phenomena as memory and an awareness of the psychometric issues surrounding the problem of whether or not tests could measure true amounts or only assist in classifications according to relative amounts. DuBois also points out that Binet, while admiring the great precision with which elementary psychological processes

could be measured (e.g. reaction time), thought it much more important to measure higher psychological processes where, in his view, individual differences were greater.

Furthermore, Binet addressed himself to the question of defining intelligence. DuBois notes that -

"The definition of intelligence was a recurrent theme in his articles and in 1890 he had this to say 'In the narrow sense of the word, that which we call intelligence consists of two phases: first, to perceive the outside world, and then to reinstate the perceptions in memory, to rework them and to think about them'" (DuBois, 1970, p.30).

Binet's views on the nature of intelligence and the content of his tests suggest a concept that involved the blending of a number of functions such as memory, reasoning skills, concentration, mechanical skills and verbal facility. All of these needed to be assessed to come to a sound estimate of 'general ability'. Galton had already proposed a general ability concept and Spearman added weight to such a concept along with a methodology for its investigation. Binet was to add a mechanism of measurement that conceptually coincided with these views.

The publication in 1916 of the Stanford Revision of the Binet-Simon Intelligence Scale by Terman acted to strengthen both intelligence testing and the Binet-Simon scale itself. Terman's revision contained improved content, normative data based on a large sample of United States children and replaced Binet's concept of mental age with that of the intelligence quotient, which had been suggested somewhat earlier by Stern (1911) in the form of a mental quotient. Following its publication the Stanford Revision was employed in the translation of Binet tests from English to a number of other languages. Smith (1974) has presented an analysis of the pattern of translation of the Binet tests. The 1916 Stanford Revision not only provided the base for all subsequent United States developments of the test, but was the model for translations into Punjab and Urdu, Kannada and Marathi, Swiss-German and German.

Nonetheless, despite the possibility of translation of the new

intelligence tests, it was recognised that the results were not necessarily dependable with subjects for whom English was not their first language, who were handicapped by speech defects, by deafness or the test content itself. The result was the development of the performance tests of intellectual ability.

DuBois records that -

"As early as the 1904 World's Fair in St. Louis, Woodworth (1910) had used a form board and other performance tests in his study of race differences. To test immigrants at Ellis Island for possible mental defect, Knox (1914) had developed a series of tests requiring imitation and other action but no language responses. The Healy and Fernald tests (1911) were developed for use as supplements to the regular scales (for) children who spoke a foreign language or who had handicaps of various kinds One of the most useful of these tests was a nonverbal application of the concept behind the Ebbinghaus completion test, involving selection of appropriate pieces to complete a picture" (DuBois, 1970, pp.52-53).

Pintner and Paterson (1917) pioneered the first formal scale of performance tests which included form boards and wooden puzzle boards, the assembly of parts to form a picture, a human figure or face, a block tapping task and digit-symbol substitution items. At about the same time as the work of Pintner and Paterson, two other important performance tests were published.

In 1915, Porteus published a series of graded mazes. These were printed on paper and the subject traces the pathway through the maze with a pencil attempting to avoid entry into a cul-de-sac or crossing lines. Perhaps the most famous performance test however, has been Kohs' Block Design (Kohs, 1920). With its coloured cubes and pattern cards it has, in various forms, appeared in almost all major performance scales since its publication.

Goodenough (1926) proposed the measurement of intelligence through the use of a subject's drawings and developed the Draw-a-Man Test. This device has not only been regarded as another performance approach to the

measurement of intelligence, but has also evolved into a 'projective' instrument tapping internal motivational and emotional conditions.

The 1930's saw the publication of two further individual performance tests of intelligence that synthesized and refined many earlier developments. These were the Arthur Point Performance Scale (Arthur, 1930) and the Leiter International Performance Scale (Leiter, 1936).

The development of the paper and pencil group test of intelligence was also being undertaken at the time when Binet published his model for the individual intelligence test. It was apparently Ebbinghaus who pioneered the group test with his 1896 publication of a syllable 'completion test' that rendered a simple numerical index of intellectual ability.

According to DuBois -

"Ebbinghaus anticipated the invention of the group intelligence test two decades later. His tests were printed on the inside of a folded sheet; on the outside the student wrote name, grade, place and age. At a signal to begin, the student opened the folder and began to work. He closed it when the signal to stop was given. Whole classes were tested simultaneously" (DuBois, 1970, p.27).

It is reported by Thomson (1968) that Burt in 1911 re-introduced the paper and pencil intelligence test when presenting the verbal analogy item, a form of item that has subsequently become an important feature of many intelligence tests. DuBois also points out that other psychologists were also working on group administration procedures for tests of varying kinds at that time. The great stimulus to the group intelligence testing movement came not from Europe however, but from the United States.

The entry of the United States into World War I in 1917 resulted in Robert Yerkes, as president of the American Psychological Association, offering the services of psychology to the nation in the war effort. The psychological examination of recruits was proposed as an area of practical service. In order to meet the challenge of examining all recruits, a committee under the chairmanship of Yerkes set about devising a group test of intelligence. Working from a prototype provided by Arthur Otis, the committee in some four months developed the now celebrated Army Alpha which became the blueprint for the majority of verbal group tests of

intelligence. Indeed, in psychological circles the name Otis has become synonymous with tests of this type. In much the same way the Army Beta, which was designed as a non-verbal group measure for illiterate and non-English speaking recruits, became a prototype for tests of this nature, although now it is probably not as well known as the Raven's Progressive Matrices Test (Raven, 1938).

By the 1920's the foundations of the psychometric movement in psychology had been laid and the technology of intelligence testing established. The theory accompanying this technology has not however, led to a closing of ranks amongst psychologists over the nature of intelligence.

In closing this section, it should be noted that as so often seems to be the case nowadays, many of the recent discoveries of the West were known in ancient China. DuBois, while focusing on the development of modern psychological testing in Europe and the United States, begins his history with an account of the elaborate competitive examinations employed to select government personnel in China from as long ago as 2200 B.C. As a mark of respect, his book is dedicated 'To those wise men of China who, thousands of years ago, invented the psychological test'. This should serve to remind us that the development of psychological measurement has not been a purely occidental affair.

Intelligence: Comments on Definitions

Defining intelligence is no easy matter - even for psychologists. In the preceding section reference was made to some early views and to changing conceptualizations as efforts to measure intelligence shifted from tests of a sensory nature to tests purporting to tap 'higher mental processes'.

Binet placed considerable emphasis upon these higher mental processes in the study of intelligence. Binet and Simon (1905 a) certainly wrote about such separate mental functions as judgement, initiative, practical sense and adaptation to circumstances, but as Cronbach (1970) points out, Binet proposed a more global, general or inclusive concept of

intelligence that was more than the sum of the separate functions. Cronbach drawing on Terman's (1916) account of Binet's work notes that Binet -

"gradually identified the essence of intelligence as "the tendency to take and maintain a definite direction; the capacity to make adaptations for the purpose of attaining a desired end; and the power of auto-criticism" (Cronbach, 1970, p.200, quoting Terman, 1916, p.45, Cronbach's italics).

In Cronbach's view, Binet made an important contribution in setting aside the idea of distinct and separate mental functions "in favour of the concept of a pervasive, general intelligence' (Cronbach, 1970, p.201). Terman, over half a century earlier, had also noted that -

"After vainly trying to disentangle the various intellectual functions, Binet decided to test their combined functional capacity without any pretense of measuring the exact contribution of each to the total product" (Terman, 1916, p.151).

In the interplay between the empirical approach taken by Binet in order to establish a set of test items that would discriminate between bright and dull children of various ages and his judgement that intelligence must involve purposive and adaptive behaviour tempered by rational appraisal, arose the major elements of the still current concept of general intelligence. However, while the term general intelligence and its synonyms have become commonplace in both psychological and educational discussions, its meaning and method of assessment is still not agreed upon.

In 1921 an important symposium was held on intelligence and its measurement, the proceedings being reported on the Journal of Educational Psychology of that year. In this symposium, 13 psychologists put forward 13 different views on the nature of human intelligence, although to be fair, there was much in common in their definitions. One of the participants was Terman who suggested that intelligence was the ability to engage in "abstract thinking". Terman also pointed out the danger of placing too much emphasis on the results of any one test. He noted that -

"We must guard against defining intelligence solely in terms of ability to pass the tests of a given intelligence scale. It should go without saying that no existing scale is capable of adequately measuring the ability to deal with all possible kinds of material on all intelligence levels" (Terman, 1921, p.131).

This point, made by Terman is as appropriate today as in 1921.

Burt (1955) has commented on this symposium noting that the descriptions suggested for the term intelligence were not intended to be definitions in the logical sense of attempts to explain the word, but primarily a practical discussion "..... to determine how intelligence appears to operate....." (Burt, 1955, p.159). Rather than trade off elements of these and other descriptions, Burt examined both the non-statistical and statistical evidence for the concept of intelligence. From his wide ranging review, Burt concluded that -

"..... there is a general factor making for efficiency in all mental activities, that this factor is essentially cognitive or directive, and that the greater part of the individual variance found in this factor is attributable to differences in genetic constitution. This triple conclusion suggested a modernized formula for the abstract conception to which so many different writers have been led, viz., 'innate, general, cognitive ability'. If, therefore, we are to retain the word 'intelligence' as a technical term in psychology, this still seems the best definition" (Burt, 1955, p.176).

The issue of defining intelligence was also taken up by Miles (1957) at this time. Miles took the view that the issue had been most thoroughly confused by arguments at cross purposes and by unnecessary disputation. He suggested that the confusion could best be resolved by understanding the different senses of the word 'definition' as applied in the study of intelligence. Six forms of definition were identified by Miles. These were definitions based on the notion of a 'real essence', lexical definitions and appeals to ordinary usage, stipulative definitions, descriptive and naming approaches, the search for key constituents and lastly operational definitions. Each type of definition was discussed at length.

On the whole, Miles appears to favour an operational definition and argues that a concept or "..... substrate has meaning only in relation to its exemplaries" (Miles, 1957, p.158). For Miles -

"Instead of 'what is intelligence?' or even 'what does the word "intelligence" really mean?' we need to ask "How do you test - or what operations are involved in testing - whether a person is intelligent?' This question can readily be answered. Psychologists have devised standardized tests. It is the items in these tests (or, more strictly, the person's behaviour in producing correct responses to these items) that are regarded as constituting the exemplaries of the word 'intelligent'. Intelligence, in other words, is what intelligence tests measure" (Miles, 1957, pp.158-159, Miles' italics).

The dictum that 'intelligence is what intelligence tests measure' is, of course, circular as it stands but as Miles points out "..... if in the definiens we substitute the names of particular intelligence tests any trace of circularity disappears" (Miles, 1957, p.159, Miles' italics). The dictum also has merit as a reminder to the fact that the construct of intelligence stems ultimately from the form of measurement adopted. This is a most critical point when it is remembered that there now exists a multiplicity of tests of intelligence and a good many models attempting to describe the nature of human intelligence.

While Burt's conceptualization of intelligence as an innate, general, cognitive ability is not, as we shall see, universally shared amongst psychologists, the place of operational definitions would appear to be well established. Even a casual inspection of material on intelligence whether from a psychometric perspective (e.g. Vernon, 1961) or a developmental-cognitive psychology approach (e.g. Bruner, Olver and Greenfield, 1966) indicates that considerable attention is given to the actual measurement operations involved in any study.

Butcher (1968) has shown considerable concern over the question of defining intelligence. He notes that Vernon (1960) had classified descriptions of intelligence into three types - biological, psychological and operational.

The biological descriptions generally involved the idea of adaptation to environmental circumstances. To Butcher -

"This is the most far-reaching and general view. But for many practical purposes this idea of intelligence is too fundamental and general. Many human beings, to whom one could hardly deny an assessment of

exceptional intelligence have been spectacularly ill-adapted to their physical and social environment. The biological picture of intelligence is therefore, although important as a bird's-eye view, not necessarily of practical use in the study of individual differences within a given culture" (Butcher, 1968, p.28).

Psychological definitions took a number of forms. They included on the one hand Burt's (1955) belief that intelligence is 'innate general cognitive ability' and on the other, Piagetian views about intelligence as a set of logically developing operations and 'schemata' through which the individual acquires, uses and creates knowledge (see Inhelder and Piaget, 1958). As Butcher points out, the psychological definitions suggest a genetic substrate to intelligence and its development, but "scores on existing tests have often been shown to be susceptible to environmental influences" (Butcher, 1968, p.28). Consequently, intelligence as defined in psychological terms "differs from intelligence as measured by tests" (Butcher, 1968, p.29). Butcher suggests that for many psychologists this creates a degree of scientific discomfort since in their view "it appears more satisfactory to be able to relate one's concepts directly to actual observations or measurements" (Butcher, 1968, p.29).

Butcher argues that this dilemma has been resolved by psychologists by one of two ways -

"Either, like Hebb and R.B. Cattell, they have distinguished two kinds of intelligence, calling them 'Intelligence A and Intelligence B' (Hebb) or 'Fluid and Crystallised Intelligence' (Cattell), or they have plumped for an operational definition in an attempt to parallel procedure in other sciences, and have maintained that intelligence should be defined as 'what the tests test'" (Butcher, 1968, p.29).

Hebb (1949) and Cattell (1963) are essentially making quite similar distinctions. Both are suggesting that one kind of intelligence can be thought of as genetic potential, while the other kind is, in the main, the result of experience, learning and environmental circumstances. The views of Cattell and especially Hebb are more fully discussed below.

The appeal of operational definitions of the form 'intelligence is what intelligence tests measure' is strongest amongst "those most eager to

establish psychology as a satisfactory scientific discipline" (Butcher, 1968, p.30). Arguments over the operational definition of concepts such as intelligence involve however, important questions concerning scientific methodology, the conduct of psychological research and the scientific status of the behavioural sciences. Butcher discusses these concerns and arrived at the following position on operational definitions -

"Operational definitions have the advantage that they prescribe standard procedures, ensure agreement about concepts between different observers and eliminate those that are too ambiguous or hazy to be translated directly or through a chain of operations into an observable and measurable physical occurrence. Their disadvantage is that they may be arbitrary and restrictive unless changed as knowledge expands" (Butcher, 1968, p.30).

Arguments against operational definitions are seen by Butcher to reflect the uncertainty over the state of psychology as a science. However, he is of the opinion that -

"Future trends are likely, almost certain, to be in favour of the hard-headed fraternity of psychologists that insists on operational definitions and distrusts 'open' concepts with their seductive penumbra of vagueness" (Butcher, 1968, p.31).

Eysenck (1973) has recently summarised the conceptual and operational issues in the study of intelligence thus -

"The man in the street, and often the unwary psychologist too, thinks of intelligence as something really existing 'out there'; something which the psychologist may or may not recognise successfully, and measure with more or less success. In these terms it would make sense to argue about whether a particular test 'really' measures intelligence. Such reification is utterly mistaken; there is nothing 'out there' which could be called intelligence, just as there is nothing 'out there' which could be called gravitation. Intelligence and gravitation are concepts, and concepts only exist in the minds of scientists; they are useful or useless, appropriate or inappropriate, in terms of their success in enabling us to form generalizations, discover invariances, and predict future events" (Eysenck, 1973, p.1, Eysenck's italics).

Eysenck goes on to suggest that E.L. Thorndike, who is said to have said "Everything that exists, exists in some quantity and can therefore be measured" was in a sense wrong, since intelligence does not exist. However,

this does not mean that it cannot be measured. Returning to the gravity analogy, Eysenck points out that gravitation can be measured very accurately although it does not exist either. Reinforcing and clarifying the issue, Eysenck puts it this way -

"Individuals exist and their behaviour can be observed and measured; these observations and measurements give rise to concepts which we reify at our peril. It makes sense to argue about the usefulness of the resulting concepts; it makes no sense to argue about the 'existence' of these conceptualisations" (Eysenck, 1973, p.2).

On the question of utility there would seem to be little doubt, for the present, about the usefulness of the concept of intelligence. Psychologists, teachers, guidance counsellors and many others who are required to make evaluations and decisions concerning individuals usually find the concept of intelligence one of the most useful and far reaching in its implications for decision making. Furthermore, in Cronbach's view -

"The outstanding success of scientific measurement of individual differences in behaviour has been that of the general mental test. Despite the over-enthusiasm and occasional errors that have attended its development, the general mental test stands today as the most important single contribution of psychology to the practical guidance of human affairs" (Cronbach, 1960, p.157).

Psychological Theories of Intelligence

As a result of the considerable efforts psychologists have poured into the study of human intelligence, a number of psychological theories or models have been advanced that attempt to describe the nature of intelligence. Some of the theories seek to portray the structure of intelligence as a developed entity while others are much more concerned with the processes of intellectual development themselves. The evidence upon which the theories are based is extremely diverse. Naturally the studies employing intelligence tests have contributed a great deal to some theories. Equally, studies of child development have been as rich in information on the growth of intellectual skills. At the same time biology, medicine, anthropology, ethology and cybernetics are all making valuable contributions,

although their influence upon current psychological theories is, at times, limited.

For convenience the theories can be broken down into four types. The theories arising from the use of conventional intelligence tests - the psychometric approach - appear to be of two types. First, the theories arising from the statistical methods of factor analysis and second, a set of viewpoints on the meaning of intelligence test data that are less clearly identified with a particular set of statistical methods.

The two remaining approaches owe far less to the results of intelligence tests as such and far more to developmental and more broad based ecologically orientated studies. The third set of theories revolves around the work of Piaget and his fellow researchers (e.g. Piaget, 1950; Inhelder and Piaget, 1958) on the stages and processes of intellectual development from birth to adulthood. Piagetian theory and the allied conceptualisations (e.g. Bruner, 1964, 1965) represent attempts to establish logical and, as far as possible, empirically validated descriptions of the internal mental operations associated with intelligence at each developmental stage.

The fourth approach is the far less ordered collection of cognitive-style theories. These theories suggest that particular patterns of thinking, or styles of mental operation, are linked to identifiable differences in socialisation, ecological and cultural circumstance. An example would be Cohen's (1969) 'relational' versus 'analytical' thinking model in which she links the acquisition of a dominant or preferred mode to particular cultural and social class backgrounds.

In as much as the major research aspects of this thesis are concerned with the measurement of intelligence from a psychometric standpoint, further detailing and comment has been limited to theories stemming from this approach.

The theories of intelligence arising from the use of factor analytic procedures are most prominently identified with the work of Spearman,

Thurstone, Guilford, Vernon and Cattell. A relatively brief account of these factor analytic theories is presented, but is done so subsequent to a cautionary note on the use of factor analysis itself.

The danger of reifying the concept of intelligence has been mentioned already and this danger is possibly increased in studies using factor analysis. Dockrell illustrates this point when he says that -

"..... this mathematical technique does not speak to the issue of the validity of a particular concept of intelligence. All it does is to make use of one of a particular group of mathematical procedures to arrive at a simpler set of hypothetical tests or factors, taken to underlie performance on a wider range of more complex real tests" (Dockrell, 1970, p.5).

He takes up the earlier arguments of Thomson (1950) and Guilford (1967) and stresses the fact that a psychological factor cannot automatically be inferred from a mathematical factor and the "mere existence of a mathematical factor does not speak to its psychological utility" (Dockrell, 1970, p.6). This should not be forgotten.

Spearman (1927) introduced factor analysis to psychology and with it his two-factor theory of intelligence which he proposed to account for the patterns of correlations among group tests of intelligence. The theory asserts that a general factor ('g') and one specific factor ('s') per test can account for performance on intelligence tests. In Spearman's view any intellectual activity involved both a general factor which was shared with all other intellectual activities and a unique specific factor which it shared with no other intellectual task.

Spearman's theory has been criticised in texts on intelligence (e.g. Butcher, 1968) for being such a beautifully simple picture that it could not possibly be true. It has also been criticised by other factor theorists who, in their turn, have been responsible for radically different theories. Spearman in his later years conceded that the theory failed to meet all the observed facts, but maintained that it still accounted for the most common and important ones (see Spearman and Wynn Jones, 1950). The critics still disagree, but as Butcher comments -

". . . . his basic concept of general intelligence, which following Galton, he made the main plank in his programme, has weathered the storms of adverse criticism more satisfactorily than many other plausible-seeming concepts in the short history of psychology" (Butcher, 1968, p.46).

The most cogent opposition to Spearman's two-factor theory of intelligence came from Thurstone (1938) with his multiple factor theory. Employing a somewhat different factor analytic procedure to Spearman in the analysis of test data, Thurstone identified seven important group factors which he called the 'primary mental abilities'. These were: verbal meaning, number facility, inductive reasoning, perceptual speed, spatial relations, verbal fluency and memory. Tests were developed to measure each of these factors in as pure a form as possible.

While Thurstone's multiple factor theory did at first seem to eliminate 'g' as a significant component of mental functioning, the primary factors were found to correlate moderately among themselves. This led Thurstone to postulate the existence of a more inclusive second order factor which may be related to 'g'.

The contrasting views of Spearman and Thurstone have become enshrined in modern psychology. Spearman and his hierarchical model of intelligence has set the pattern for the 'British school of thought' on the subject in much the same way as Thurstone with his multiple factor theory has set the pattern for the 'American school of thought'.

Butcher has commented on this trans-Atlantic divergence in these words -

"American psychologists, since the earliest empirical work on abilities have, in contrast to their British counterparts, tended to favour an 'oligarchic' or 'anarchic' view (in Spearman's language) rather than a 'monarchic' one. It is tempting to ascribe the difference in approach to pre-suppositions arising from national temperament and national culture-pattern Be this as it may, there is no doubt that on the whole theories of general intelligence as the prime mover, perhaps modified or 'perturbed' by subsidiary influences, have flourished in Britain, whereas the generally preferred picture in the United States has been of multiple abilities of more or less equal status and influence" (Butcher, 1968, p.51).

The multiple factor approach has been carried forward by Guilford (1956, 1959a, 1967) and his associates at the University of Southern California. Guilford has developed the cuboid Structure of Intellect model as a way of organising the components of intellect into a system. The model was presented as a kind of 'periodic table of the mind', reputedly derived from intuition and the speculation that intelligence was a complex multi-dimensional affair.

A three-way classification of the components of intellect according to their properties under the broad categories of contents, products and operations has been proposed. Thus, intellectual tasks can be understood by the kind of mental operation performed, the type of content on which the mental operation is performed and the resulting product. The model lists five different kinds of operations (cognition, memory, divergent thinking, convergent thinking and evaluation), four types of content (figural, symbolic, semantic and behavioural) and six products (units, classes, relations, systems, transformations and implications). Thus 120 cells ($5 \times 4 \times 6$), or factors as Guilford refers to them, are postulated, ninety-eight of which Guilford and Hoepfner (1971) claim to have been empirically demonstrated.

There are arguments both for and against Guilford's complex classificatory schema. Butcher sees advantages stemming from its systematic and seemingly scientific appearance in contrast to "..... much research in the field which has rightly been criticised as blindly operational and undirected by theory" (Butcher, 1968, p.60). It is also noted by Butcher that Guilford's theory and tests have stimulated new and more varied approaches to the study of intelligence.

The major disadvantage with the schema as noted by Butcher is the absence of demonstrated predictive validity for the many tests covering the multitude of factors. Hunt (1961) reviewing multiple factor tests and theories has gone so far as to state that "..... in no situations are the minute, splinter like factors of predictive significance" (Hunt, 1961,

p. 301). Butcher has summarized the thrust of the criticisms of the Structure of Intellect model as follows -

"Undoubtedly Guilford goes to extreme lengths in extracting a very large number of factors that are at the opposite end of the scale from 'general intelligence' in terms of generality, and apparently also at the opposite end in predictive power" (Butcher, 1968, p.60).

A continuing problem with the Structure of Intellect model, Guilford's efforts notwithstanding, has been the difficulties it presents test constructors. While Guilford and Hoepfner (1971) have indicated how test items might be classified, the relationships between the various factors and the behaviours to be summarized or predicted need to be illustrated more clearly to guide the development of tests of the respective factors. In addition it does seem that for the development of tests for ordinary everyday use, the model is over-precise to the point of impracticality. Not surprisingly therefore, modifications to the model have been proposed.

Adcock and Webberley (1971) have offered a modification to the Structure of Intellect model involving its reduction to a more manageable 24 cells. They have suggested a two dimensional matrix consisting of three forms of operation (spatial, verbal, symbolic) and eight types of content (short term memory, long term memory, perceptual speed, fluency, speed of closure, flexibility of closure, reasoning, insight). While possibly a valuable reduction in the number of the theoretically independent aspects of intellect psychologists might wish to consider, the task of providing a satisfactory measure of each skill remains, for the present, beyond the capabilities of tests designed for general use.

Humphreys (1962) commenting on the proliferation of factors thought to be involved in human intelligence noted, with a measure of irony, that Thurstone, who first proposed a multiple factor model and believed the whole domain to be well sampled by seven tests, has now been overtaken by models advocating many more factors than he originally had tests!

The classic reappraisal of the general versus multiple factor

theories of intelligence is still however, contained in McNemar's (1964) paper 'Lost: Our Intelligence? Why?' McNemar has traced the development of the two views under the guidance of Spearman in Britain and Thurstone in America. He has commented, somewhat scathingly, on factor analytic procedures and then in some detail reviewed first, the development and utility of multiple aptitude test batteries and second, the contribution of general intelligence in a number of fields of creative endeavour. Of immediate interest are McNemar's views on the general versus multiple factor debate.

McNemar recounts the position of Spearman with his concept of general intelligence 'g' and that of Thurstone with his primary mental abilities. Noting that these theoretically independent primaries inter-correlated, McNemar points out that Thurstone had to admit that "..... a general factor was needed to explain the interrelatedness of the primaries" (McNemar, 1964, p.872). A resolution of sorts to this difference came by way of Thurstone's method of factor analysis whereby oblique axes represented the primary mental abilities as first-order factors while the more pervasive general factor was termed a second-order factor. For McNemar -

"It began to look as though Spearman was being revisited, except for the little matter of labelling; anything called second-order could not possibly be regarded as of much importance. Furthermore, it could always be said that, in the ability domain, it is less difficult to attribute psychological meaningfulness to first-order than to second-order factors so why pay much attention to the latter? Thus it was easy for most American factorists to drop the concept of general intelligence and to advocate that tests thereof, despite their proven usefulness over the years, should be replaced by tests of the primaries" (McNemar, 1964, p.872).

As is well known, the British group of factor theorists preferred statistical methods that allowed for the emergence of a sizable general first factor. Group factors could also be obtained and have borne names the same or similar to the primary mental abilities, but these factors have been considered to be less important. Again in McNemar's words -

"Apparently the British are skeptical of the multitude of ability factors being 'discovered' in America. The structure of intellect that requires 120 factors may very well lead the British, and some of the rest of us, to regard our fractionization and fragmentation of ability, into more and more factors of less and less importance, as indicative of scatterbrainedness" (McNemar, 1964, p.872).

In making his point, McNemar has no doubt overstated his case a little. There is the prospect of some systematic theoretical advances from the multiple factor and structure of intellect approaches as Butcher suggests. A prospect which needs to be weighed against the social usefulness of the general intelligence concept when trying to decide whether or not the latter should be retained. It is McNemar's conclusion that -

"..... the concept of general intelligence, despite being maligned by a few, regarded as a second-order function by others, still has a rightful place in the science of psychology and in the practical affairs of man" (McNemar, 1964, p.880).

A rapprochement between the diverging general factor and multiple factor models is suggested however, in the hierarchical theory of intelligence proposed by Vernon (1960, 1961). Vernon's theory has been built upon his own extensive factorial studies and the earlier work of Burt (1949). It is proposed that at the highest level of organisation there is a general intelligence factor or 'g'. This general factor is divisible into two major group factors. One is labelled verbal-educational (v:ed) and the other spatial-mechanical (k:m). The major group factors can in turn be broken down into a series of minor group factors which would include such skills as verbal fluency, numerical abilities, spatial reasoning, mechanical information and so on. The minor group factors are themselves based on a large array of specific factors which depend upon the unique aspects of each test.

Vernon points out that this hierarchical model is at best only a general approximation in which the details, especially at the minor group factors and specific factors levels, are of necessity vague. This is because details are not fully known and in any case the details in

factorial studies will vary according to circumstances. Both Burt and Vernon have stressed the relativity of factor structures and their dependence upon both the nature of the test battery and the characteristics of the sample employed in any factor analysis.

Vernon's model is a relatively open conceptualization when compared with the specificity encouraged in the Structure of Intellect approach. This openness in conjunction with the stress on the relative nature of any factorial structure has at least meant that this model has enabled the general factor theorists and multiple factor theorists to resolve some differences.

Referring to Vernon's model, Butcher has had this to say -

"The hierarchical theory has major advantages over almost all other models of human abilities. It accounts for the proliferation of apparently conflicting findings and the multiplicity of ability factors that have been described and labelled" (Butcher, 1968, p.50).

For Butcher this model has assisted in two important ways. It has helped in resolving confusions that have resulted from comparing factors which were operating at different levels in the hierarchy, especially where there have been important sample and test battery differences. Second, it has given recognition to the fact that different factors, and their associated tests, will have very different roles in assessment and prediction. Thus -

"Other things being equal, measurement of a general factor will account for more variation in performance and provide prediction over a wider range of tasks than measurement of a major or minor group factor. This is of course the fundamental reason for the continued use of measures of general intelligence But for particular families of skills and for greater accuracy in a more limited area it may well be necessary to use tests that measure a major or minor group factor" (Butcher, 1968, pp.50-51).

Cattell (1963, 1971) has also made extensive use of factor analytic procedures in his studies of intelligence test data. His view on the organisation of human abilities is also a form of synthesis of the general factor and multiple factor models. He believes in the importance of a

general intelligence factor, but holds that this is best understood and assessed as a second-order factor through Thurstone's primary mental abilities.

Cattell has also proposed that general intelligence is of two kinds - 'fluid intelligence' ('gf') and 'crystallised intelligence' ('gc') - which are distinct but correlated. 'Fluid intelligence' is seen as a basic capacity for learning and problem solving independent of education and experience. It is general to many different fields of intellectual activity and is employed in tasks requiring adaptation to new situations. On the other hand, 'crystallised intelligence' develops through the interaction of an individual's 'fluid intelligence' and his or her cultural experiences, thus it consists of all learned knowledge and skills.

Cattell argues that both learning opportunities, interests and maturation play a part in the balance between 'gf' and 'gc' within individuals and has attempted to experimentally demonstrate the consequences of his conceptual distinction between the two forms of intelligence (see Horn and Cattell, 1966, 1967). In doing so, Cattell has produced a classification of the components of 'fluid' and 'crystallised' intelligence in an 'Ability Dimension Analysis Chart' as a schema for considering the action phases, content aspects and process parameters associated with activities deemed intellectual.

Having outlined the central features of the major factor theorists it is apparent that the fundamental differences between the viewpoints concern models where a general intelligence factor predominates over and above a number of subsidiary factors as opposed to models which emphasize a set of multiple factors all equally important to a full understanding of intelligence.

It may be helpful to think about the differences and relationships between the models in terms of what is trapped in the 'sieve of analysis and thought' employed by each type of theorist.

The general intelligence, 'g', theorists (e.g. Spearman) use a very coarse sieve with a wide mesh. They are after the biggest rocks they can find - the boulders in fact. This group might be thought of as the Big Rock Theorists.

The multiple factor theorists (e.g. Thurstone) use a moderate mesh sieve. The boulders stay along with a lot of gravel in between and it is this that attracts their attention. This group might then be termed the Gravel Theorists.

Lastly, there are the multi-cell theorists (e.g. Guilford) who use an extra fine sieve. Rocks might be in their sieve, gravel as well, but most of all there is sand and it is that they prefer to play with. This group had best be the Sand Theorists.

Consideration of the factor analytic approach to the study of intelligence opened with a caution from Dockrell. In trying to put the above theoretical positions into some form of perspective, it is appropriate to call on Dockrell again. To the question of whether or not intelligence should be viewed as one broad general ability along with some smaller less important skills, or as a set of seven equally important primary abilities, or even as 120 independent abilities Dockrell gives this answer -

"If factors are thought of as convenient generalisations, the question is not whether this is one ability or many, but which model is useful in a particular context for a particular purpose"
(Dockrell, 1970, p.6).

Alongside the factor analytic theories of intelligence there are a number of viewpoints which have arisen from consideration of data from psychometric studies, but which have not been so dependent on factor analysis itself.

Thorndike (1927) took the view that intelligence consisted of a multitude of separate elements, each representing a distinct ability. A number of mental activities however, had elements in common which combined to form clusters. He identified three important clusters.

Those concerned with dealing with things making up 'concrete intelligence', those concerned with dealing with verbal and mathematical symbols called 'abstract intelligence' and those concerned with dealing with people which he called 'social intelligence'.

Burt (1955) has already been mentioned in connection with attempts to define intelligence. Collectively the evidence from many studies, including his own, led him to the conclusion that intelligence was in the nature of a general factor and largely innate. Evidence for the high heritability of intelligence came from the pattern of correlations obtaining between intelligence test scores and various categories of kinship. Burt, up until his death, continued to publish papers reviewing data on the inheritance of general intelligence making empirical, statistical and methodological contributions (see Burt, 1958, 1969, 1971, 1972).

Kamin (1974) has recently raised serious questions as to the scientific validity of many kinship studies. Burt's work in particular was found wanting in view of statistical inconsistencies between publications and poor reporting of critical detail. Nonetheless, despite this controversy Burt's contention that genetic factors may have an important bearing upon general intelligence cannot be lightly dismissed. Jensen's (1972a, 1973a) reviews which synthesize the evidence from psychology and related disciplines remain in support of Burt's proposition. Furthermore, Vandenberg (1962) and Bock and Vandenberg (1968) using multiple factor measures such as Thurstone's Primary Mental Abilities and the Differential Aptitude Tests have, by the twin-study method, obtained high heritability estimates for a number of the skills measured by these test batteries.

Viewpoints that contrast sharply with the high heritability claims concerning intelligence have, however, been expressed with both force and conviction. A case in point would be the views of Wesman (1968). Wesman has advanced a strong environmental and learned-behaviour viewpoint on the nature of intelligence. He argues that confusion has arisen over defining

and measuring intelligence because it has been treated as an entity rather than an attribute and because of a failure to appreciate that intelligence is but the summation of all of the learning experiences of the individual.

Wesman contends that for the most part, many of the differently named intelligence tests are measuring similar abilities and all are measuring what the person has learned. So far as factors obtained in factor analytic studies are concerned these, Wesman argues, must be treated as descriptive categories and not reflections of supposed underlying entities - genetic or otherwise.

The most recent and extensive theorizing on intelligence resting upon psychometric studies comes, however, from Jensen (1970,a,b,c,d, and 1972a). As a result of his studies on memory and intelligence and surveys of the literature on the assessment of intelligence, Jensen has synthesized a number of important issues and evolved a theory of mental functioning as an explanation of the findings.

Jensen (1970c) considers intelligence to be an attribute of persons much like temperature and in the same way as the thermometer enables the objective measurement of temperature, the intelligence test enables objective measurements of intelligence to be made. However, unlike temperature, intelligence is multi-dimensional and therefore, people will be ranked differently on different kinds of intelligence tests. He suggests that from a psychological point of view this multi-dimensional view of intelligence must include the primary abilities of conceptual learning ability, abstract or symbolic reasoning ability and abstract or verbal problem solving ability. While these abilities account for only part of the range of intellectual abilities, they are those most frequently emphasized in intelligence tests because they predominate in most educational and occupational activities. It is through the similarity of composition of various intelligence tests that the observed high intercorrelations amongst such tests occurs leading to the large common and general factor identified

Jensen points out that while the intelligence quotient is related to socially valued criteria, noting the positive correlations between measured intelligence and scholastic success, upward social mobility and occupational status, the data on the predictive validity of an intelligence test score is much less accurate for an individual than for groups. The effects of unassessed traits or unpredictable and unusual future circumstances may radically alter both the pattern and direction of an individual's intellectual development. There is therefore, good reason for caution in the psychological assessment of individuals and especially in attempting to predict future intellectual accomplishments.

The concepts of intellectual 'breadth' and 'altitude' have been suggested by Jensen (1970b) to help in understanding the content of various intelligence tests. General information and vocabulary type tests are said to measure intellectual breadth while those of a problem solving nature measure intellectual altitude. Intellectual breadth depends on the amount and range of an individual's exposure to learning opportunities in conjunction with interests and values. The altitude dimension is seen to be much more dependent upon innate endowment. Breadth and altitude measures are highly correlated with the concept 'g', but it is primarily the test content reflecting intellectual breadth that comes under extensive environmental influence.

Concurrent with the views illustrated above, Jensen (1970 a and d) has proposed a two-level theory of mental abilities. First, there is Level I ability or 'associative learning ability'. This is "essentially the capacity to receive or register stimuli, to store them, and to later recognise or recall the material with a high degree of fidelity" (Jensen, 1970 d, p.154). Level II ability or 'conceptual learning and problem solving ability' in contrast "is characterized by transformation and manipulation of the stimulus prior to making the response" (Jensen, 1970 d, p.155).

Both abilities are seen as being based upon underlying genetic

processes, but these are probably different in nature. Thus, while the abilities may be correlated, their rate of development may vary. In addition, while the conceptual learning ability is partly dependent upon the associative learning ability, the reverse does not hold.

It is Jensen's contention that current intelligence tests measure both abilities to varying degrees. Associative learning abilities would be measured by tasks such as digit span, free recall, serial learning and paired associate learning. Conceptual-learning abilities are tapped by the more complex abstract reasoning tasks such as mathematic items involving transformations within a defined symbol system, matrix items and block design items. He argues that what are needed are tests that reliably measure the separate Level I and Level II abilities.

Jensen has written extensively on the role of socio-economic factors and 'cultural disadvantage' upon intellectual development employing his two-level theory. Associative learning abilities are regarded as being under the control of genetic mechanisms and largely independent of socio-economic influences. As a result the hypothetical growth curves for Level I abilities for children from middle and lower socio-economic groups would be almost identical. Conceptual learning abilities, while based to a degree upon genetic mechanisms, require experience and learning for their development and manifestation in performance. The hypothetical growth curves for Level II abilities for children from lower and middle socio-economic groups are seen as progressively separating with the widening difference in development favouring children from the middle socio-economic backgrounds.

Jensen has advanced this theory as a possible explanation of the progressively poorer scholastic performance of children from low socio-economic backgrounds relative to their middle-class peers as they proceed through school. It is argued that -

"As the content of the school's curriculum becomes increasingly abstract and conceptual with advancing grades, the child with below-average Level II ability, regardless of his status on Level I, will be at an increasing disadvantage. The cumulative deficit effect will then snowball because of the child's

discouraging experience of diminishing returns from his efforts in school"(Jensen, 1972a,p.239).

While Jensen has brought together considerable evidence in support of his theory, it has not passed without criticism. Humphreys and Dachler (1969a) in one empirical test of the theory have found measures of Level I type abilities to be positively correlated with socio-economic status and I.Q. in four samples established on the basis of high or low socio-economic status and high and low measured intelligence using Project TALENT data. These relationships with Level I type abilities are contrary to Jensen's theory and the interaction terms between socio-economic status and intelligence obtained in this study were smaller and differed in direction. It was also argued that an inappropriate experimental design has been used in Jensen's own empirical studies of the theory.

Jensen (1969b) has replied to the criticisms suggesting that the data was unsuitable, that the design question^{was} open to debate and that other data supported the theory. In a rebuttal, Humphreys and Dachler (1969b) assert that their data was adequate, that the design question has an important interpretative bearing upon the theory and that certain additional data cited by Jensen does not apply to the theory itself. An important theme in the criticisms of Humphreys and Dachler concerns the generality of Jensen's theory. They suggest that the theory may explain particular fixed effects, but that this in itself might not constitute evidence of a general theory of intelligence, applicable to the total population.

Criticisms notwithstanding, Jensen's theory and related writings have acted as an important stimulus for the fields of psychology and education. For his part, Jensen has proposed that educational curricula and teaching methods are in need of revision to more adequately cater for children from low socio-economic backgrounds with below-average conceptual learning and problem-solving ability. This would be achieved by providing instruction more in accord with the associative learning processes rather than the conceptual learning and problem-solving processes currently emphasized.

While Jensen's educational prescriptions have been seriously challenged by sections of the educational community for providing a justification for continued segregation and discrimination in education (see Brazziel, 1969; Cronbach, 1969) and for ignoring the evidence on 'developmental plasticity', to use Hunt's (1969) term (see Voyat, 1970), he has brought into the open issues critical to both education and psychology. Perhaps more than any other theorist in this field of late, Jensen has made an effort to address the question of just how tests of intelligence might lead to an understanding of mental ability itself in a manner bearing more directly upon teaching methods and educational processes.

Theories aside for the moment the problem of giving meaning to the term intelligence still persists. Fortunately, Hebb (1949) and Vernon (1955) have assisted by describing three different meanings associated with the word intelligence. On the basis of research on neural processes, Hebb distinguished Intelligence A from Intelligence B.

Intelligence A refers to the innate capacity of an individual. It is genotypic and can not be measured directly. It is therefore inferred. The level of Intelligence A for an individual would be lowered by deficiencies in the general pattern of biological development and in genes relevant to special aptitudes and abilities.

Intelligence B refers to the observed behaviour of an individual. It is phenotypic, resulting from the interaction of genes with the pre- and post-natal environment. It can be impaired by constitutional factors such as brain damage and by environmental factors, such as limited satisfaction of both biological and social needs.

Hebb does not view Intelligence A and B as being totally separate and independent of one another. Intelligence A enters into and is a necessary component of Intelligence B.

The intelligence test as we know it is a method for making systematic observations of behaviour and Vernon has added the term Intelligence C to

refer specifically to intelligence test scores, since a formal test can only include a limited sample of skills that would be part of Intelligence B. It stands for the sample of behaviour obtained, the actual measurements made with a specific set of test items, that may be employed in inferences about Intelligence B. This is a deliberately restrictive meaning which would differ from what most people regard as intelligent behaviour. For Vernon, Intelligence C depends upon the method of assessment, the test and its content. Consequently, a test may be more or less pertinent to inferences about Intelligence B depending upon environmental and cultural circumstances. Vernon (1969) has used this Intelligence A, B and C distinction effectively to present a clear account of both the research and emerging theory relating to comparisons of ability test performances across cultures.

It could be argued that Hebb and Vernon have added a further element of complexity to an already confused domain of psychology. Alternatively, their contribution can be seen as a long overdue and much needed clarification.

The Cross-Cultural Study of Intelligence

Interest in the comparative performances of various human groups deemed indicative of aspects of intelligence is as old as the psychological test itself and arose well before the philosophical and empirical considerations that have led to the modern theories of intelligence. While ancient China was making considerable advances in the field of individual assessment, it would seem that final examinations were much preferred in Palestine where perhaps the first cross-cultural comparison also took place. Thus, it is recorded that -

"..... the Gileadites took the passages of Jordon before the Ephraimites: and it was so, that when those Ephraimites which were escaped said, Let me go over; that the men of Gilead said unto him, Art thou an Ephraimite? If he said

Nay: then said they unto him, Say now Shibolet; and he said Sibboleth; for he could not frame to pronounce it right. Then they took him, and slew him at the passages of Jordon; And there fell at that time of the Ephraimites forty and two thousand" (Judg. 12:5-6, Holy Bible, King James Version).

Times and psychology have changed a little since then and with the development of psychometrics in the last hundred years, tests of a good many psychological functions have been applied to subjects from extremely diverse social and cultural backgrounds. To realistically deal with the many issues involved in the application of tests of intelligence in cross-cultural contexts it is necessary to be selective and to impose some structure upon the discussion.

First, the major theories or viewpoints lying behind the issue of intellectual differences or similarities across culture groups are commented upon.

Second, consideration is given to cross-cultural studies of intelligence of a psychometric nature. The development of these studies is covered relatively briefly since a number of reviews are available. Attention is then given to the general developments taking place in ability test adaptation and administration when cross-cultural application is envisaged. This is followed by a discussion of some of the theoretical perspectives and revisions that have taken place in cross-cultural psychology of late which have an important bearing upon the interpretation of comparative psychometric data.

Third, a short statement on the 'race and intelligence' issue is made in the hope of clarifying the evidence and the relationship of the issue to the research reported in this thesis.

Theoretical Viewpoints

Historically, it is difficult to identify any one theoretical orientation as being the most influential in the cross-cultural study

of intelligence. Cole and Scribner (1974) taking a wide frame of reference point out that "Over the centuries, sociologists, anthropologists, philosophers, linguists, and psychologists have all put forward theories linking culture and mind" (p.12).

In a succinct account of developments and research in this field they note that interest in cultural influences on intellectual abilities began understandably with concern over differences. European exploration and colonisation of new lands brought with it continuing and intense contact between different cultures posing problems for the developing bureaucracies and the new social scientists.

"Under such circumstances, concern with cultural differences all too often took the form of comparisons between 'them' (the 'uncivilised' in the colonies) and 'us' (the 'civilised' in the mother countries)" (Cole and Scribner, 1974, p.13).

Explanations of the observed differences in such cross-cultural comparisons relating to intellectual tasks have been advanced by a number of disciplines, although their impact has varied in conjunction with evidence and circumstance. Cole and Scribner have identified biological, sociological and psychological types of accounts.

Biological accounts of intellectual differences between cultural groups had their origins in the evolutionary theories of Darwin and Spencer promulgated in the latter half of the nineteenth century.

"Spencer held that during the course of man's experience he acquires certain mental traits that favor his continued existence and are passed down from generation to generation" (Cole and Scribner, 1974, p.14).

This view was widely held in scientific circles at that time and through such aphorisms as "ontogeny recapitulates phylogeny", meaning that a child's development passes through equivalent stages to those traversed by the human species during its evolutionary development, evidence concerning the 'childlike' intellectual abilities of primitive

peoples gained easy acceptance.

This biological orientation manifested itself "..... in the ubiquitous identification of cultural differences with racial differences" (Cole and Scribner, 1974, p.17), and served doctrines of racial determinism. The fledgling social sciences, in their attempts to measure intellectual abilities and evaluate cultures, added weight to popular prejudices.

This century has seen important changes in both biological and social theory such that, "No one today would seriously maintain that 'mental peculiarities' caused by habit become organic and hereditary" (Cole and Scribner, 1974, p.18). Furthermore, the rate and extent of social change among nations and social groups in the last seventy to one hundred years has highlighted the inadequacies of social evolutionary theories based on natural selection.

The equation of race with culture, which had acted to support biological accounts of difference in intelligence, has long been the subject of debate. In their survey of this issue Cole and Scribner have pointed out that -

"The classic attack on the identification of race with culture was made by Franz Boas in The Mind of Primitive Man more than fifty years ago. At the end of a long survey of the historical antecedents of modern societies, Boas reached the conclusion that there is no foundation for the equation of race and culture" (Cole and Scribner, 1974, p.18).

More recently, Herskovits (1965) in a reappraisal of the work of Boas, in conjunction with developments in anthropology and world events, came to the conclusion that the very concept of race represents a scientific dead-end in explaining cultural phenomena.

Comparisons between the thought processes of the 'primitive' peoples and the Europeans were also of interest to the early sociologists. The concept of the social collectivity played a central part in the

writings of Comte and Durkheim and was held to shape an individual's character and behaviour. Using this framework Levy-Bruhl set about studying mental functioning, contrasting 'primitive' and 'civilised' forms.

In his view, cultures could be best described in terms of the general beliefs or "collective representations" which regulated the thought processes of the individual belonging to each particular group. For the European these collective representations imposed a sharp distinction between intellectual, motoric and emotional aspects of thought and behaviour. This was not the case, however, with the 'primitive'.

"Their mental activity is too little differentiated for it to be possible to consider ideas or images of objects by themselves apart from the emotions which evoke these ideas or are evoked by them" (Levy-Bruhl, 1910, p.23).

In addition, Levy-Bruhl held that primitive thought was "prelogical" in the sense that it need not of necessity avoid contradiction which he took to be a singularly important condition of European thought.

Levy-Bruhl's views have been seriously challenged on a number of counts. Boas (1911) questioned seriously the ethnographic material used by Levy-Bruhl and also criticised the approach adopted whereby the traditional beliefs and customs of a people were used as the basis for inferences about thought processes.

Psychologists were also critical of Levy-Bruhl's propositions. Bartlett (1923) thought it fallacious to compare 'primitive' thought with a select and limited mode of scientific thought maintaining that there is much in common between the mental functioning of the ordinary members of both 'primitive' and 'modern' societies.

In recent years one of the most telling challenges to Levy-Bruhl's position was Horton's (1967) analysis of traditional African belief

systems in relation to the logic of Western scientific thought. He has argued that the similarities in these modes of thought were much greater than Levy-Bruhl recognised. From his studies, Horton proposed as a basic premise that -

"..... all peoples try to understand their world by developing explanatory theories (and he proposed) compelling analogies between the theories that underlie traditional African belief systems on the one hand, and Western so-called scientific beliefs on the other"(Cole, Gay, Glick and Sharp, 1971, p.7).

Criticisms on psychological grounds also came from Köhler (1961) who, from his Gestalt psychology background, described how instances of 'primitive' mystical thinking and animistic beliefs cited by Levy-Bruhl could be accounted for by some of the same dynamic principles of perception established with European subjects. Consequently, Köhler viewed it as unnecessary to postulate the existence of fundamentally different thought processes between human groups.

Psychological accounts of intellectual differences between culture groups have also been advanced. Werner (1948; Werner and Kaplan, 1956), a developmental psychologist, has characterized non-Western thought in a manner reminiscent of Spencer and Levy-Bruhl. To Werner there are orderly and directional changes in mental functioning that occur both between and within species. The stimulus for these changes is not, however, a strict biological or social mechanism, but a general concept of development which rests on the assumption that life means growth and development which must take place in a systematic orderly sequence with increasing differentiation, articulation and hierarchic integration. Rather than suggesting an identical recapitulation of developmental sequences in animals, children and cultures, however, Werner only points to parallels.

In the sphere of intellectual functioning, he refers to primitive mental activities which are evidenced in the failure to differentiate,

to articulate and to form hierarchic integrations of cognitive material. Thus, abstract thought is not achieved. Using evidence similar to that of Levy-Bruhl, Werner believes that pre-literate primitive peoples, children and psychiatric patients all display these failures. They are all, therefore, limited to his primitive mental activities.

Little in the way of cross-cultural research has come from Werner's views, probably because the hypothesis that all development is the unfolding of a regulated and presumably genetically controlled sequence is of little assistance in studying the relationships between intellectual abilities and cultural experience.

A more productive psychological theory on intellectual differences between culture groups has been propounded by Bruner (see Bruner, Olver and Greenfield, 1966). Bruner's view on intelligence is somewhat similar to Wesman's in that intelligence is seen as largely being the internalisation of the "tools" of the particular cultural system experienced by the individual. The tools include both the technologies and the symbol systems of the culture in question. Cultural differences on tasks requiring the exercise of intellectual skills are the result of differences in the nature and range of tools and social institutions developed to preserve and transmit knowledge and tool-using skills within cultures. To Bruner an institution with considerable impact upon intellectual development is the school which in its Western form has structured learning in such a way that it is largely divorced from the practical everyday activities pursued by the members of a culture group. Under these conditions language, especially in the written form, becomes the medium for analytic processes requiring abstract modes of thought.

Over a number of years Bruner and associates (Bruner, Olver and Greenfield, 1966) have pursued programmes of research in West Africa and the United States investigating aspects of intellectual development in relation to cultural circumstance. The research has led Bruner to

the conclusion that -

"..... some environments 'push' cognitive growth better, earlier and longer than others. What does not seem to happen is that different cultures produce completely divergent and unrelated modes of thought" (Greenfield and Bruner, 1969, p.654).

It is this environmental 'push' which is held to be responsible for the observed differences in the performances of subjects from different cultural groups. As Bruner rightly points out, any decision on whether or not to judge such "differences on some universal human scale is a matter of one's values" (Greenfield and Bruner, 1969, p.654).

Psychology has always had close links with the biological sciences so it is perhaps not surprising that differences in intelligence test performance between culture or racial groups have been often interpreted as being the result of genetic differences. This has occurred despite debates over the utility of the concept of race in the social sciences (see Montagu, 1969), persistent questions concerning the adaptation of psychological measures for use in differing cultural contexts (see Brislin, Lonner and Thorndike, 1973) and the many methodological issues in understanding and accounting for environmental effects in comparative studies (see Stinchcombe, 1969). Then too, there are the not unimportant issues of the appropriateness of both the data and the heritability models employed in studies supporting racial genetic factors as explanations for observed group differences in intelligence test performances (see Fehr, 1969; Feldman and Lewontin, 1975). Contemporary views on the 'race-intelligence' debate are discussed in brief below. For the present it is sufficient to note, however, that assumed genetic differences have been invoked by psychologists from time to time in order to explain both differences in the levels and patterns of intelligence test performances of racially identifiable groups of subjects. In their turn, these 'explanations' have been used in support of arguments concerning the racial inferiority

of particular human groups. Within psychology perhaps the clearest example comes from the writings and activities of Henry Garrett.

In his text General Psychology Garrett (1955) demonstrated his rather simplistic and indeed erroneous views about the nature of heritability with his illustration of the family tree of Martin Kallikak. This was included to show the dysgenic and eugenic consequences of 'dallying with a feeble-minded tavern girl' and of also 'marrying a worthy Quakeress'. Given these views on the process of genetic inheritance it is little wonder that Garrett (1962, 1964) went on to assert that the race differences in intelligence between American Blacks and Whites were so persistent and regular as to convince him of their genetic base. In his declining years Garrett has turned his attention to the popular press. It is reported by Brazziel (1970) that he is -

"..... now winding up his career as tout and chief pamphleteer of White Citizens Council type organisations such as the Patrick Henry Press of Richmond, Virginia. One of the more recent productions of this firm was a Garrett pamphlet that presented test scores to 'prove' racial inferiority" (p.7).

Alongside theories or viewpoints that try to account for cultural group differences in intellectual ability and cognition, there are a set of theories and perspectives that stress instead features of human intellectual activity that are seen to be universal in character.

In the main, contemporary anthropological views on cognitive or intellectual processes tend to emphasise the universal features of such processes across culture groups. The obvious differences among cultures does, of course, seem to suggest that there must be underlying intellectual differences. However, anthropologists following the 'structuralist' school of thought pioneered by Lévi-Strauss (1963) consider the observed cultural differences or dissimilarities to be simply different manifestations or expressions of underlying cognitive processes and structures that are common and universal.

In his studies into belief systems in a number of cultural contexts Lévi-Strauss (1966) has attempted an analysis of the similarities and differences in the category systems employed by various cultures in attempting to comprehend the world. The evidence according to Lévi-Strauss is that the thought processes identified in 'primitive' cultures do not occupy some lower stage in the development of human intellect.

"Rather, he suggests, they represent different strategies by which men make nature accessible to rational inquiry. Both Western and non-Western strategies seek objective knowledge of the universe; both proceed by ordering, classifying, and systematizing information; both create coherent systems. These and other similarities have led Lévi-Strauss to conclude that the two types of thought systems are based on 'the same sort of mental observations'" (Cole, Gay, Glick and Sharp, 1971, p.8).

The differences between the two thought systems lies for Lévi-Strauss in the actual material, the content, the attributes used in forming classes of objects, not in the underlying intellectual processes. The contribution of Lévi-Strauss lies not in any direct study of possible psychological processes within the individual, but in his attempt to demonstrate the more universal aspects of human intellectual activity. As Cole and Scribner (1974) suggest -

"His principal significance for the study of culture and cognition is his demonstration that ethnological material embodying the endlessly varied products of many cultures still testifies to common underlying human operations" (p.27).

Support for generalized and universal features to intellectual functioning can also be found in recent developments in the field of linguistics. The science of language has in recent years stressed the importance of the structural features of language that are held in common by all languages. The complexity of all language systems is evident through this approach which has in turn countered the view that languages could be graded upon a scale of increasing complexity. Furthermore, the belief that the cognitive structures of a linguistic

group could be understood via the comparative analysis of the vocabularies of languages could not be maintained. Consequently, the view that cognitive or intellectual differences are a concomitant of linguistic (cultural) differences has become more difficult to sustain.

This revolution in the linguistic viewpoint has been brought about principally by Chomsky (1968) through his theory of grammar. The theory proposes that any and every sentence arises from the application of a set of basic human linguistic components in conjunction with a complex system of learned rules governing the grammar of a language. Thus -

"Any human speaker who is competent in any human language, according to this theory, must store and use productive rules in a complex and nonmechanical fashion. The implication of this approach is that the thinking processes of an individual cannot be less complex than the rules required for his speech production" (Cole and Scribner, 1974, p.27).

These linguistic propositions have been important in counter-acting the tendency to describe the thought and intellectual capacities of non-Western and pre-literate peoples almost exclusively in terms of deficiencies. The linguistic and anthropological approaches that stress human universals in cognitive processes have not, however, focused upon individual and actual cognitive operations or intellectual capacities. Understandably, it has been theories of a more psychological nature that have addressed this issue.

Of the psychological theories stressing underlying universals to the growth and organisation of intelligence, Piaget's (1950) theory must be given a measure of prominence. The theory postulates the existence of a set of successive logical structures governing cognition which are developed through a continuous interplay between the environment of the individual and his/her development throughout childhood and adolescence. For Piaget -

"The characteristics of these structures and their order of appearance are considered to be universal. They are the outcome of adaptive processes between human organisms, whose biological heritage is the same the world over, and environments, whose fundamental physical properties are identical" (Cole and Scribner, 1974, p.28).

The theory suggests that cultural factors act to accelerate or retard the developmental process thus bringing about age variations in the appearance of the successive stages of cognitive development and their associated operations. In the light of cross-cultural data bearing upon the theory, Piaget (1972) has recently suggested that the impact of cultural factors in some contexts may be insufficient to bring about the final stage of cognitive development of formal, propositional thinking. This suggestion has not, however, undermined the universalistic nature of the theory.

Piaget's theory has generated a considerable amount of cross-cultural research, especially at the period of concrete operations and on the topic of conservation as Dasen's (1972) review shows. As far as one can generalise from the cross-cultural research the stages of development and the sequencing as postulated by Piaget are supported. What has not been confirmed, and indeed at times refuted, are propositions asserting that -

"..... (1) all normal individuals in all cultures pass through all stages, and with regard to every form of cognitive task. Appropriate environmental support or stimulation appear to be necessary for the achievement of certain levels of development; (2) the ages for reaching stages conform to the Genevan norm; and (3) the order of development is invariant, for example, reaching conservation of quantity before weight, and weight before volume" (Brislin, Lonner and Thorndike, 1973, p.172).

Furthermore, as Cole and Scribner (1974) note, there is the possibility that Piagetian theory itself may not in fact address itself to universal cognitive structures, but may simply be an attempt at a universal theory built upon the logical structures of thought as conceived of in the Western intellectual tradition.

Inasmuch as some psychologists have been prepared to interpret intelligence test performance differences between culture groups as evidence in favour of genetic differences, others have subscribed to the view that there are more likely to be underlying universals with regards intellectual ability. For these psychologists the manifest differences on measures of intelligence are the result of cultural and environmental circumstances that may also be compounded by both conceptual and measurement problems associated with the concept of intelligence itself.

Vernon's (1969) cross-cultural psychometric studies reported in his text Intelligence and Cultural Environment reflect the latter approach. On the basis of studies conducted in the United Kingdom, North America, East Africa and the West Indies Vernon concluded that notwithstanding the inadequacies of many measures of ability, especially when employed cross-culturally, the results tended to confirm 'everyday observations'.

"Each group certainly shows variations in patterns of abilities..... But the average performance on quite a wide range of tests only too strikingly fits in with the observed inequalities of mankind" (Vernon, 1969, p.213).

The inequalities Vernon has in mind are not, however, primarily genetic. He concedes that such differences may exist,"..... but (that) they are probably small and we have no means of proving them" (p.215). What Vernon is at pains to point out are the great variations in environmental circumstance between culture groups that influence intellectual development and thus performances upon particular measures of intelligence. Pre- and post-natal factors influencing intellectual development have been reviewed in some detail by Vernon. His account covers health and nutritional conditions, perceptual and sensory-motor development, language acquisition, child-rearing and parenting patterns, schooling and more general societal and cultural factors.

Although Vernon does not specifically align his studies and review with the views expressed earlier on the universal character of human intellectual activity, this attitude is present in his work. The demonstration of the importance of cultural and environmental influences upon intellectual development and performance can be seen as an argument in favour of a set of underlying general intellectual processes whose elaboration and expression are very much dependent upon such influences.

While the work of Vernon has been highlighted in this brief discussion, many others involved in cross-cultural psychometric studies have adopted similar views when interpreting comparative data. Examples would be Biesheuvel (1952), Cryns (1962), Ferron (1965), Irvine (1969a,b), MacArthur (1968), McElwain (1969) and Ord (1970). Again the approach taken when interpreting comparative intelligence test data can be summarised as being one where differences between culture groups on tasks assessing intellectual development and ability arise principally from the cultural and environmental circumstances within which development is taking place and abilities are being applied.

In summary, this approach does not argue emphatically for the presence of a cognitive process or ability amongst one culture group and its absence in another, especially on the grounds of hypothetical genetic factors. Given the explicit or implicit acceptance of underlying cognitive universals then just determining intellectual potential, capacity or the existence of particular cognitive processes amongst culture groups is not the whole issue. Psychologists and others of this orientation would give as much if not more importance to the task of comprehending the environmental, cultural and personal circumstances involved in the development of intellectual ability and the patterning of cognitive processes.

Cross-Cultural Psychometric Studies: Notes on Developments and Issues

Before the turn of the century the anthropometric methods pioneered by Galton and others in Europe and North America were taken by McDougall, Myers and Rivers to Papua New Guinea. The studies formed part of the Cambridge Expeditions under the direction of Haddon which began in 1898 and were the first systematic attempt to understand the psychological functions relating to vision, perception and classificatory behaviour amongst so-called 'primitive' peoples (see Haddon, 1901-1935). The anthropometric approach to the study of cognitive abilities was largely swept aside in psychology however, with the advent of intelligence tests in the tradition of Binet, Otis and others.

As recounted earlier, translations of the Binet tests and the revisions offered a partial solution to the problem of employing the tests in culturally and linguistically different communities. Nonetheless, it was felt that difficulties related to the biasing influence of cultural content needed to be overcome in order to more properly investigate the intellectual abilities of differing cultural and ethnic groups. Consequently, the performance tests of intelligence pioneered by Woodworth, Knox, Pintner and Patterson, Porteus, Kohs, Goodenough, Arthur, Leiter and Raven often came to the fore in many cross-cultural studies. As will be seen, however, the performance test approach has not entirely resolved the many issues encountered in cross-cultural comparative research.

Armed then with a great variety of intelligence tests, researchers principally from Anglo-American backgrounds have, over the past half-century, collected enormous amounts of psychometric data from non-European subjects in the hope of saying something intelligent about the intelligence of non-Western peoples.

Apart from studies of an obviously comparative nature where the performances of two (or more) distinct culture or ethnic groups are contrasted, there are also those that are what might be called 'implicitly comparative'. Here, data collected in some specific cultural context is interpreted against the backdrop of similarly generated data in other cultural contexts. Then, too, there are the comparative studies undertaken within countries evidencing cultural diversity quite possibly as marked as that, say, between the youth of New Zealand and the adolescents of Ambunti in the Sepik River valley. Often the differentiating criteria may be in the nature of social stratification factors as much as ethnic and associated cultural differences.

The problem of setting some boundaries to a discussion of even the cross-cultural psychometric studies of intelligence are considerable. It would seem more appropriate to follow the lead given by Price-Williams (1975). He found the whole field of cross-cultural psychology more easy to characterise than define or to set boundaries upon. Therefore, rather than attempt an exhaustive review of all the potentially relevant studies (an exercise most probably doomed to failure), attention is given instead to the major theoretical and methodological developments that have taken place within this field. Details on a good many of the more central studies can be found, however, in the following texts and papers. Shuey (1966), Jensen (1973a) and Samuda (1975) cover in considerable detail data and issues in this field arising within the United States. Vernon (1969) reports principally upon his studies in Africa, Canada, the West Indies and the United Kingdom but also refers to many other related studies. Ord (1970) in the first instance reports on his research in Papua New Guinea, but his text also includes a review of studies on the assessment of intellectual abilities in Africa, America and Asia. Ord (1972b) in a major monograph on the role of psychological and educational tests in

developing countries for selection purposes has presented a review covering more than 400 sources. The cross-cultural adaptation and use of measures of ability is a central theme of the monograph. Kearney (1966b) has reviewed the great majority of the Australian studies and a review of much of the New Zealand research in this field is included in this thesis (see Chapter 3).

Arising out of the research, and as a direct result of some of the findings and interpretations, there has been a series of important methodological and theoretical developments in the cross-cultural study of intelligence employing psychometric techniques. It has been a relatively quiet revolution that has led in practical terms to establishing certain ground-rules for cross-cultural test adaptation and in theoretical terms to the issues of psychological equivalence and cultural relativism.

The first three to four decades of this century saw the accumulation of a considerable body of psychometric data suggesting that measured level of intellectual ability amongst the non-Western and pre-literate peoples of the world was lower than that found amongst Anglo-American populations. This sort of finding was not just restricted to comparisons made between groups of extreme ethnic and cultural diversity, but was also found in studies contrasting the test performances of sub-cultural and ethnic minority groups within particular societies. Vernon (1962) in summarizing this evidence has noted that "Test scores are undoubtedly lower in non-technological or underdeveloped societies, that is among peoples who are not of Western European or North American white stock" (p.94). The critical questions for Vernon, however, are not those concerning possible existence of racial or national differences in intelligence in the genetic sense which, in his view, has led to "..... a great deal of fruitless controversy" (p.94), but those to do with environment

and cultural circumstances "..... which probably have a much greater influence on observed group differences in intelligence" (p.94). Coupled with this focus upon cultural influences has been the test adaptation movement which has sought to modify and develop many psychological assessment procedures in order to either accommodate to or to 'control' for the effects of cultural differences. This movement has been singularly important in shaping the content of measures of intelligence for use in cross-cultural contexts and has in recent years provoked serious and technically sophisticated discussion on the underlying issue of test bias.

By the early 1940's the usefulness of many of the traditional intelligence tests in situations where obvious and important cultural differences existed between groups was being questioned. A response to this growing dissatisfaction emerged in the culture-free test movement. Existing tests were faulted on the grounds of both underlying theory and content leading to Cattell's (1940) proposal to "..... wipe the slate clean of these earlier results, and begin afresh with sounder tests" (p.162). The sounder tests were hopefully going to be such instruments as Cattell's (1944) Culture-Free Tests of Intelligence or along the lines of the tests arising out of the work of Eells, Davis, Havighurst, Herrick and Tyler (1951) who also attempted to develop measures of intelligence free of cultural influences. It was their contention that once cultural factors had been controlled then there would be no significant differences between the average intellectual ability of children from different socio-economic backgrounds.

At this time, the culture-free test approach was attempting to strip away, as it were, the cultural veneer of a subject in order to reveal his or her true and inherent abilities. The view implicit in this approach was that "..... native intelligence lies buried in pure form deep in the individual and needs only to be uncovered by ingenious mining methods" (Wesman, 1968, p.269).

The persistence of cultural, social and ethnic group differences on the culture-free tests, the growing awareness that the nature-nurture issue was more complex than a simple dichotomy, and the realization that a truly culture-free test was unlikely to ever be achieved led to the reformulation of the movement into one emphasising culture-fairness. The task was now seen as being not one of totally eliminating cultural influences from a test, but rather one of endeavouring to prevent particular cultural factors from permeating through the test. This was to be achieved by establishing the test items only from those experiences, knowledge and skills common to different cultures. In comparison with the conventional intelligence tests, those tests in the culture-fair genre (including the culture-free tests before them) attempted to de-emphasise factors believed to mitigate against the performances of subjects whose cultural, social and ethnic backgrounds were at variance with that of subjects from the mainstream, middle-class and Caucasian communities of the United States and Europe. This was to be achieved through tests of a primarily non-verbal nature, employing simple oral or mimed directions and with materials consisting of pictures, drawings, diagrams and the like that were presumed to have some universal quality through being either equally common or uncommon to various groups of subjects. Speed of performance also tended to be de-emphasised.

Like the culture-free concept, the culture-fair approach was also found wanting. Marquart and Bailey (1955) reviewing Cattell's tests summed up the problem by pointing out that "It is difficult to know whether cultural contexts can be completely removed from any scale. Almost all human reactions are culturally influenced" (p.357). Similarly, Kidd (1962) viewed the culture-free, culture-fair test enterprise as suspect "..... because even methods of manipulation of

material objects are culturally determined" (p.352). General support for the culture-fair concept and discussion of the issues involved in the use of culture-fair tests continued into the mid-sixties (see Anastasi, 1964), although by this time some of the absurd consequences of the concept were being demonstrated, albeit satirically, by commentators such as Bernardoni (1964) with his Culture Fair Intelligence Test for the Ugh, No and Oo-La-La Cultures.

Brislin, Lonner and Thorndike (1973) have summed up this period as one where -

"..... many psychologists were captivated by the fascinating prospects and promises of 'culture free' and 'culture fair' tests. Stimulated largely by the nature-nurture controversy, such tests were deemed necessary to help find universals in mental functions. The quick successes (or fresh approaches) that such tests presented had a brief intoxicating effect, but contemporary psychologists are now correcting the legacy of errors inherited from this very forgivable passion. Researchers are retreating and regrouping, virtually unanimously agreeing that a test unequivocally fair to all peoples in all cultures is a distant hope" (p.109).

Concurrent with the efforts to develop culture-fair tests of intelligence for use in culturally and ethnically heterogeneous countries such as the United States, there has also been a considerable growth in the use of psychological tests in the 'Third World' countries of Africa, Asia and the Pacific. Decolonisation after World War II brought with it new and extensive manpower needs. Selection and placement problems arose in the developing educational systems operating continuously under the stress of often severe physical, financial and human resource limitations. In many such situations psychological tests of ability have been pressed into service (see Ord's (1972b) extensive review) and in the process many principles have evolved for the adaptation, application and interpretation tests of intelligence in non-Western settings.

On a number of counts, there has been a convergence of views and information from these two strands in the development of measures of ability for use with subjects from culturally different contexts. The lessons learnt from the culture-free, culture-fair test movement and from the efforts to adapt tests for use in numerous non-Western settings have been both practical and theoretical.

The practical lessons involve the development of various guidelines and cautions, first concerning the process of test adaptation and second, relating to the process of test administration itself. The theoretical lessons have meanwhile been concerned with the issue of test result interpretation in cross-cultural contexts which in turn have led to some quite radical reconsiderations of important aspects of psychological theory and method.

The appraisals and adaptations of test content and the elaboration of administration guidelines for test use with ethnically and culturally different populations have been the work of a number of researchers. The contributions of Biesheuvel (1949, 1952, 1969), Brimble (1963), Deutsch, Fishman, Kogan, North and Whiteman (1964), Doppelt and Bennett (1967), Hicks (1969), Irvine (1965), McElwain and Kearney (1970), Ord (1970), Ortner (1963 a,b), Schwarz (1963), Schwarz and Krug (1972), Silvey (1963), Vernon (1969) and Wober (1967) have all been important in this regard. All of the above papers and texts have dwelt upon the issues of test adaptation and administration in cross-cultural situations. Brislin, Lonner and Thorndike (1973) have recently summarized much of this material.

With regards test adaptation (and development) Brislin, Lonner and Thorndike have come to the following conclusions:

(a) Considerable attention must be given to the selection of stimuli (tests and test items). If non-verbal stimuli are employed there should be an attempt to justify the assumption that the same intentional

meaning is attached to them by subjects as by the test user and test developer. Regard should always be given to the appropriateness or otherwise of the selected stimuli in particular cultural contexts.

(b) Where there are few or no precedents on the use of particular stimuli evidence in support of their inclusion must be obtained.

Brislin, Lonner and Thorndike (1973) suggest that anthropological and ethnographic evidence should be studied and also point out that members of particular cultures "will be valuable judges and sources of information concerning the selection of tests, or the modification of any parts of the tests" (p.140).

(c) When the stimuli are verbal the issues of translation and 'understandability' arise. Brislin, Lonner and Thorndike have detailed procedures for assessing both the accuracy of translated items and their cultural relevance. They suggest that psychologists working in the cross-cultural domain must assume that translated test stimuli will contain flaws and that only through application and analysis will the difficulties be understood.

(d) Linked to any consideration of the adaptation of test items or stimuli is the issue of the overall content validity of the instrument in cross-cultural contexts. Here Brislin, Lonner and Thorndike point to the need for evidence to be brought forward in support of the behavioural representativeness of the content of the test in relation to the cultures in which it is to be employed. It is their view that -

"Unless the tests (items, stimuli) being used make a reasonable claim of culture-fairness (then) determine their appropriateness before they are administered. After administration, an examination of the responses should be made to verify their appropriateness" (p.141).

On issues relating to test administration Brislin, Lonner and Thorndike list some nineteen points and procedures ranging across such matters as the use of visual aids, methods for instructing subjects in

the task requirements through to appropriate behaviour on the part of the test administrator in relation to interpersonal contacts with the subjects. They have summarized the formidable list by suggesting that -

(a) The test administrator must adopt a "minimax" strategy, that is, minimise "..... the potential deficiencies in administration while maximising the gain (desired response elicitation)" (Brislin, Lonner and Thorndike, 1973, p.140).

(b) To achieve this they point out that nothing can be taken for granted and that therefore the points and procedures concerning test administration must be both considered and acted upon.

(c) Furthermore, cross-cultural studies often place sets of subjects in somewhat unique situations with comparative strangers. There will be inter- and intra-individual consequences of the encounters that may have a bearing upon the data.

(d) Finally, it is suggested that -

"If because of administration difficulties there is a reasonable doubt that administration, and hence measurement, fell short of minimum expectations, do not use the results" (Brislin, Lonner and Thorndike, 1973, p.141).

These issues of test adaptation and administration can, of course, be looked at in other ways. Accepting that cross-cultural and sub-cultural (e.g. social class) comparative studies will continue to be undertaken (a point raised below) then one of the major problems remains the development of tests or measures that will enable the drawing of safe inferences concerning the compared groups. Speaking in relation to intellectual abilities, Vernon (1969) summed up the problem as one where subjects "..... have to display their abilities through a common medium of comparison if we are to measure them, and if they are not equally at home in this medium, we cannot compare them" (p.96). The attempt to find a 'common medium of comparison' in relation to the measurement of

intellectual ability draws attention to factors likely to influence test performance. Biesheuvel (1952) spoke about two classes of factors. First, there are those factors which are 'intrinsic' and unlikely to be easily modified in the short term, but which nonetheless affect measured intellectual ability. Many of the cumulative biological and environmental effects upon an individual fall into this class. For example, a history of poor nutritional balance or being deprived of (denied) adequate and appropriate educational experiences.

Second, there are the 'extrinsic' factors which pertain to the measuring device and the measurement operations. It is presumed that these extrinsic factors are modifiable and can thus be manipulated in order to reduce distortions in the measurement of, in this case, intellectual ability. The procedures and recommendations on test adaptation and administration in Brislin, Lonner and Thorndike (1973) are a summary of the accumulated evidence and experience of a good many researchers who have focused upon the modification of various extrinsic factors in the search for a more 'common medium of comparison'.

As will be seen, many of the procedures and recommendations made in this regard apply to the major measure of intellectual ability employed in this research (see Chapter 5).

Alongside this concern with the practical aspects of test adaptation and administration there have been a number of important developments taking place in the theoretical perspectives adopted in relation to cross-cultural comparative research.

The emic-etic distinction must be part and parcel of any consideration of cross-cultural psychological research. Briefly, emic analysis refers to studies where the intention is "..... to document valid principles that describe behavior in any one culture under study, taking into account what people themselves value as meaningful and

important" (Brislin, 1976, p.16). On the other hand etic analysis refers to studies attempting "..... to make generalizations across cultures that take into account all human behavior" (Brislin, 1976, p.16). The prime purpose of such etic analysis is theory building. However -

"The problem for psychologists is that by administering a test standardized in one country (usually their own), they may be imposing an artificial etic and losing the emic or meaningful aspects of the other culture as practiced by their members" (Brislin, Lonner and Thorndike, 1973, p.24, Brislin, Lonner and Thorndike's italics).

The question of emic and etic approaches in cross-cultural psychological research has received a good deal of attention of late (see Berry, 1969; Brislin, Lonner and Thorndike, 1973; Price-Williams, 1975; Triandis, 1976).

Brislin, Lonner and Thorndike in their consideration of this issue have given support to Berry's (1969) proposals on how cross-cultural research might proceed. The proposals attempt to strike a balance between the emic and etic perspectives without sacrificing the long term theory building goal of psychological research. An important point made by Berry concerns the ideas of the 'imposed etic' and 'derived etic' approaches. Berry is suggesting that many cross-cultural studies are of an 'imposed etic' nature which attempt to study some (presumed) universal aspects of behaviour using descriptive categories and methods for obtaining data which have arisen from only one cultural framework. On the other hand, he sees the 'derived etic' approach as one where an attempt is made to establish shared descriptive categories having a greater commonality between cultures in the methods of obtaining data. The goal of theory building has not been abandoned but is approached differently. As Brislin, Lonner and Thorndike (1973) argue -

"New instruments and research techniques can be devised and validated using the derived etic as a conceptual base. These instruments must, of course, be equivalent in meaning to the members of the cultures under study" (p.25).

A substantial proportion of Brislin, Lonner and Thorndike's text on cross-cultural research methods focuses on the development of derived etic procedures. It considers not only what might be done with regards the general issues of test adaptation and administration but also reviews specific test instruments considered suitable for cross-cultural use and which, in varying degrees, reflect derived etic considerations. Included are measures of intellectual ability studied in both the Papua New Guinean and Australian contexts which now form the Queensland Test, the instrument at the centre of the research reported in this thesis.

While the emic-etic distinction raises important methodological questions concern for the purpose and interpretation of cross-cultural research has been strongly influenced by the cultural relativist viewpoint.

The doctrine of cultural relativism has been popular in anthropology for over half a century (see Herskovitz, 1973) and has recently come to the fore in cross-cultural psychology. Berry (1969, 1972) and Price-Williams (1975) have both addressed this issue. The key argument is simple enough. Cultural relativism embodies the argument that there is no way of rationally justifying standards that attempt to span the boundaries of time, society and culture. Consequently societies cannot be ranked morally, culturally or cognitively. Berry (1969) initially proposed the derived etic approach to cross-cultural research in the hope that concepts and instruments applicable to more than one culture system might be developed. He has since identified this as a 'gradual' approach and suggests a more 'radical' solution. Thus he has argued that -

"..... rather than attempt to modify or merge our assumptions, we can conceptually wipe the slate clean and approach the problem from a nihilistic stance. This position accepts the statement: 'We can assume no psychological universals across cultural systems'. It may be termed a position of radical cultural relativism" (Berry, 1972, p.78, Berry's italics).

Obviously, Berry's argument on the absence of psychological universals runs counter to the evidence discussed above on cognitive universals from the fields of anthropology, linguistics and psychology. However, the more immediate issue is the doctrine of cultural relativism itself which, as Popper (1966), Gellner (1973) and Jarvie (1975) have noted, is leading the social sciences into some philosophical deep water.

Jarvie (1975) has provided a useful summary of the issues. He sees relativism as being peculiar to Western social scientific thought and containing within it a confusion between questions of knowledge and questions of belief. This has led, at times, to a failure to recognise that "There is a world, which is the way it is and not another way" (Jarvie, 1975, p.347, Jarvie's italics).

Nonetheless, cultural relativism is an attractive argument since it gives the appearance of liberalism and tolerance. Afterall, as Jarvie points out "The open-minded tolerance and sympathy with which a relativist approaches other cultures seems humane and rational" (p.349). However, the issue is a little more complex and the arguments reviewed by Jarvie led him to the conclusion that -

"The relativist is tolerant if he discovers intolerance among the natives; he is intolerant of intolerance among his compatriots. This is not only the double standard; it is also patronizing" (Jarvie, 1975, p.349).

Gellner (1973) has also challenged the cultural relativist position. He has argued that the weight of evidence would suggest that the peoples of the third world and from non-European cultural traditions are not themselves relativists. This has not been the case

either within traditional cultures or in modernizing cultures. It is Gellner's view that traditional cultures are as ethnocentric as the best and that in the modernizing cultures a cluster of Western values and artifacts associated with industrialization are rapidly being adopted. This points to the condescension underlying much relativist talk and thinking. Thus -

"We are told that natives do not think as we do, do not suffer as we do, are not interested in formal freedom, intellectual inquiry, etc. Indeed, the very refusal to extend to native peoples the intellectual courtesy and consideration we extend to each other is condescension" (Jarvie, 1975, p.350).

Berry's (1972) radical cultural relativist argument, if accepted, suggests that no psychological universals can be assumed and that an emic view is the only legitimate view. This would mean an end to cross-cultural comparative research as presently pursued be it on intelligence or any other psychological construct. This would be the case if the relativist's arguments were sound. As has been shown, however, this is not so. Cultural relativism in confusing social tolerance with epistemological tolerance runs the risk of doing cross-cultural psychology something of a dis-service for as Jarvie (1975) notes -

"It is fine to be socially tolerant of other views and their holders; (but) wretched to declare them equally true in their own way when they are nothing of the sort" (p.350).

Perhaps relativism has run its course, although in doing so has alerted psychology to some important theoretical and methodological questions. General theory building remains a goal of cross-cultural psychology and comparative research a necessary procedure. As Triandis (1976) has noted "Psychologists make many comparisons. In fact, it has been argued by some that all knowledge is knowledge of differences, yielded by comparisons" (p.155). The problem is that some comparisons present special methodological dilemmas. Attention has already been given in general terms to the efforts made in relation to the special methodological dilemmas associated with the cross-

cultural measurement of intelligence. The detailed notes on the test employed in this research also address many of the same dilemmas. Even so, it would be unwise to suggest that the difficulties have all been resolved.

A very useful contribution to the theoretical issues surrounding intelligence test data from cross-cultural studies has been made by Irvine (1969a, b; 1970).

From studies in central Africa employing a large battery of ability measures Irvine (1969a, b) has found considerable similarity in the constructs derived through the use of factor analysis on the data from different culture groups. Cautiously Irvine suggests that the factorial method of analysing and classifying ability tests can be used in non-Western settings, particularly in relation to tests assessing intellectual skills and abilities fostered through schooling. In interpreting the factors obtained in the various African culture groups, Irvine employed many of the constructs developed through the use of similar ability measures with Western populations, e.g. 'general ability', 'numerical-educational', 'verbal-educational' and 'mechanical-spatial'. The shared nature of the constructs were in Irvine's view primarily the result of the common pattern of education received by the culturally diverse groups of subjects. He concluded that in large measure variations in African languages, cultural traditions and social relationships were not obviously related to the acquisition of educational skills in a second language and that -

"Cognitive tests such as those used in the surveys reported have not begun to tap modes of thought that are the product of African languages and social relationships" (Irvine, 1969a, p.27).

Interest in this latter question led Irvine (1970) into a study of Mashona proverbs, omens and beliefs. Taking these as the ground rules for intelligent and purposive acts within that society

he classified the kinds of knowledge needed for these to have meaning, and the consequences of non-observance. After a knowledge of the natural environment (including objects and animals) were the 'social' kinds of knowledge which gave proverbs, omens and beliefs meaning. The consequences of non-observance rebounded on self, kin and community and was evidence to the Mashona of unintelligent behaviour. Serpell (1974) and Wober (1974) have also employed community definitions and estimates of intelligence in African studies and noted the strong affective emphasis associated with such definitions and estimates in traditional cultures.

Having suggested that some important dimensions of intelligent behaviour may not be tapped by conventional measures of intelligence in non-Western settings Irvine (1970) has proposed that in the first instance the construct Intelligence X must be applied to the interpretation of ability test data. Earlier the concepts of Intelligence A, B and C as proposed by Hebb (1949) and Vernon (1955) were discussed. Intelligence A is viewed as genotypic and inferred, Intelligence B as phenotypic and observed while Intelligence C applied strictly to intelligence test scores. However, Irvine's view is that discussion of intelligence data even in Intelligence C terms may be inappropriate in non-Western settings. Hence his construct Intelligence X, proposed to describe test scores from the use of these tests in such settings. He has reasoned that these tests may not produce results equivalent to Intelligence C because the test scores may reflect variance stemming from indigenous thought systems. It is Irvine's contention that -

"..... the values of non-Western societies with a considerable oral tradition for the transmission of skills will prescribe the criteria for intelligent acts. Such acts may require a completely different use of knowledge from that of the individually competitive, industrial west" (Irvine, 1970, p.24).

It is the nature of the relationship between these different ways of behaving intelligently that in Irvine's view remains unknown. Or, as Price-Williams (1975) puts it - "..... the task of making Intelligence X congruent with Intelligence C (assuming this construct to apply to Western populations alone) is a problem not yet overcome" (p.52). The diversity of human culture systems might suggest that this congruence is unlikely to be achieved in any general way in the short term.

However, Vernon (1969) is inclined to think otherwise. He has argued that the demands of social change and technological and industrial development in non-Western societies will give rise to the development of Western-type intellectual abilities. Consequently, measures based on Western constructs of ability are of practical assistance in the selection of personnel for education, training and employment. Although this moulding effect may initially apply only to a section of the society the implication of Vernon's view is that expanded education will lead to increased construct congruence across cultures.

Price-Williams (1975) sees the changes as being not simply a case of the advance of technology alone, but of a more pervasive cultural diffusion process that is having a world-wide effect. He acknowledges that this thought may be depressing but adds that it "..... probably sums up the picture for an educated elite in a good many areas of the world and does have consequences for intelligence testing" (Price-Williams, 1975, p.56).

Irvine (1966) has himself argued that as school systems become more homogeneous with the development and industrialization of non-Western societies the relationships between test scores will come to resemble those found in Western societies. Subsequent studies by

Irvine (1969a, b) have in fact shown that when ability measures focusing upon educationally-induced (and possibly 'over-learned') skills are used with different culture groups, then there is considerable similarity in the underlying constructs. Furthermore, the descriptive statements of the proposed construct follow those employed with similar data from Western contexts. It would appear that Irvine is prepared to treat such data in the Intelligence C mode and interpret it accordingly. This does not necessarily undermine his original Intelligence X construct. Rather, it suggests that a researcher or test user must first consider whether or not the ecological, environmental, educational, social and cultural circumstances between particular groups of subjects are sufficiently similar in specific ways as to predispose them towards similarities in intellectual processes. Second, evidence on the psychological equivalence of ability measures across culture groups must be sought - not assumed. Granted that some advances have been made in adapting test content and developing the methods of administration for cross-cultural use it still remains to be shown that "..... the test scores for the groups have the same psychological meaning" (Irvine, 1966, p.28).

The question of the psychological (and statistical) equivalence of measures of intellectual ability for differing culture groups has been extensively discussed in relation to test fairness and test bias. Darlington (1971), Humphreys (1973), Irvine (1973), Linn and Werts (1971), McNemar (1975), Schmidt and Hunter (1974) and Thorndike (1971) have all considered aspects of this question. Cleary, Humphreys, Kendrick and Wesman (1975) have presented a useful summary paper on test fairness reviewing historical issues, questions concerning theories of intelligence, test misuse and misinterpretation, and strategies for

the evaluation of test fairness. New approaches to assessment are also considered. This report has received some critical comment (see Jackson, 1975; Bernal, 1975) but the major arguments are defended by Humphreys (1975).

The issue of test bias has been much discussed in the United States but has also become an issue in other countries as cultural, ethnic and social diversity has become recognised. Simply put, the bias argument is that ability tests by their form and content measure things that exist to a greater degree in the privileged (often European) population than in particular ethnic or cultural minority populations (often non-European). The same argument is invoked when tests are applied to sub-populations differing in wealth, socio-economic status and educational opportunity. The discrimination in the tests might be obvious or subtle. Concern over this possible discrimination has led to calls for the complete banning of tests.

However, as has been pointed out in many of the above papers, it is incorrect to argue that a test is biased simply because differences are found between cultural or ethnic groups. In fact, it can be argued that a test can not itself be biased. Where bias intrudes is the critical areas of test score interpretation and test score use. It is the interpretations and uses of test scores that can be biased. The critical issue is then one of the validity of a test for use with particular sub groups of a population. Cleary, Humphreys, Kendrick and Wesman (1975) have discussed content, construct and criterion-related validity in relation to test bias. Criterion-related validity is given special attention. It is with the practical use of tests in selection and placement that important questions on bias or fairness have come to the fore. In this regard it has been clearly shown that the relationships between criterion scores and test scores

for sub groups must be researched. Test score use can be substantially biased for particular subgroups if a single selection norm is employed with all subjects when the relationship between the test scores and the criterion measure is markedly different between the subgroups. Thus, it can be said that -

"A test is considered fair for a particular use if the inference drawn from the test score is made with the smallest feasible random error and if there is no constant error in the inference as a function of membership in a particular group. These components of fairness are directly translatable into regression statistics" (Cleary, Humphreys, Kendrick and Wesman, 1975, p.25).

These authors then illustrate biased and unbiased types of relationships between test and criterion scores employing regression techniques and argue for a definition of test fairness where neither consistent over- or under-prediction of criterion scores from test scores occurs for the particular subgroups in question. It will be noted that a definition of fairness in test usage of this nature is independent of group differences in mean scores. Cleary, Humphreys, Kendrick and Wesman (1975) suggest that ideally -

"..... whenever a precise assessment of the fairness of a test for a particular use arises for members of distinct subgroups of the population, it must be determined whether the within-group regression lines are identical. A first requirement is that the lines be parallel, but a second critical requirement is that they have the same intercepts. If the regression lines are identical, the test can be used in the combined group with no advantage or disadvantage occurring to either group" (p.27).

Reasonable as this may sound, it is not possible to attain perfect fairness. The presence of small nonzero differences in observations, the difficulties in assignment of subjects to groups and statistical effects associated with changes in test reliability or in employing composite predictor tests (see Linn and Werts, 1971) result in departures from perfect fairness. If there is a difference between the regression lines for subgroups, even one attaining statistical

significance, Cleary, Humphreys, Kendrick and Wesman (1975) argue that the social significance of the difference is a far more important consideration. They then discuss ways in which the importance of the size of any difference might, in practical terms, be assessed.

The attention given the issue of test bias of late heralds a number of important consequences for test use in both research and practice in cross-cultural settings. First, it is evident that generalizations from test scores to non-test behaviours for different groups cannot be supported in the absence of explicit validation studies. The transfer of test validity across groups cannot simply be assumed. Second, the generality in the interpretation of test performance must be much more circumscribed in the absence of clear evidence on construct congruence. In addition to these direct consequences the question of test bias and test comparability between culture groups has stimulated the field of psychometrics.

Drenth and van der Flier (1976) have discussed the Rasch model for estimating subject ability and item difficulty and also research approaches that take as their starting point the non-comparability of test item responses when attempting to separate individuals and form homogeneous subgroups. The development of these procedures in Drenth and van der Flier's view will be an important psychometric contribution to the study of test comparability. Furthermore, they argue that test scores must meet a number of requirements if they are to be considered suitable for cross-cultural comparative studies. Drawing on Poortinga's (1971) studies on test comparability Drenth and van der Flier (1976) state three conditions -

- "1. Functional equivalence (the test has to measure the same attribute in different cultures).
2. Score equivalence (the test has to measure this attribute on a similar scale).
3. Item equivalence (the same requirement as score equivalence, but applied on an item level)" (p.142).

The importance of these conditions is that they -

"..... originate from the necessary theoretical assumption that the test variance in both groups can be explained by the same theoretical construct(s) and that these constructs have the same position in a nomological network and an interrelated set of constructs" (Drenth and van der Flier, 1976, p.142).

Once again these requirements represent an ideal and it remains to be seen how well measures of ability employed in cross-cultural contexts fare against such standards.

The attention given to the psychometric properties of the Queensland Test with European and Maori samples in this thesis bears directly upon the issue of psychological equivalence.

Race and Intelligence

Given the nature of the research reported in this thesis, it is necessary to include a brief statement on the race-intelligence debate. This issue has received considerable attention since the publication of Jensen's (1969a) paper suggesting that the lower performance of American Blacks relative to American Whites on measures of intelligence and scholastic attainment might be ascribed more to genetic rather than environmental factors. Jensen's (1972a, 1973 a, b) texts contain elaborations of his 'racially linked genetic differences in levels of intelligence hypotheses'. Volumes by Brace, Gamble and Bond (1971), Cancro (1971), Kamin (1974), Loehlin, Lindzey and Spuhler (1975), Richardson and Spears (1972) and the collected papers in the Harvard Educational Review, Reprint Series No. 2 and 4 (1969a, b) all consider Jensen's arguments from a number of perspectives and review the evidence for and against the genetic determination of intelligence and its bearing upon ethnic group differences in test performance. These publications consider the methodological difficulties of heritability studies, criticise much of the psychometric data and theory underpinning the race differences in intelligence hypothesis and

also draw attention to the potentially disastrous social consequences of a thesis linked as much to doctrines of social, economic and political domination by particular ethnic groups over others as to any indisputable scientific evidence.

While Jensen's thesis has drawn support from some quarters (see Eysenck, 1971; Shockley, 1972) the predominant reaction to his arguments and evidence has been negative in that his case is considered to be 'not proven'. Loehlin, Lindzey and Spuhler (1975) reviewing the evidence from many of the texts and the all too numerous papers on this question arrived at the following conclusions.

First, that the observed average differences in ability test scores between (U.S.) ethnic groups arise in part from test inadequacies and biases, in part from environmental differences between groups and in part from group genetic differences. These factors are not necessarily independent, in all probability they interact.

Second, the evidence (or lack of it) does allow for different relative weightings to be given these three factors. They suggest that the weightings might be sensibly varied in relation to the abilities, groups and tests under study.

Third, irrespective of the weight given these factors it was found that ability differences among individuals within an ethnic (or socio-economic) group far exceeded in magnitude any average difference between such groups.

Loehlin, Lindzey and Spuhler emphasize the relative nature of their conclusions by noting that they relate only to (U.S.) conditions in the recent past. Changes in these conditions may well alter the importance of and relationships between the factors. Furthermore, they point out that observed between-group ability differences need not always be in the same direction as probable environmental differences, genetic differences and psychometric biases. The limited nature of the

conclusions on such a controversial topic are acknowledged. As the authors say, "It does not appear to us that the state of the scientific evidence at the present time justifies stronger ones" (Loehlin, Lindzey and Spuhler, 1975, p.239).

This is essentially the position reached by the author (St. George and St. George, 1975) in an earlier statement on this issue where it was concluded that what little is known about genetic factors in intelligence does not allow us to immediately leap to the conclusion that Race A is genetically inferior to Race B just because between-group differences were found upon a particular ability measure.

The research reported in this thesis does not address itself to the question of the likelihood of genetic factors being implicated in any between ethnic group differences in test performances of European and Maori subjects. As will be seen, the question has been asked by some commentators. However, the issue is not pursued in this research because basically it is viewed as being relatively unimportant.

It is unimportant because ethnic groupings in the research context stress social affiliation as well as biological connectedness. It is unimportant because of the increasing genetic diversity of the research populations. It is unimportant because any group estimates of the heritability of intelligence have no implications for an individual and no utility in a situation of changing genetic and environmental circumstances (points made clear by Hirsch (1968) some years ago). Furthermore, the basic correlational data on measured ability, measured environments and kinship just does not exist for the research populations in question. Comment on the race-intelligence issue without such data could only be speculation - and dangerous speculation at that.

Summary

The first part of this chapter reviewed the concept, definition and measurement of intelligence in psychology. The philosophical discussions on the nature of the human mind were noted along with the early attempts to measure intellectual capacities through simple sensory processes. The break with this rather unsuccessful approach came with Binet's measures of higher mental processes and his concept of a pervasive general intelligence. The new field of psychometrics grew in power and sophistication along with the growth of new tests of intelligence.

However, defining intelligence has continued to present difficulties. Stipulative definitions suggested such qualities as abstract thinking, general reasoning capacity and adaptation. However, over the years stipulative definitions have tended to give way to the more circumscribed operational definitions linked to particular measurement operations. The study of intelligence test data has long been linked to factor analytic procedures and has led to the postulation of a number of theories of intelligence. These theories conflict principally over the emphasis that should be given to a central and general construct as opposed to more limited and special types of abilities. Vernon (1960, 1961) is one author who has suggested a useful rapprochement between these types of theories. Theories less dependent on factor analytic procedures were noted along with viewpoints that are largely independent of the intelligence test movement and instead approach the study of intelligence in terms of qualities evident in its development and organisation.

Nonetheless, the intelligence test and the theories associated with it are seen to be both important within psychology and to the discipline's practical contribution to human affairs. Furthermore,

it was suggested above that the hierarchical general intelligence theory in the tradition of Vernon provides, for the moment, one of the more useful models for the formulation of constructs from intelligence test data. When coupled with the Intelligence A, B and C meanings of intelligence as set down by Hebb (1949) and Vernon (1955) there is now the possibility of arriving at a clearer understanding of the meaning of intelligence test data.

The second part of this chapter has reviewed many of the theoretical and practical issues in the cross-cultural study of intelligence. The meaning of cultural group differences on all manner of tasks indicative of ability has posed considerable theoretical problems. Theories stressing underlying intellectual differences and those stressing underlying intellectual universals were considered. The anthropological, linguistic and psychological evidence tends to favour the latter viewpoint and the evidence, such as it is, on 'racial'-genetic differences in intelligence is inconclusive.

Attention was given to the adaptation and administration of intelligence tests in cross-cultural settings. The emic-etic distinction, the cultural relativist argument, indigenous concepts of intelligent behaviour, Irvine's (1970) Intelligence X concept and recent views on test fairness and bias were all discussed. It was noted that there was much that could be done in the test adaptation and administration fields to improve tests considered for cross-cultural application. Clear guidelines are available and instruments of a 'derived-etic' nature are, in theory, possible. The relativist position was considered and on some counts found wanting as an argument against the total abandonment of the cross-cultural study of intelligence. The suggestions arising from recent reconsiderations on test bias for establishing the functional equivalence of measures

it was suggested above that the hierarchical general intelligence theory in the tradition of Vernon provides, for the moment, one of the more useful models for the formulation of constructs from intelligence test data. When coupled with the Intelligence A, B and C meanings of intelligence as set down by Hebb (1949) and Vernon (1955) there is now the possibility of arriving at a clearer understanding of the meaning of intelligence test data.

The second part of this chapter has reviewed many of the theoretical and practical issues in the cross-cultural study of intelligence. The meaning of cultural group differences on all manner of tasks indicative of ability has posed considerable theoretical problems. Theories stressing underlying intellectual differences and those stressing underlying intellectual universals were considered. The anthropological, linguistic and psychological evidence tends to favour the latter viewpoint and the evidence, such as it is, on 'racial'-genetic differences in intelligence is inconclusive.

Attention was given to the adaptation and administration of intelligence tests in cross-cultural settings. The emic-etic distinction, the cultural relativist argument, indigenous concepts of intelligent behaviour, Irvine's (1970) Intelligence X concept and recent views on test fairness and bias were all discussed. It was noted that there was much that could be done in the test adaptation and administration fields to improve tests considered for cross-cultural application. Clear guidelines are available and instruments of a 'derived-etic' nature are, in theory, possible. The relativist position was considered and on some counts found wanting as an argument against the total abandonment of the cross-cultural study of intelligence. The suggestions arising from recent reconsiderations on test bias for establishing the functional equivalence of measures

across culture groups as a precondition to their use are seen as being much more positive and practical. Recognition was given to the observation that cultures may differ in their conceptions of intelligent behaviour and that the concept of Intelligence X may in some contexts prove to be useful. It will all depend on the social situation existing between the particular culture groups in question and the evidence on the functional equivalence and construct congruence for a particular measure of intelligence when researched cross-culturally.

CHAPTER 2

MAORI EDUCATIONAL UNDERACHIEVEMENT

Introduction

The questions of access and involvement of the Maori people of New Zealand in almost every aspect of New Zealand life have been the subject of wide study and debate.

The continuing publication of material has fostered the appearance of both general and specialized bibliographies to assist students with the extensive literature (e.g. McKenzie, 1970; Taylor, 1965; Taylor, 1972). Equally, the place of substantial texts seems well assured (e.g. Barrington and Beaglehole, 1974; King, 1975; Metge, 1976; Ritchie, 1963; Schwimmer, 1968; and Sutherland, 1940 to name but a few).

Of the many fields of inquiry, education would rank high as an area where the question of Maori access and involvement in an aspect of contemporary New Zealand society has received considerable attention. It has generated its own specialised research concerns, its own series of official and semi-official reports, its own general and specialised texts and its own cadre of professionals seeking to both understand and ameliorate Maori educational difficulties. Commenting in broad terms on the question of Maori education, St. George and St. George (1975) were moved to say that -

"Schooling in New Zealand appears to have acquired, or perhaps created for itself, at least one seemingly intractable problem - namely 'Maori schooling'. It is unlikely that any other topic in the educational scene within New Zealand has attracted as much comment, as many reports, as much research, as many publications or, for that matter, as many prophets" (p.302).

It is not the intention of the author to attempt to review thoroughly all the historical and contemporary literature on the Maori education question. Useful synopses can, however, be found in Bray and Hill (1973, 1974), Ewing and Shallcrass (1970) and Metge (1976).

The questions of more immediate concern are those relating to the apparent educational underachievement of the Maori population relative to the European population, the nature of the substantiated or suggested reasons

for this underachievement and in particular the place of intellectual ability factors amongst these reasons.

The following discussion has been drawn from material where the term underachievement has referred to the relatively poorer performances of Maori students on many indices of educational attainment where compared with their European peers. It is recognised that this use of the term does not altogether satisfy the concept of underachievement as discussed by Thorndike (1963). Thorndike rightly argues for the clear specification of the expected or predicted achievement, in conjunction with an appreciation of the errors of prediction, prior to statements about underachievement.

The Evidence on Underachievement

In terms of a good many educational attainment criteria currently utilized throughout the New Zealand educational system, Maori children do not do as well as their European counterparts. Repeated reviews and investigations testify to this state of affairs. Ramsay (1975) for instance, has produced figures on school attainments by European and Maori pupils for the years 1955, 1966 and 1972 from Department of Education sources. He notes that -

"While the proportion of Maoris leaving school without an academic qualification has declined significantly in recent years, the gap between Maori and non-Maori has remained relatively constant. For example, in 1955 40 per cent more Maoris left school without any qualification than did non-Maoris. Although this figure dropped to 37.8 per cent in 1964, it has since risen to 39 per cent where it appears to have stabilised" (p.289-290).

Furthermore, as Ramsay points out, the academic qualifications obtained by Maori school leavers are at the lower end of the range. This point should not be overlooked in the rush to suggest that 'substantial gains' have been made when the wider society is, realistically or unrealistically, making demands for higher qualifications for entry into many areas of the work force and tertiary education.

For Ramsay the conclusion to be drawn is obvious -

"..... the Maori, generally speaking, leaves school with lower qualifications than the non-Maori. As there is a close connection between the level of academic qualifications and the kind of job held, it should follow that the Maori will be in a disadvantaged position in the labour force" (p.290).

This Ramsay investigates and demonstrates to be the case.

The possible reasons for the educational disadvantage of the Maori pupils have long been the subject of investigation and speculation.

Metge (1976) has provided a useful, chronologically organised account of the explanations advanced for the educational underachievement of Maori children. Her analysis suggests an awakening of concern about educational underachievement in the 1950's with "..... a general tendency to locate the causes in the personality and home experience of the Maori children" (p.154). Interwoven were the effects of rural residence, the question of the low socio-economic circumstances of many Maori families and from time to time the suggestion of low interest by Maori parents in their children's educational progress. Metge gives some consideration to each of these arguments.

The 1960's saw the explanations for Maori educational underachievement being couched in terms of a cultural deprivation conceptualization stressing the intellectual and/or linguistic impoverishment of the home environment. Extrapolations from research in the United States and the United Kingdom were not uncommon (see Mitchell, 1968, 1970) culminating in the publication Maori Children and the Teacher (Department of Education, 1971).

This handbook for teachers adopted a strong cultural deprivation perspective on a number of critical issues in Maori education. This was especially so over the question of language development. In a very simplistic manner, Bernstein's (1961) 'restricted' and 'elaborate' code terms were employed and imbued with powerful explanatory significance in relation to the educational attainments of Maori children. The arguments and views presented in this book were however, quickly challenged by Oppenheim (1972) on the grounds that they were but sweeping generalizations

and that ethnocentric assumptions and 'folklore' concerning the background, abilities and experiences of Maori children appeared to have gained more prominence than the facts.

Hawkins (1972) also commenting on this same publication, presented a detailed analysis of the 'restricted' and 'elaborate' code question in relation to the linguistic skills of Maori children. His conclusions were to the effect that the whole issue of the English language skills of Maori children was a far more complex matter involving questions concerning both the development of English dialects and the extent of Maori and English bilingualism in New Zealand. Certainly the instructional consequences arising from a recognition of both dialects and degrees of bilingualism amongst the Maori population would be somewhat different from those arising from the assumption that the language difficulties of Maori pupils stemmed only from a 'restricted' English language code. Pride (1974), George (1974), Anne Salmond (1974) and Elizabeth Gordon (1974) in a collection of short papers have also commented on the use of restricted linguistic code explanations of Maori educational underachievement and find these to be wanting. Research is now however, focusing on the question of bilingualism and dialects amongst Maori children (see Benton, 1972). In time this may counter certain linguistic deficit explanations. Meanwhile the process of education itself is being subjected to criticism on the grounds that it is in many ways premised upon monocultural assumptions and practices (see Dewes, 1968; Walker, 1973). These arguments are being expressed with considerable conviction by Maori commentators.

These changes in the nature of the explanations advanced for Maori educational underachievement no doubt reflect developments both within New Zealand and overseas. Increasingly, biculturalism, multi-culturalism or cultural pluralism are being spoken of as both desirable and possible forms of social evolution for New Zealand in view of its increasingly varied ethnic and cultural composition (see Piddington, 1968; Havighurst, 1974; Pitt and Macpherson, 1974).

An alternative approach to considering the explanations advanced for Maori educational underachievement would be by studying the frequency with which various types of explanations or causal factors have been suggested in the literature. A most useful study along these lines has been undertaken by Harker (1971a).

Harker has presented a systematic summary of the types of explanations offered for Maori educational underachievement from twenty-three publications appearing between the years 1957 and 1969. Some twenty determinants of Maori educational underachievement were identified by Harker in his search of these publications and in a table he has indicated whether or not these determinants were accepted as causal factors. These determinants were grouped into three sub-classes - genetic, social and cultural. The genetic sub-class included the single determinant of 'lower intellectual functioning' which was rejected as a causal factor in six publications and failed to gain mention in the remainder. The social sub-class included twelve determinants such as school system inadequacies, low level of parental education, low income, lower class status and racial prejudice. Such social factors have all been advanced from time to time as reasons for Maori educational underachievement, the most popular reasons being school system inadequacies and the low level of parental education.

The sub-class of cultural determinants included seven factors covering language difficulties, differences in cultural background and socialization, the absence of a professional tradition, the rural orientation of the Maori world, culturally induced differences in intellectual functioning and lower intellectual functioning as a result of cultural factors. The most frequently advanced causal factors in the cultural determinants area were language difficulties and the different cultural background of Maori children.

In reviewing these studies, Harker was led to the conclusion that there were only a relatively small number that could be considered to "..... provide a 'hardcore' of empirical data" (p.46) on the question of

Maori educational underachievement. Harker's view was that -

"..... the sum of (the) combined efforts presents a rather slender body of empirical research upon which to base either explanations of the level of Maori achievement or policy decisions concerning the education of Maori children" (p.46).

Further, Harker pointed out that the authors of many of the other publications in this field have relied on this 'slender body of empirical research' or on a priori assumptions in suggesting causes for Maori educational underachievement.

A case in point which Harker refers to concerns the question of intelligence. He noted that many assertions have been made concerning possible differences in the distribution of intelligence between Maori and non-Maori without regard to the definition of the concept or its method of measurement. As well, Harker contended that perhaps unwarranted assertions have been made concerning the vexed question of the genetic components of intelligence without "..... any regard for the extremely uncertain situation within this field of research" (p.46).

To these criticisms Harker has also added the observation that -

"..... a number of writers suggest only one or two factors in explaining Maori-Pakeha differences in school attainment. Even where a number of factors are suggested, the complex of interactions between the various factors is not systematically explored" (p.48).

As a response to this limited perspective, Harker has postulated an interaction model of the major suggested factors contributing to the educational underachievement of Maori children. One of the four major factors is identified by Harker as being a 'different pattern of intellectual functioning'. In this factor he included the culturally induced lower level of intellectual functioning suggested by Ausubel (1961) and Lovegrove (1966). He also included cognitive style considerations and then addressed himself particularly to the 'relational' versus 'analytical' cognitive style model of Cohen (1969) which he was subsequently to employ in research with European and Maori children. The other interacting factors proposed by Harker were 'lower class status', 'ethnocentric school system' and 'different cultural

background'.

So far as explanations of Maori educational achievement are concerned, it would appear that two features are worth noting. First, over time there has been a movement away from explanations premised on the superiority of the life style and culture of the dominant New Zealand society. Increasingly it has become apparent that explanations offered in cultural deficit terms implicitly, if not explicitly, adopted a superiority stance over the experiences, life style and culture of the so-called culturally deprived group. Reference to cultural difference has increased of late. An example would be Gadd's (1976) text for teachers Cultural Difference in the Classroom: Special Needs of Maoris in Pakeha Schools.¹ Nevertheless, it is still unclear as to whether or not New Zealand is prepared to fully endorse and accept the consequences in the educational sphere and elsewhere of fostering cultural differences and diversity. Time alone will tell.

The second feature to emerge from this consideration of the explanations advanced for Maori educational underachievement is Harker's point that empirical research on the matter is both small in amount and fragmented.

Given that Maori educational achievement is a matter of some concern, and the continuing research and comment referred to above suggests it is, then further research that attempts to accommodate the cultural difference perspective and contribute to the 'body of empirical research' in, hopefully, an integrative manner would certainly seem justified.

Cultural Deficit and Cultural Difference Views

Along with changing viewpoints on the nature of the possible relationships between social groups in multi-ethnic and multi-cultural societies there have been important changes in the premises underlying much research and comment on the educational performance of pupils from ethnic and

¹ The Maori word Pakeha is in common usage denoting a New Zealander of European (Caucasian) descent.

cultural minority backgrounds.

The concept of cultural deprivation, cultural disadvantage or cultural deficit gained prominence in the early 1960's and throughout the decade the related educational strategies of compensatory and 'enrichment' education dominated efforts to bring about improvements in the educational attainments of minority group pupils. Texts such as Riessman's (1962) The Culturally Deprived Child did much to popularise the cultural deprivation perspective. The content and the shortcomings of the deficit concept are perhaps best illustrated by the nature of the criticisms that were to be levelled against it.

Gallimore and Howard (1968), for instance, attacked cultural deficit models on the grounds that they were based on culturally restricted views of appropriate behaviour. It was their view that action premised on such models had led agency or institution personnel to regard the so-called culturally deficit person as being 'sick' or as an 'empty-cup' that must be filled with the right kinds of skills, attitudes and motives.

Very similar views were also starting to be expressed in sections of the New Zealand educational research community. Watson (1967) saw a striking parallel between what was being assumed and said about "..... the home and mind of the Maori child....." (p.13) and what had been assumed and said about the American Indian child. In the latter case, the assumptions and comments had been challenged on the grounds that they were premised on a 'vacuum ideology'. In a subsequent paper, Watson (1968) extended his arguments against the adoption of a cultural deprivation perspective to the whole question of Maori education arguing that -

"..... it makes no sense to label the Maori child as culturally deprived, whatever his contemporary disadvantages in acquiring a capacity for enjoying, imagining, choosing, criticizing or altering the culture in which he finds himself in the future. Moreover, if (educational) proposals were to imply an abandonment of Maori culture we would not only be guilty of an austere, bleak and negative answer to the basic problems of our community, but we would also be committing our pupils to a psychologically improbable and unsatisfying future" (p.14).

Focusing specifically upon the implications for psychological theory, Cole and Bruner (1971) have addressed themselves to the question of deficit and difference models. They chronicled a long list of deficits that were said to afflict the children of poverty and those from minority group backgrounds. They noted that -

"Such data seem to compel the conclusion that as a consequence of various factors arising from minority group status (factors affecting motivation, linguistic ability, goal orientation, hereditary proclivities to learn in certain ways - the particular mix of factors depends on the writer), minority group children suffer intellectual deficits when compared with their 'more advantaged' peers' (p.867-868).

Cole and Bruner then argued that from both anthropological and linguistic perspectives there is "..... a body of data and theory that casts doubt on the conclusion that a deficit exists in minority group children" (p.868). They suggest also that the evidence "even raises doubts as to whether any non-superficial differences exist among different cultural groups" (p.868, Cole and Bruner's italics).

In terms of educational practice, Cole and Bruner see at least two clear implications. First, they argue for a "..... recognition of the educational difficulties (of children from disadvantaged sub-cultural groups) in terms of a difference rather than a special kind of intellectual disease" (p.874, Cole and Bruner's italics). Second, they suggest that teachers much discard the impression that they are there to "..... create new intellectual structures and start concentrating on how to get the child to transfer skills; he already possessed to the task at hand" (p.874, Cole and Bruner's italics).

A further and important perspective to the deficit versus difference debate has been introduced by Valentine (1971). Valentine has reviewed the tenets of both models as applied to Afro-American culture and finds both wanting. Deficit models are rejected on the grounds that -

"Any theory of class or racial deficits of biological origin is quite undemonstrable, indeed scientifically untestable, in an ethnically plural and structurally discriminatory society. The necessary separation of biological and socio-cultural factors is methodologically impossible in this setting" (p.138).

In addition, he suggests that deficit models rest on assumptions concerning Afro-American pathologies and shortcomings erected by arrogant cultural elites.

Difference models which have now become the respectable alternatives in more recent educational texts (e.g. Ginsburg, 1972; Keddie, 1973) have however, also been called into question by Valentine. He describes in his paper the difference model particularly as propounded by Baratz and Baratz (1970) but claims that the model has been over-extended "..... into a simplistic portrayal of cultural separation" (p.140).

In Valentine's view -

"The central theoretical weakness of the 'difference model' is an implicit assumption that different cultures are necessarily competitive alternatives, that distinct cultural systems can enter human experience only as mutually exclusive alternatives, never as intertwined or simultaneously available repertoires" (p.141).

It is therefore Valentine's thesis that "Both the 'deficit' and 'difference' models neglect and obscure the important concept of 'biculturalisation'" (p.141).

It would appear on reflection then that there exists a measure of congruence between what is being said about the nature of the evolving social structure of New Zealand in relation to its ethnic and cultural communities, the questions that are being directed towards assumptions and practices within the educational system and the emerging theoretical premises likely to underscore contemporary research in this area.

The Involvement of Intelligence

In the many publications that examine the differences between the educational attainments of European and Maori pupils, the question of intelligence regularly gains a mention. As Harker has noted, numerous assertions have been made on the presence or absence of differences in the level and/or distribution of intelligence between European and Maori pupils with little regard to the empirical evidence, such as it is,

the definition of the concept or the method of assessment. This being the case, the question of the measurement of the intellectual ability of Maori students is obviously of some importance to the broader issue of educational underachievement. This is no doubt why Harker included 'intellectual functioning' as a discrete factor in his interaction model of suggested factors involved in the low educational attainment of Maori pupils. The question of the assessment of the relative intellectual abilities of European and Maori pupils is, of course, far from new as the review presented in the following chapter shows. The long history of investigations bearing upon this question suggests that it has been of importance to the New Zealand educational research community. A number of events in the last decade may also have added to the significance of the question.

Reference was made above to some of the recent facts and figures concerning Maori educational attainment as reported by Ramsay (1975). The continuing prominence given to the question of Maori educational achievement suggests that there is little room for complacency.

Forster and Ramsay (1969) in examining the place of the Maori population in New Zealand society, adapted Myrdal's (1957) idea of interlocking and overlapping factors in a spiral of 'cumulative causation' which have conspired to limit the opportunities of Maori people economically, occupationally, politically, educationally and socially in the context of the wider New Zealand society. Their analysis would appear to be supported by observers such as Colgan (1972) and Garrett (1973) who have suggested that New Zealand is in danger of creating a very distinct 'brown proletariat'.

Ramsay (1975) has again recently employed Myrdal's 'cumulative causation' when addressing himself to the 'school's role in the revolution' needed to reverse the downward path of the spiral. Here Ramsay was prepared to point out that -

"A question which is usually avoided by writers when discussing the Maori situation is the relative intelligence, or levels of intellectual functioning, of the various ethnic groups. Although many authors specifically reject the notion of Maoris being intellectually inferior, they do so on a basis of very slim evidence" (p.298).

The issue of Maori ability, skill and talent has also come to the fore in ways other than by reference to the question of measured intelligence as such.

Watson (1967) was responsible for a challenging review of many issues in Maori schooling and argued that in human resource terms, the Maori population was one of the large and important pools of untapped talent. Addressing himself to the question of the economic well-being of New Zealand, Watson suggested that -

"In a period of rapid economic growth and acute labour shortages, employers have begun to appreciate that the Maori population contains a great reserve of talent and skill awaiting development and deployment" (p.4).

This view echoed sentiments expressed in a Department of Industries and Commerce (1967) publication on the need for the increased utilization of the skills and talents of the Maori population in a much wider range of economic activities than had previously been the case.

The under-utilization and under-investment in the skills and talents of the Maori population was also referred to in both the New Zealand Educational Institute Report and Recommendations on Maori Education (1967) and the New Zealand Post Primary Teachers' Association Report on Maori Education (1970). In each case figures were produced to show the cost to the nation of the continued failure to make, in their view, adequate investments in the field of Maori education. It was argued that through increased investment in this area New Zealand would benefit through both the direct economic gains from a more skilled and talented workforce and from a general reduction of social economic and political inequalities which might otherwise further divide the nation on ethnic or 'race' considerations.

Recent events external to New Zealand have also intruded into the question of the comparative performances of European and Maori samples on tasks attempting to assess intellectual abilities. The suggestion by Jensen (1969a) that hereditary factors might be involved in the observed ethnic group differences in intelligence test performance has not passed unnoticed. Articles and correspondence bearing upon Jensen's argument appeared in the Christchurch Press (Beattie, 1971: Countryman, 1971a, 1971b; Gregson, 1971a, 1971b). Furthermore, in subsequent research Harker (1971a, 1973), Brooks (1975) and St. George and St. George (1975) have all found it necessary in studies involving the measurement of intellectual abilities of European and Maori samples to address themselves to the hereditary-environment debate which has been thrust into recent prominence in psychological and educational discourse. Aspects of this question were considered in the preceding chapter.

Ethnicity or Social Class?

An important and continuing issue in relation to European and Maori differences on many variables, including educational attainment, is whether or not such differences are to be attributed primarily to social-class or to ethnic group factors.

It has long been observed that the Maori population experiences higher rates of infant mortality, shorter life expectancies, inferior standards of housing, a higher rate of offences against the law, lower per capita income, disproportionate levels of employment in unskilled and semi-skilled occupations and a lower level of educational attainment when compared with the European population (see Hunn, 1960; Walsh, 1973).

To some observers this state of affairs is seen as a reflection of the lower-social class position occupied by Maoris rather than the fact of 'Maoriness' in any cultural or ethnic sense (Burch, 1967; Mitchell, 1970; Adams, 1970, 1973; Gregory, 1974; Hermansson, 1974). Burch (1967) commenting on the position of both New Zealand Maoris and

American Blacks argued that -

"Many of the observed and reported behavioural characteristics of ethnic groups are not necessarily a function of skin colour, nor of racial category, nor even of immutable cultural traits, but rather are a function of one's place in the social structure, his class position" (p.120).

Similarly, Mitchell (1970) writing on the education of the disadvantaged child in New Zealand held the view that -

"Although there is a disproportionately high incidence of cultural deprivation among the Maori population, it seems to me that the validity of this distinction between the problems of the Maori and the problems of the culturally deprived is open to serious question. In focusing on a particular, statistically convenient, ethnic group there is a twofold risk of overlooking the possibility that certain environmental factors of significance to education cut across racial divisions and of ascribing to the total ethnic group the characteristics possessed by some or even a majority of its members" (p:181).

Mitchell went on to argue that one could seriously question the utility of a racial distinction over socio-economic or psychological divisions when trying to examine or explain differences in educational achievement and cited the studies of Lovegrove (1966) and Barham (1965) to support his contention that when matched for socio-economic status, then educational achievement and English language competency differences between Maori and European were few. It is important to note, however, that educational achievement and English language competency differences were found in these studies despite the matching on socio-economic status, and that the methods of analysis adopted by Lovegrove have been seriously questioned by Harker (1976). Furthermore, subsequent studies have shown ethnic group differences in educational attainments where socio-economic status matching was attempted (e.g. Codd, 1972; Geraldine McDonald, 1975).

Since Lovegrove is often cited by researchers and commentators in support of socio-economic determinants of educational achievement over ethnic determinants, some of Harker's points of criticism should be noted. First, he points out that Lovegrove's use of the term 'significance' is confusing in that sometimes Lovegrove is referring to the statistical

significance of a finding, while at other times the term is used in an intuitive sense in attempting to give weight to a finding for the reader.

Second, Harker takes issue with Lovegrove's use of one and two-tailed tests of significance. While the former were used in the within-group analyses, the latter were employed in the between-group analyses. Harker points out that Lovegrove had previously formulated hypotheses concerning achievement differences and that therefore one-tailed tests were more suitable for the between-group comparisons. Harker has re-analysed these data employing directional statistical tests finding significant differences at the .03 level. The problem as pointed out by Harker is that in quoting Lovegrove in support of socio-economic over ethnic determinants in educational achievement, writers have been content to use Lovegrove's reporting of the data - not the data itself.

Both Gregory (1974) and Hermansson (1974) have also expressed the view that socio-economic and ethnic factors are intertwined, but in the final analysis argue that the educational difficulties of Maori pupils stem from their lower social class background.

It has been Adams (1970, 1973) however, who has given this controversy most attention in its own right. In two papers Adams (1970, 1973) has argued that if Maori educational underachievement cannot be attributed to some 'built-in propensity' such as intellectual inferiority then the causes must be sought in the environment. An 'environmental' path towards educational advancement and achievement is then plotted as seen through European sociological eyes.

On the issue of Maori educational underachievement, Adams intimates that the 'built-in propensities' are genetically determined, but they are not examined further. Turning to environmental factors he argues that 'Maoriness' in the life-style sense needs to be invoked and that the assumption must be made that this life-style is deprived since educational disadvantage and underachievement are the results for the participants.

Adams notes, however, that deprivation is not necessarily the prerogative of the Maori people and suggests therefore, that the problem is primarily one of understanding and alleviating deprivation or disadvantage in general. It would appear that Adams sees this disadvantage as being primarily socio-economic in character and cites Lovegrove's studies to the effect that -

"..... when socio-economic level is controlled and families with roughly the same income are compared, ethnic origin (Maoriness) does not predict to failure or disadvantage" (p.65).

Adams has concluded his argument with the assertion that "Our problem is to alleviate the disadvantage that exists, and that problem is not, of itself, an ethnic problem at all" (p.67).

A number of the above arguments are, however, open to question. Adams appears to have adopted a rather restricted perspective on the genetic versus environmental factors question. In the study of behaviour-genetic relationships dichotomy models are giving way to interaction and interpenetration models when considering the development of human characteristics both within and between subgroups (see Hirsch, 1968). Thus, Adams' reading of 'built-in propensities' as necessarily genetic and the genetic-environment debate as an 'either/or' issue is, therefore, a too radical simplification.

Furthermore, the implication of Adams' argument is that once disadvantage has been overcome and educational achievements equalised between two groups then occupational success - and presumably all other forms of success in European eyes - will be achieved. In a small way this proposition has been put to the test by Geraldine McDonald (1975).

McDonald found that when a sample of fathers of Maori and European children were matched on the basis of level of formal education their occupational statuses were not, in fact, the same. Thus in her words -

"If education created equal opportunity for Maori and Pakeha then they (the fathers) should be distributed equally in occupational statuses. But they are not. And unless education does create equal opportunity for Maori and Pakeha then there is little point in saying that ethnic origin 'does not predict disadvantage'. Therefore, finding that lack of academic success is associated with low socio-economic status in the Maori group cannot provide a sufficient explanation for the facts" (p.79).

McDonald notes that Adams put the question "What useful social purpose (does) the Maori-Pakeha differentiation serve?" (Adams, 1973, p.66). Her answer is that -

"The Maori-Pakeha differentiation serves no useful purpose to the Pakeha majority. In fact, it embarrasses it. Its purpose for the Maori population is that it reflects a social fact." (p.80, italics added).

In discussing studies concerned with Maori educational under-achievement, McDonald also argues against some other dichotomies and distinctions that have been employed. Thus she finds fault with the urban/rural dichotomy arguing that particular communities and groups must be categorised in their own terms according to differences in cultural expression and experience. Her point is that the outwardly similar cultural elements (e.g. level of education, socio-economic status, family type and so on) may be imbued with very different meanings and result in different patterns of social interaction for persons of different ethnic groups.

Aspects of McDonald's paper have drawn comment from Adams (1975). While not really contesting the possibility that low socio-economic status is not a sufficient explanation of low Maori educational achievement, Adams questions the adequacy of the evidence to date and calls for more research.

Adams also takes issue with the treatment of his remarks concerning the social purpose of differentiating between Maori and European. He argues that his remarks were made in order to point to the underlying importance of identifying causes of educational disadvantage rather

than offering 'explanations' on, as Adams implies, arbitrary ethnic categories.

Here Adams may have missed the point of McDonald's paper, for she points out that the division is not arbitrary but a social reality - especially for Maoris. In addition Adams seems to have adopted, in this instance, a rather narrow view of what authors attempt to convey by the term 'ethnic group'. A reading of the research of Bray (1971), McDonald (1974a, 1974b) or Ritchie (1963) for instance, leaves the impression that it is not ethnic group in any narrow racial sense of strict biological descent that is being talked about, but rather a group now bound together by (i) common attitudinal and value orientations, (ii) common conceptual frameworks, (iii) behavioural repertoires that display important similarities, and (iv) a shared sense of history and assumed common descent. Thus ethnicity as 'race' itself is not invoked to explain events, but rather ethnicity implying communalities in affect, cognition, behaviour, history and assumptions about descent. In time these variables might be incorporated into suitable causal-comparative research models. For the present however, there can be no debate with Adams over his point that the research has not been done.

Because of lingering doubts, it is appropriate to consider two issues in this area in a little more detail. First, it has already been noted that McDonald questioned the imposition of dichotomies or scales from one cultural context upon another. In the case of socio-economic status scales, we find not only some disagreement over their content for the European New Zealand population (see Brooks and Cuttance, 1973; Elley and Irving, 1974) but also a little evidence that status in Maori society is influenced by factors not included in such scales.

Fitzgerald (1968) in a study of the social position of Maori university graduates found the most frequently mentioned determinant of status in Maori society to be one's ancestry and genealogical ties. Thus kinship and the importance of kin over-rode factors such as level of

education and income and the idea of equating Maori and European on a single socio-economic status scale might well be suspect.

This suspicion has received some support in the recently reported study by Harker (1976) into the cognitive style preferences of Maori and European children in relation to school achievement and environmental circumstance. Harker used fathers' occupation and the Elley and Irving (1972) scale to determine socio-economic status rankings. Controlling for ethnic group and urban or rural samples, Harker compared the performances of high and low socio-economic status groups on school achievement criterion.

It was found that socio-economic status was a significant discriminator for the European group in terms of school achievement, but not the Maori group. In Harker's words -

"..... it appears that the assertion which explains Maori-Pakeha achievement differences in terms of differential distribution on the SES hierarchy is untenable, as SES makes little or no difference within the Maori group" (p.68).

The second concern lies with the question of whether or not Maori and European differentiations should be made at all. McDonald believes the differentiation to be a social reality for Maori people. The evidence on the holding of stereotypes, of prejudicial attitudes towards Maoris by Europeans, not to mention discriminatory behaviour (see Archer and Archer, 1970; Vaughan, 1972) is more than adequate testimony to the social reality of the Maori-European differentiation for Europeans as well.

This would suggest that ethnic group membership does act as a 'social marker' in New Zealand. For European and Maori alike, ethnicity is a visible and convenient characteristic that may act as both a trigger and determinant of subsequent interpersonal events. It might not be surprising then if in everyday life socio-economic considerations are easily subordinated to the readily perceived ethnic group identifiers borne by the population of New Zealand.

Perceived ethnicity as a social marker is an idea explored more fully by Banton (1967) through his concept of 'race as a role-sign'. This embodies the view that perceived ethnicity in both the biological and behavioural sense acts as a role-sign which may have the effect of patterning subsequent interpersonal and social events between people.

Banton suggests that "Race is a role sign only in multi-racial societies or in situations of racial contact in which expectations of behaviour have crystallised into patterns of some sort" (p.58). It certainly is the case that New Zealand is a multi-racial society and the data on European and Maori stereotypes of one another is indicative of 'expectations of behaviour crystallising into a pattern'. The evidence and arguments reviewed do lend weight to the proposition that ethnicity continues to be an important parameter and reflects genuine divisions recognised and acted upon by New Zealanders.

In two recent papers Havighurst and Glazer have added interesting perspectives to the question of social class and ethnicity. Havighurst's (1976) paper considers the question of the relative importance of ethnicity and social class in human development. He contends that in a complex society it is the ecological structures of social class and ethnicity that are primarily responsible for differences in life style and development. In his view "Members of a complex society are members of a given social class and also of a given ethnic group, at least potentially, though they may play down or refuse to recognise membership in one or the other sub-cultures" (Havighurst, 1976, p.56).

A social class is seen by Havighurst as possibly containing several ethnic groups within it, just as an ethnic group, if large, may be composed of more than one social class. As a result "..... there is a kind of competition between social classes and ethnic groups" (Havighurst, 1976, p.56). This raises the question of which takes precedence, which

most accurately describes the person, ethnicity or social class? To Havighurst the answer is simple - preference will be shown by the behaviour of the individual.

With evidence from the United States, Havighurst then illustrates the ways in which social class and ethnic group factors might interact or dominate the socialization process and life-styles of various communities. It is his view that contrary to the suggestion that group differences are primarily social class differences, ethnicity is "..... a rather powerful influence on socialization processes" (Havighurst, 1976, p.63). Furthermore, he has put forward the view that the cultural pluralism movement has been advanced as a way of "..... maintaining democratic co-operation among ethnic groups that are interacting in various ways with the social class structure" (Havighurst, 1976, p.63).

Glazer (1975) in a wide ranging essay on ethnicity, discusses the development of ethnic allegiances as a new world phenomenon in group relationships, and conflict, which has made the identification of groups of people on the criteria of race (in the biological sense), religion or social class less meaningful than previously was the case.

Glazer gives attention to the question of ethnicity and social class in terms of conflicts of interests that have arisen within societies. It is his view that -

"..... the Socialist hope for a trans-national class struggle, based on class identification, never came to pass. Instead, it has been replaced by national and ethnic conflicts to which combatants have often tried to give a 'class character' (Glazer, 1976, p.41).

In Glazer's view, and the evidence and examples he cites certainly supports his view, the terminal community that effectively commands loyalty "..... is increasingly the ethnically-defined community rather than any exclusively interest-defined group" (Glazer, 1975, p.42, Glazer's italics).

In summary, it would appear that the Maori/European distinction is both real and appropriate to the present social circumstances within New

Zealand. Such a distinction based upon an ethnic group concept, with both cultural and assumed common descent components, is pertinent to many issues in New Zealand including educational achievement. Both nationally and internationally the increasing importance of ethnic group distinctions suggests that they will parallel and indeed supersede divisions within societies once based primarily upon social-class considerations.

Summary

The evidence presented in this chapter demonstrates that in terms of contemporary scholastic criteria Maori children perform less well than their European peers. It has been an issue that has been subject to extensive commentary and research. It was also seen that over the years the tenor of the 'explanations' offered for Maori educational underachievement have shifted from deficit to difference viewpoints.

A question that has been raised on a number of occasions has been whether or not ethnic group differences in the levels of intelligence and modes of intellectual functioning might be involved. As will be seen from the review of studies in the following chapter, much of the research in this field is inconclusive and few, if any, of the tests employed have been developed and adequately researched for cross-cultural use.

The importance of the ethnic group differentiation of the New Zealand population was also considered. It was argued that it is a social reality and a primary social force in this country. Consequently, it is important to continue to monitor and attempt to understand social, cultural and psychological phenomena in ethnic group terms.

CHAPTER 3

MAORI INTELLECTUAL ABILITIES

Anecdote and Opinion

The European colonisation of New Zealand brought with it numerous accounts and commentaries upon the customs, habits and attributes of the indigenous inhabitants. Writings on the ways and life of the New Zealand Maori ranged from the serious such as the meticulous documentation found in Cook's journals (see Beaglehole, 1955-1969) to the more sensational as in the account by Travers (1872) on the life of Te Rauparaha. The writings covered every facet of Maori life - physical features, social organisation, economics, politics, arts, crafts and the material culture. Attempts were also made to describe the 'Maori mind' and the level of intelligence of these 'New Zealand Aborigines'.

The preamble to the Native Trust Ordinance of 1844 expressed the view that -

"The Native people of New Zealand are by natural endowments apt for the acquirement of the arts and habits of civilized life, and are capable of great moral and social advancement" (cited by Barrington and Beaglehole, 1974, p.39).

Beaglehole (1955) in his thesis on the beginnings of Maori schooling records that -

"There was general testimony to the keenness and ability of the Maori pupils. Mr T.D. Jowell told the 1838 Committee of the House of Lords on New Zealand, that they were as intelligent as any children he ever saw anywhere....." (p.17).

In much the same vein Sutherland (1929) reports the Reverend Samuel Marsden as saying of the Maori -

"They are a noble race, vastly superior in understanding to anything you can imagine in a savage nation I do not believe that there is in any part of the world, or ever was, a nation in state superior to the inhabitants of New Zealand in mental endowment and bodily strength, nor any who would in a shorter period render themselves worthy of being numbered with civilized nations, provided they were favoured with the ordinary means of instruction by which men are gradually refined and polished" (in Sutherland, 1929, p.130).

John Thornton (1892) headmaster of Te Aute College shared Marsden's high opinion of the Maori. He found the Maori to be -

"..... loyal, intelligent, good natured and gay. He possesses moreover, a keen sense of justice, and, as regards honesty and truthfulness, he at least stands on a platform equally high with the average of mankind" (p.202).

More recent Maori commentators have also expressed similar views. For instance Love (1943) stated that -

"The mentality of the Maori was impressive considering his isolation from other races. His powers of reflection, his comprehension of all things in his life and his knowledge in the realms of astronomy, mythology and tradition were considerable" (p.61).

However, this is not to say that contrary views were not expressed. Lovegrove (1964) noted that not all who came into contact with the Maori during the early period of European settlement held the Maori and his abilities in high regard. To some he was "a distasteful hurdle to be overcome" (Lovegrove, 1964, p.6). The Daily Southern Cross of 11 July, 1851, said of the Maori that "The sultry, scowling, bilious-eyed folks are, perhaps, the most disagreeable of the Pacific Islanders" (in Lovegrove, 1964, p.6).

Lovegrove (1963) has written specifically on these early views and attitudes of European settlers on the intellectual abilities of the Maori and concluded that the predominant view was that "In terms of intellectual capacity there was little difference between Maori and European....." (Lovegrove, 1963, p.81).

Early commentary on the intelligence of the Maori was concerned not only with its likely level but also with its organisation. The faculty psychology of the times no doubt influenced the attempts to compartmentalize the qualities of the 'Maori mind'. For the early observers the sovereign quality of Maori intellect was memory or retentive capacity. In keeping with the oral tradition of the culture prior to European contact, commentators marvelled at the lengthy and complex

genealogies Maoris were able to recount unaided. Shortland (1851) describing Tuhawiki, a Maori chief, remarked that he -

"displayed that remarkable power of memory at which I have often wondered in the New Zealander, repeating a long list of miscellaneous property, which he and others had received at different times, specifying what share each had obtained on division" (p.80).

Thomson (1859) who wrote widely on New Zealand at this time commended the Maori for his singular power of memory, but went on to condemn him for his lack of imagination and inventiveness. Thomson argued that the Maoris were deficient in reasoning power and thus crippled intellectually. Furthermore, Thomson asserted that the 'minds' of the Maori were unregulated, being deficient in the powers of steady and continuous attention, of association and mental industry. Thomson's views on the lack of mental diligence were supported by Polack's (1840) earlier comments that the Maori was at times found wanting "in stamina of mind for long and pertinent deliberations" (Vol 2, p.116).

The view that the organization of intellectual qualities was somewhat different for the Maori was to be found amongst Maoris themselves. In 1929 Sir Apirana Ngata said -

"The writer as a Polynesian accepts without reservation the dictum of students of Maori mentality, that the race had not attained to such a command of ideas and of language to express them, as to have been able to use abstraction and generalisation" (cited in Lovegrove, 1963, p.78).

Although, as Sutherland (1929) was quick to point out, Ngata's form and method of argument was in point of fact a clear demonstration of the invalidity of the assertion.

Much as comment varied on the level of Maori intelligence, so too did comment on intellectual qualities, with perhaps the exception of 'memory'. Unlike Thomson or Polack, Martin writing in 1845 and reported by Renwick (1963) thought the reasoning and reflecting powers of the Maori to be equal, if not superior, to those of most Europeans. In Martin's view no civilized people were more painstaking than the Maori in reaching decisions through full discussion.

No doubt the range of views expressed on the intelligence of the Maori during the first decades of European settlement reflected the circumstances and experiences of the commentators and the absence of a consensus over what was to constitute evidence of intelligence per se or the exercise of particular qualities of intellect.

While 'faculty psychology' and phrenology greatly influenced early views on the 'Maori mind' other views based on historical and cultural differences, rather than distinctive cranial formations, were also to be found. Marsden anticipated that the Maoris would soon be 'civilized' as a result of a correct and proper education. Nonetheless, both views were essentially deterministic in their outlook. The 'faculty - phrenological' explanation of differences in intellect because cranial typology was essentially innate and the 'historical - cultural' explanation because the inexorable path to civilisation was that charted by the Europeans.

Both perspectives on the reasons for the differences between Maori and European intellect could be reconciled with the 'social Darwinism' of the day, a perspective which Sinclair (1971) asserts was "widely held and believed among Europeans in New Zealand....." (p.125). It mattered little whether the criterion was cranial protuberances and associated faculties or the state of civilisation of the indigenous peoples. In the grander scheme of things the 'New Zealanders' were slotted into the hierarchy of mankind. A hierarchy in which the predominant position had been confidently claimed by the men and women of Europe.

While accepting 'social Darwinism' the European settlers of New Zealand were much comforted by the opinion that the Maori was in all ways superior to all other known savage races. Renwick (1963) writing on Victorian views of the Maori reported that -

"There was general agreement that, compared with the Australian aborigine or the American negro, the Maori was particularly well endowed to respond to civilizing influences" (p.365).

Sinclair (1971) in analysing race-relations in a number of countries has included such arguments as perhaps one of the contributing factors to the better state of race-relations in New Zealand. He points to an interesting twist in the argument noting that -

"The pakeha, who claim to be descended from 'selected' British stock (and to be definitely non-Australian) take pride in the superiority of their Maori compatriots over Aborigines or other (supposedly inferior) races" (p.124, italics added).

Perhaps it was not surprising then that the most select of British colonists should have had the good fortune to civilize the most select of savages!

The merits of positions adopted and arguments put forward on the intelligence of the Maori by the early commentators must, of course, be seen in the light of their times. As far as one can summarize these early views, something of a mixed position has to be accommodated. On the one hand it was widely held that the Maori was the intellectual equal of the European while on the other hand, aspects of his culture and behaviour gave weight to the view that in some ways he must 'think' very differently. It might be suggested that the European settlers as egalitarians and pragmatists were content then, as now, to adopt a crude 'environmentalist' position. While acknowledging the equality of Maori and European in intellectual potential, this potentiality was seen as being at times unrealised by the Maori. But there was always the hope that time and education would change his ways, his thinking and ultimately all other differences. (see Lovegrove's, 1972, account).

Reasons for the failure of the Maori to fulfil this intellectual potential were seen by some to have resulted from the absence of a stimulus such as European civilization. Thomson (1859) expressed the view that the lack of intellectual activity on the part of the Maoris in pre-contact times had caused the brain to shrink. It is not clear whether Polack (1840) speaking of post-contact Maoris had witnessed a sudden increase in brain size, but he did remark that the Maoris were

perceptive enough to realise the benefits from contact with European civilization and as a result "an improvement has taken place in their mental as well as their physical capacities" (cited in Lovegrove, 1963, p.79).

A somewhat different view involved the idea of a 'fall from a state of intellectual grace' on the part of the Maori. Sutherland (1929) reports William Colenso saying of the Maoris that -

"Their intellectual and moral faculties, as a race, were of a high order; however stunted, warped and debased they may have been through custom, habit or their strong and unrestrained natural propensities" (cited by Sutherland, 1929, p.137).

While Colenso was commenting on the Maori of the nineteenth century, elements of this view that the Maori has misused or abused his intellectual talents have persisted. For instance, Ball (1940), a prominent teacher and administrator in the field of Maori education stated that "The Maori is intelligent, but at present his intelligence is a powerful tool more often than not used for petty purposes" (p.280).

Notwithstanding the above variation in the arguments, the great burden of comment and speculation favours a view of equality between Maori and European in general intellectual ability. This position has become an almost foundation premise to the many reports concerned with Maori social advancement in general and Maori educational underachievement in particular. (A short list of reports and articles in this area can be found in Ramsay, 1975, p.299-301).

The only voice of dissent in the literature appears to be that of Ausubel (1961). It was his contention that in New Zealand the view that the Maori was inherently inferior in intellectual endowment was frequently advanced. Certainly in terms of commentators both past and present his contention is not borne out. The beliefs of the general population on this matter remain, of course, another question.

Besides opinion and anecdote, there has grown up a considerable body of empirical evidence on the comparative intellectual abilities of

Maori and European. The comparative studies are reviewed in the next section.

Comparative Studies on Levels of Intellectual Ability

Probably the earliest comparative study of Maori and European intelligence that could lay claim to a 'scientific' frame of reference was that by Thomson (1859). Reflecting the influence of phrenological theories and relying on anthropometric devices Thomson reported that -

"It was ascertained by weighing the quantity of millet seeds skulls contained and by measurements with tapes and compasses that New Zealanders' heads are smaller than the heads of Englishmen, consequently the New Zealanders are inferior to the English in mental capacity" (p.81).

Thomson, by way of explanation, went on to remark that -

"This comparative smallness of the brain is produced by neglecting to exercise the higher faculties of the mind.....it is only natural that generations of mental indolence should lessen the size of brains" (p.81).

As far as can be ascertained the next comparative studies did not take place until the late 1920's and early 1930's when McKenzie (1930, 1931) undertook the 1927 and 1929 'Taranaki School Surveys' covering Standard 3 to Standard 7 pupils. With the assistance of Ball the surveys were extended to a large sample of Maori pupils in 1930. In the period between the work of Thomson and that of McKenzie the field of psychometrics, the intelligence test and the achievement test had all come of age and the new science was already being put to use in New Zealand (see Marsden, 1925).

For their studies, then, McKenzie and Ball had at their disposal both measures of intelligence and educational attainment based upon psychometric principles. In their studies they administered the Haggerty Tests of Educable Capacity along with supplementary tests of reading, arithmetic and spelling finding that in the main the European ('British') pupils performed better than the Maori (and 'foreign') pupils, and it was duly noted in the 1931 report of the Inspector of Native Schools that on

assessments of "educational capacity" European children were superior. (AJHR, E-3, 1931). The evidence was not, however, seen as conclusive and reasons for the differences were suggested. McKenzie (1931) concluded that on the basis of the surveys, the 'evidence of history' and his personal knowledge -

"The whole evidence tends to prove that the Maori is the intellectual equal of the European. Any apparent disparity is due to the language difficulty and to the Maori's relative lack of social, educational and vocational opportunities" (p.203).

Butchers (1930) in reviewing education in New Zealand also noted these surveys and like McKenzie argued that the poor command of English on the part of Maoris must be taken into account in interpreting differences on the survey measures. A decade later Ball (1940) reiterated the view that the lower performances by Maoris on the survey measures stemmed from language difficulties. Ball also attempted to view the results more positively by stating that -

"All the tests were in English and the results were particularly creditable because the Maori children, in a foreign language, so nearly reached the norms established by English children in their own tongue" (p.283).

Investigations concerned with the intellectual abilities of the Maori had gained some recognition through the work of McKenzie and Ball and the research effort gathered momentum. However, as will be evident, the problems of research in this field have persisted and the 'answers' remain elusive.

Cooke (1933) found the poor command of English by Maori children a restrictive element in his research. Cooke intended to study aspects of the English ability of both Maori and European children "of equal ability as measured by IQ....." (p.5) but was forced to conclude that the linguistic nature of tests of ability precluded their satisfactory use with Maori pupils. It was inevitable then that the so-called non-verbal tests would attract the attention of New Zealand researchers and it was Winiata (1944) who took the first steps in this regard. Winiata was intent upon constructing "a Non-verbal Test of General Ability for

Maori children" (p.1). This he attempted by modifying Mann's (1935) Fiji Test of General Ability, after trials with samples of Maori children. Winiata did not actually undertake comparative studies with European subjects and there is no evidence that his studies were extended in this direction. For his own part Winiata argued for a non-verbal measure of intelligence to identify able Maori children for special and different educational treatment. A task, in his own words "..... vitally related to the scientific approach to Maori problems" (p.28).

Hearnshaw is reported by Ernest and Pearl Beaglehole (1946) to have administered tests of dexterity, attention, observation and general ability to samples of Maori and European males. On all measures the European males were superior and that the differences were attributed in this case to the failure of Maori subjects to adapt to the test situation. No other details of these studies appear to have survived.

Following World War II Downes (1948) reported on the ability testing of Maori and European male army recruits aged between 20 and 45 years. Downes had available to him data from Raven's Progressive Matrices, Bennett's Test of Mechanical Aptitude and Knowledge and the 'SP25' - a War Office verbal ability scale. It was reported that mean performances favoured Europeans on all three measures.

Comparative studies of the intellectual ability of Maori and European subjects increased in number during the 1950's. Calvert (1950) undertook a study into the English language difficulties of Maori adolescents and as part of her testing programme administered the Otis Higher, Raven's Progressive Matrices and the A.C.E.R. Non-verbal Test. On all three instruments the median performances were higher for European subjects though the differences were less pronounced on the Raven and A.C.E.R. tests. In addition, Calvert investigated the performance of a further sample of Maori subjects on the Otis test alone. The distribution of scores evidenced a marked positive skew with a median IQ of 88 and the item difficulties were at variance with those established

for European samples.

Logan (1952) like Calvert, made a study into the English language difficulties of Maori subjects. He compared samples of Maori and European Teachers College students on the Otis Higher, Raven's Progressive Matrices and the Mill-Hill Vocabulary Test. On all three assessments the mean performances of the European subjects were higher. Logan also made the point that the item difficulty order of the Otis Higher test was unsuitable for Maori subjects. In the same year Baird (1952) reported a study comparing rural Maori and European children on the Otis Intermediate Test with the results again favouring the European children.

Comparative studies employing individual ability measures were initiated by Walters (1953) and Adcock, McCreary, Ritchie and Somerset (1954) in the mid-1950's. Walters (1953) administered the Wechsler-Bellevue Intelligence Scale to Maori and European prison inmates matched for age finding that the European group obtained significantly higher Full Scale IQs. Adcock, McCreary, Ritchie and Somerset (1954) compared a small rural Maori sample, a Maori Teachers College sample and European Teachers College sample matched for age on the same intelligence test. The European group obtained a higher average Full Scale IQ than either Maori group. The comment was made, however, that the 'Performance Sub-Scale IQs' of the Maori samples more closely approximated the European scores than comparison between the 'Verbal Sub-Scale IQs'. A set of data from an unselected Maori sample was also noted. A year later another small study employing group measures was undertaken. Simpson (1955) reported lower median IQs for Maori secondary pupils when compared with Europeans on the Otis Higher and A.C.E.R. Non-verbal Test.

The major investigation in this field during the 1950's was, however, that of Leone Smith (1957). Her study was into the influence of English reading achievement on the intelligence test performance of Maori children. The intelligence tests employed were Cattell's Culture-

Fair Test, the A.C.E.R. Intermediate C and D, and Thurstone's Primary Mental Abilities Test. Reading achievement was assessed by Schonell's Graded Word Vocabulary Test and Test of Silent Reading. Two Maori samples were tested - urban and semi-rural, and three European samples - urban, semi-rural (Culture Fair Test only) and rural. The subjects were all aged between 11 years and 13 years 11 months and were from Standard 4 to Form II class levels.

Smith conducted analyses between Maori and European samples when matched for 'reading achievement' finding no significant differences on the intelligence test performances. In analyses within ethnic groups both Maori and European urban samples generally performed better than either semi-rural or rural groups. An exception to this was on the Word-meaning Sub-test of the Primary Mental Abilities Test where no difference was found between Maori urban and semi-rural groups.

Smith also made a number of comparisons between the performances of Maori and European subjects on the three intelligence tests in terms of the urban, semi-rural and rural groupings. A complex set of results was reported. These have been difficult to interpret clearly since data were not available for all equivalent groups. Lovegrove (1966) was to also note subsequently that there was a five month time lapse between the collection of the rural and urban data. This would have afforded the urban groups opportunities for additional educational experiences.

In the same year Ritchie (1957a) reported a further study with the Wechsler-Bellevue Intelligence Test (Children's Scale). Rural secondary school Maori and European samples were compared with the European Full Scale IQ being significantly above that of the Maori.

Walters (1956,1958), who had previously undertaken comparative research with the Wechsler-Bellevue Scale, returned to the area with a study employing Thurstone's Primary Mental Abilities Test and a 'Non-Language Test' of his own design. His study compared the performances

of Maori rural, market-garden settlement and urban samples with those of an urban European sample. No significant difference was evident between the Maori urban and European urban samples on the Primary Mental Abilities Test. However, the same comparison on the 'Non-Language Test' produced a significant difference in favour of the European urban sample.

Walters, like Smith, made comparisons between ethnic groups in terms of the urban, market-garden settlement and rural distinctions. A somewhat confused pattern of results emerged. In the light of his research Walters questioned the use of non-verbal tests with Maori and European subjects. The adequacy of his own 'Non-Language Test' was not, however, seriously debated in his reports.

McClew (1958), like others before him, sought to investigate the English language difficulties of Maori children. In the course of a study on reading achievement he compared matched samples of Maori and European primary school children on the Otis Intermediate, Jenkins Non-verbal and Thurstone Primary Mental Abilities Tests. He found only slight differences in the mean performances of the samples. McClew made a point of noting that on the Jenkins Non-verbal Test the Maori sample actually scored 2 IQ points above the European sample.

The 1960's brought further research and an expansion of the debate on the comparative intellectual abilities of Maoris and Europeans.

The Commission on Education (1962) considered at length the question of Maori education and acted as a stimulus for further comparative studies. The Commission itself received evidence on comparative studies during its deliberations and the report has acted as a baseline for subsequent research.

The Ministry of Defence (1960) made available to the Commission a report of results from nearly 8,000 18-year-old male National Service Cadets on the Army General Classification Battery. Data from this battery were reported in terms of a converted score known as a 'selection-

grade'. A comparison was made between 6,875 Europeans and 755 Maori cadets in terms of the distribution of selection grades. It was noted that while the distribution of selection grades was reasonably normal for the European cadets there was a marked positive skew for the Maori sample. Bar graphs of means and standard deviations based on normalised scores were presented for each ethnic group on the five subtests of the battery - namely, Mechanical Aptitude, Clerical Speed and Accuracy, General Intelligence, Verbal Ability and Mathematical Attainment. On all of the subtests the European mean was higher. The report went on to consider differences in the type of secondary schooling, level of educational attainment and the occupations of the cadet population. It did not, however, elaborate on the meaning or implications of the ethnic group differences.

Ausubel (1961) in Maori Youth reported the Otis Test performances of 'matched' samples of urban and rural secondary school Maori and European males. He found that the mean Maori urban IQ more closely approximated the mean European urban IQ in comparison with the rural samples, but that between group differences were still in themselves significant. In attempting to explain his results, Ausubel argued that tests such as the Otis reflect the "current level of intellectual functioning as influenced by both genetic and environmental factors" (p.92, Ausubel's italics).

Contrary to some earlier views on Maori-European differences on intelligence tests Ausubel contended that -

"It is futile to argue that because of language retardation or insufficient environmental stimulation, a Maori pupil's true intelligence is really higher than his score on an intelligence test indicates" (p.92, Ausubel's italics).

Arguments favouring the acceptance of test results from Maori subjects as valid had some support. Ashcroft (1963) published a short note on the use of the Otis, Raven's Progressive Matrices and the Good-enough Drawing Technique in assessing the intelligence of Maori secondary

school students. Maori mean performances on these measures were reported as being consistently below that of European students. As far as educational purposes were concerned, however, it was stated that the Otis test gave " a most useful and valid picture of both Maori and Pakeha pupils when used as a measure of general mental ability" (p.30).

A comparative study focusing on relatively young children was undertaken by Rogers (1964). He investigated the performances of Maori and European children aged 6 to 11 years on the Raven Coloured Progressive Matrices Test. Lower mean scores were reported for Maori children at each age level when compared with the combined European and Maori samples. It was also found that the Maori mean scores showed considerable fluctuation across the age range in contrast to the steady improvement in mean scores with European children with increasing age.

At about the same time Ritchie (n.d.) undertook a small research project in a Northland secondary school on the feasibility of administering Thurstone's Primary Mental Abilities Test and the Wechsler-Bellevue Intelligence Scale to Maori students. A report was made to the Maori Education Foundation which presented a limited amount of numerical data and noted certain difficulties encountered with the respective tests. More important, however, were some of the recommendations Ritchie made.

It was suggested that the Maori Education Foundation become actively involved in promoting research on the measurement of Maori intellectual abilities. An appeal was also made to the psychological profession in New Zealand to take some action in this regard. Furthermore, Ritchie saw a clear need for previous research endeavours in this field to be reviewed as a base for the co-ordination of future studies.

Probably the most influential study of the 1960's was that of Lovegrove (1964, 1966) into Maori and European scholastic achievement. With matched samples Lovegrove found no significant differences between Maori and European pupils in terms of scholastic achievement, but diff-

erences favouring European samples were evident on the intelligence tests administered. These were a battery comprising the Otis Intermediate and the Lorge-Thorndike Verbal and Non-verbal Tests - Level 3.

Lovegrove saw cause for concern in this finding of intellectual differences. Adopting an environmental stance he sought to explain his results thus -

"It appears likely that Maori parents are able to provide sufficient stimulation and support to assist in ensuring that their children make adequate scholastic progress during primary school years, but on the other hand it seems that the environments they provide are not conducive to the development of the complex intellectual processes assessed by tests of intelligence" (p.34).

Lovegrove found, however, that Maori subjects performed significantly better than European subjects on the listening skill and speed skill measures included in his study. He viewed these skills as 'intellectual adjuncts' and argued that high performance in such skills contributed more to the scholastic achievement of Maori children than their level of intellectual ability per se. In drawing conclusions from his findings Lovegrove expressed the opinion that "The fact that Maori children employ 'intellectual' abilities to a lesser extent in their school work should not be interpreted as indicating immutability" (p.35).

He went on to suggest the adoption of a "developmentalist attitude towards mental functioning; (and the) direct cultivation of the intellectual abilities of Maori children" (p.35). Furthermore, he noted that as a result of the level of performance of Maori subjects on tasks designated as 'intellectual adjuncts' more consideration must be given to such 'non-intelligence test' factors in education.

McCreary (1966) published a paper on the use of reading tests with Maori children. His study was undertaken in an isolated rural setting and here he found that on the Otis Intermediate test his Maori

primary school sample obtained a mean IQ of 87. This he suggested was well below the expected European mean.

In the 1950's both Calvert and Logan had suggested that the Otis Tests might be unsuitable for assessing the intellectual ability of Maori school pupils. On the other hand, Ausubel and others saw the Otis results as a fair reflection of the intellectual ability of Maori students. The fires of this little controversy were again fuelled for educationalists by du Chateau (1967). He had obtained a set of Otis Higher Test data from a sample of Form III pupils. The Maori mean IQ was found to be 84 in comparison with a European mean IQ of nearly 94. As a result du Chateau concluded that "there is likely to be a ten-point IQ gap favouring Pakehas at the third form level on group verbal tests of intelligence" (p.157).

It was argued that to accept the results of Maori children on the Otis Test at face value would "... result in a quite unrealistic demand for special class enrolment for Maori children" (p.158). It was also suggested that the definition of giftedness for Maori pupils was in need of review "... perhaps by lowering the verbal IQ limit" (p.158).

Rumour has it that staff of some schools found the solution to the 'ten-point gap' by simply adding 10 IQ points to the scores of all Maori pupils!

Gustafson (1967) undertook a small investigation into reading comprehension levels of Maori and European secondary school pupils. As part of the study data on intelligence was obtained using the Jenkins Non-verbal Intelligence Test. It was found that the mean IQ for Maori pupils was 93.5 in comparison with a mean IQ for European pupils of 113.

Comparative studies employing individual ability scales were very limited in the 1960's. The only one appears to have been that of Pera Jackson (1966) who undertook a limited study into the suitability of the WISC-Performance Scale. Data were obtained from a sample of children with the sample split into European and 'non-European' (mostly Maori) groups.

In terms of comparative performances Jackson noted only that some 'non-European' children had difficulty with the Picture Arrangement subtest.

This decade has so far seen an upsurge in research in this field. Archer, Oppenheim, Karetu and St. George (1971) joined the debate over the Otis test, albeit satirically, with the MOTIS test (Maori-Otis). Although in English the items required some knowledge of 'things Maori'. A small study showed European secondary students to have a mean Motis IQ of only 67 compared with that of 104 for Maori students. The distressing results led the authors to propose programmes of compensatory education for Europeans.

Meanwhile, Zimmerman (1971) reported a limited set of data on the Queensland Test performances on Maori and European Borstal detainees. No significant differences were found between the mean raw scores for each group.

On the other hand, Codd (1972) in a more elaborate study reported differences of 10 IQ points on both the Otis Higher and Raven Progressive Matrices favouring a European over a Maori Form III female sample. This result Codd viewed as evidence for du Chateau's (1967) 'ten-point' gap existing to the same extent on non-verbal as well as verbal measures. Codd went on to 'match' 40 pairs of subjects on Raven's IQ and socio-economic status using the Elley and Irving Scale (1972). He then compared these groups on the Otis Higher and the Progressive Achievement Tests of Reading Comprehension, Reading Vocabulary and Listening Comprehension. No difference was found in levels of performance on the Otis and Reading Vocabulary tests while significant differences occurred on the Reading Vocabulary and Listening Comprehension tests. Codd concluded that "when socio-economic status is controlled, there is little measurable difference between (Maori and European on) conventional measures of verbal and non-verbal intelligence" (p.32). It must be kept in mind however, that Codd 'matched' on Raven's results therefore his

measure of 'non-verbal intelligence' could not be treated as a dependent variable.

Codd's general conclusions were to the effect that there were many difficulties in identifying specific causal factors when studying the cognitive development of Maori and European children and that from his study it was important to note that ethnic differences in aspects of cognitive development remain after the 'control' of socio-economic factors.

Three detailed studies on measures of intelligence and their cross-cultural application have recently been undertaken by Alison St. George, Margot Klippel and Brooks. These studies have all focused on pre-school and primary school children of Maori, Pacific Island and European descent.

A. St. George (1972, 1974) reviewing comparative studies of the intellectual abilities of pre-school and early primary school Maori and European children found only two previous studies of direct relevance. The first study was that of Rogers (1964) mentioned above and the second was that of Mitchell (1968). Mitchell contrasted a group of 'culturally advantaged' with a group of 'culturally disadvantaged' five-year-olds. One comparison variable was that of intelligence as assessed by tests and teacher ratings. Differences in intelligence favoured significantly the 'advantaged' group which was two-thirds European over the 'disadvantaged' group which was three-quarters Maori.

St. George's own study focused on the Pacific Infants Performance Scale (PIPS) (Ord and Schofield, 1970) and compared the levels of performance of young Maori and European children, obtaining reliability and validity figures and deriving preliminary norms. The PIPS itself is a short individually administered non-verbal performance test of general cognitive ability.

PIPS data were collected from samples of Maori and European children in 5 six-month age groups from 4 years 6 months to 6 years 11 months. Analysis of variance was employed to test for the effects of age and ethnic group. A significant age group effect was found, as expected,

with scores increasing with age. A significant ethnic group effect was also obtained. This difference was investigated further. It was evident that Maori children did tend to perform at a slightly lower level on the PIPS in four of the five age groups. These within age group differences were tested for significance by the application of post hoc t tests. It was found that for only one age group - 5 years to 5 years 5 months - did the comparison reach the required level of significance.

In explanation it was suggested that -

"..... over the whole age range the trend is for Maori children to perform less well on the PIPS When age groups are considered separately, the variance within each ethnic group is such that there is considerable overlap between the performances of the two ethnic groups and a consistent difference fails to emerge" (St. George, 1974, p.54).

The significant difference at the 5 years to 5 years 5 months age group favouring the European sample was seen by A. St. George as possibly being a sampling artifact. She pointed out that while her younger samples were all obtained from pre-schools, the 5 years to 5 years 5 months age group may have contained children without pre-school experiences. If this were the case then proportionately fewer Maori children would have attended any pre-school facility. The proportional differences in pre-school attendance between Maori and European children though declining over the years, was evidenced in the recent survey by Barney (1975). Interestingly after six months of schooling no difference in performance was evident in the PIPS data. While it would be tempting to conclude that both pre-school and primary school experiences were responsible for the diminution of any differences in the intellectual development of Maori and European children, A. St. George pointed out that these experiences may contribute as much to favourable test-taking behaviours as to intelligence per se.

Klippel (1973, 1975) employed both conventional 'product-oriented' measures of intelligence and 'process-oriented' measures in her study into

the intellectual abilities of European, Maori and Samoan five-year-olds. The 'product' measures included the PIPS, the Stanford-Binet, Raven's Coloured Progressive Matrices, the Peabody Picture Vocabulary Test (PPVT) and a memory scale. The 'process' measures were a battery of nine Piagetian-type tests developed by Lunzer, Dolan and Wilkinson (1971). Overall, Klippel found "few significant differences between the ethnic groups in the level of performances on the fourteen tests" (Klippel, 1975, p.369). The major ethnic group differences were found on the PPVT and the 'All and Some' and 'Haptic Recognition' scales of the Piagetian-type measures. While the European children performed significantly better than Maori and Samoan children on the PPVT, the only significant comparisons on the Piagetian-type measures were between the European and Samoan groups. On some of the variables sex of subject was found to be influential with Polynesian males performing at a higher level than Polynesian females, but with a reversal to female superiority with European subjects. Various sub-tests of the measures were also analysed for ethnic group, sex and interaction effects. While some further small differences were found through these analyses it was apparent that -

"In contrast to overseas studies, there appeared to be little confirmation of different ability patterns of different ethnic groups" (Klippel, 1975, p.373).

Klippel's overall conclusion was that "The results, showed few significant differences between the ethnic groups, in their performance on intelligence tests" (p.375).

Brooks (1973, 1975) researched the performances of Maori and European four-year-old children on six non-verbal tests and three verbal tests in relation to ratings of subject's test behaviour, data on home environment variables, socio-economic status of subject's families and pre-school attendance. Two of Brooks' non-verbal measures were similar to subtests of the PIPS, but modified for use with children younger than

those in the studies of A. St. George or Klippel. These were the Knox Cubes and Design Construction tests. The other non-verbal measures were Figural Reasoning, Sorting, Series Continuation and Form Assembly Tests. The three verbal scales were Picture Vocabulary, Digit Repetition and Sentence Repetition tests.

On the non-verbal tests significant differences were found between Maori and European children on the Design Construction and Sorting Tests. Higher performances were obtained by European children on the Design Construction Test whereas Maori children were superior on the Sorting Test. Klippel, when analysing PIPS subtest performances also found European children obtained significantly higher scores on the very similar Design Making Subtest. A. St. George, analysing combined age-group data for each PIPS subtest did not, however, find any significant differences between ethnic groups on this subtest. On two of Brooks' verbal tests, Picture Vocabulary and Sentence Repetition, European children performed significantly better than Maori children.

On the basis of his results Brooks concluded that there were no major differences in Maori/European non-verbal test performances, but that such differences were to be found on verbal measures. Relating the test data to that on home environment, socio-economic status and pre-school attendance he found that higher test performances were obtained by children whose homes laid greater stress on stimulating intellectual development, whose fathers had higher education and higher earnings and by children who attended some form of pre-school facility. As might be expected, these relationships with test performance were more strongly evident with the verbal ability tests.

Comparative Studies on Intellectual Qualities

It was noted earlier that the study of the intellectual abilities of New Zealand Maori subjects has focused on both the overall level of

general intellectual ability and the organization of such ability. Early speculation on the qualities of the 'Maori mind' were noted. The review of studies in the preceding section has focused primarily on research comparing the levels of intelligence test performance of Maori and European. Exceptions would be the comments of Lovegrove (1966) on the use of 'intellectual adjuncts' on the part of Maori children in scholastic tasks and Klippel's (1973) inclusion of Piagetian-type measures in her research. There has, however, been a small but nonetheless important interest in the study of possible differences in intellectual organization between Maori and European.

It will be recalled that Ngata expressed the view that the Maori was limited by language and experience in the powers of 'abstraction and generalisation'. Ngata went on to suggest that in lacking the 'faculty of abstraction' the Maori was essentially "mytho-poetic" in thought, being incapable of scientific and more philosophic forms of reasoning. In reaction to this view Sutherland (1929) pointed out that Ngata's paper by the elaborateness of its composition served to disprove the notion that the Maori was incapable of engaging in abstract debate. Sutherland also contended that Ngata had overlooked the role of the Whare Wananga - the schools of esoteric learning in pre-contact Maori society.

Sutherland (1931) in a subsequent paper on "Maori Mentality" argued against Levy-Bruhl's (1910) view of the 'primitive mind' dominated by prelogical and mystical modes of thought applying to the Maori. Instead Sutherland put the view that the Maori could think as logically as the European but that the contents and concerns of his thought were different as a result of a markedly different cultural tradition.

Some thirty years passed until the question of Maori intellectual organization was again raised in the literature. Ritchie (1963) on the basis of Maori Rorschach records collected as part of the Rakau studies commented on Maori modes of thought. He observed that 'objectivity' was

a negatively sanctioned style of thought in the Maori tradition, being regarded as cold, unfeeling and depersonalised. He suggested that the cognitive style of the Maoris of the Rakau community showed a marked preference for definitiveness, avoiding ambiguity, employing gestalt-type rather than analytical-type thought, to value constancy in attitude and belief, and to be inner-directed and socio-centric although 'rich in affective significance'. Younger school-aged Maoris were seen as lacking in originality, as being ultra-cautious and somewhat deficient in imagination in comparison to their European peers.

The Rakau studies have been criticised on a number of grounds concerning the absence of comparative or control groups, the suspect reliability and validity of the projective measures employed and the fragmented and atypical nature of the Rakau community itself (Campbell, 1958; Metge, 1958; Burch, 1967; Watson, 1967). Critics of the Rakau studies felt that it was difficult to claim that the conclusions advanced on the organisation of intellectual abilities were intrinsically Maori. Nonetheless these studies were to provide a useful base for research in this field some years later.

During the mid-nineteen-sixties two small studies were reported that focused on the intellectual organisation of Maori subjects. Martin (1966) investigated differences in 'word fluency' between Maori and European secondary pupils. He found that Maori pupils were superior in 'word fluency' production and that within both Maori and European samples, females were superior to males. Martin argued from correlational data that 'word fluency' was an ability independent from general intelligence, but possibly related to personality variables. No qualitative distinction amongst the words produced was made in this study. Relying on a quantitative measure alone may have masked the educational implications of the data.

Bonnevie (1967) reported a very light attempt to evaluate creative thinking on the part of Maori children through school work

samples of poetry and prose. The study was unfortunately devoid of a clear methodology and adequate empirical data. No satisfactory conclusions could be drawn from the study.

The question of ethnic group differences in intellectual organisation or cognitive styles as suggested by Ritchie has however, attracted more serious attention of late.

Chapman (1973) on the basis of the reported differences in child-rearing patterns between Maori and European (Ritchie and Ritchie, 1968) investigated possible cognitive style outcomes in terms of Witkin's (1967) field dependence/independence model. No significant differences were found between Maori and European male subjects in terms of field dependent/independent orientations as assessed by the Rod and Frame Test. This was contrary to the expectation that Maori subjects would display greater field-dependence in view of the repeated observation that Maori society places much stress on group cohesiveness and group values which is thought to induce a field-dependent orientation.

Harker is currently engaged upon research into possible cognitive-style differences between Maori and European subjects. In two papers concerned with Maori educational attainment (Harker, 1971a, 1973) he reviewed a number of studies in the area noting the suggested causes for the differences between Maori and European in educational attainment. Harker pointed out that an interactive model was needed to adequately understand the relationships amongst possible causative variables. One variable stressed was the likelihood of differences in the pattern of intellectual functioning.

In support of this argument reference was made to Ritchie's (1963) observations at Rakau and to North American studies (e.g. Stodolsky and Lesser, 1967) which have suggested differences in the organisation of intellectual abilities between ethnic and cultural groups. Harker suggested that "schools may not be so much ethnocentric (as has

been claimed, see Dewes, 1968) as cognito-centric and that education-
alists should explore ways of teaching which fit various cognitive
patterns" (p.50, Harker's italics).

In order to study such cognitive patterns, Harker has recently
investigated Rosalie Cohen's (1969) 'relational versus analytical
cognitive styles' proposal. Cohen has argued that factors in lower
social-class environments, which in North America also means Black
communities and cultures, foster a relational cognitive style in
children whilst the school as an agency of an ethnically and socially
different sector of society demands analytical modes of thinking of
its pupils.

Noting the concordance between Ritchie's suggestion that Maoris
eschew analytical problem solving and Cohen's differing cognitive
styles proposal, Harker set out to research possible cognitive style
differences employing a cognitive style test adapted from Kagan, Moss
and Sigel's (1963) Conceptual Style Test.

Summarising that part of the research concerned with cognitive
style differences however, Harker (1976) notes that "The main finding
of this section of the study was a lack of any significant relationship
between cognitive style and ethnicity" (p.viii).

Comments on New Zealand Test Standardisations

Despite the considerable interest in the comparative performances
of Maori and European subjects on measures of intelligence, the opport-
unities provided by the national standardisation of a number of tests
have not been pursued in this regard. Marsden (1925) reported on the
extensive studies undertaken by the Department of Education in preparing
the Terman Intelligence Tests for use in New Zealand during the mid-1920's
(see also Department of Education, 1924). However, no mention was made of
comparative studies between Maori and European pupils. Similarly Redmond

and Davis (1940) when making the first national standardizations of the Otis Intermediate Test and an A.C.E.R. Non-verbal Test undertook no separate Maori and European analyses. This appears to be a continuing feature of national test standardizations.

The data collected on the Metropolitan Readiness Tests in the 1950's made no mention of ethnic group performances (N.Z.C.E.R., n.d.). Likewise, the norming of the Revised Tomlinson Junior School Test (Callander, Marsh and Milne, 1967), the restandardization of the Otis Intermediate Test (Elley, 1969) and the standardizations of the more scholastically oriented Progressive Achievement Test Series (see Elley and Reid, 1969, 1971; Reid and Hughes, 1974) produced no separate Maori and European analyses. The opportunities to investigate these instruments comparatively and also in terms of a number of important psychometric characteristics for relatively large and representative samples from each ethnic group appear to have been lost.

More restricted studies have also been made into other important measures of intelligence in New Zealand. Fitt (1952) summarised research by a number of students on the suitability of the Revised Stanford-Binet Scale for New Zealand children. Odell (1944) who undertook one of the studies remarked that -

"all children were New Zealand born, all were 'white' as far as could be discovered foreign elements were kept at a minimum, and the Maori element excluded" (p.3).

This exclusion of Maori children was common to all of the studies concerned with the Revised Stanford-Binet.

Rogers (1956) undertook a standardization of Thurstone's Primary Mental Abilities Test. No separate Maori and European analyses were given or even details on the inclusion or exclusion of Maori children in the sample. Similarly Shouksmith (1960, n.d.) in developing the Canterbury Reasoning Tests and norming the Differential Aptitude Tests made no

mention of Maori subjects.

Keeling and Seddon (1969) have for a number of years now been engaged in the development of a test battery for use at the Form 3 level. Progress reports (Seddon and Keeling, 1968; Keeling, 1969; Keeling, 1971) to date make no mention of ethnic group differences being investigated during the development of the test battery.

Even more recently Tuck, Hanson and Zimmerman (1975) have begun investigations into the suitability of the Wechsler Intelligence Scale for Children - Revised. Once again, no mention was made of the possibility or desirability of undertaking separate ethnic group analyses.

Conclusions, Issues and Questions

Notwithstanding the differences between the numerous studies concerned with the intellectual abilities of Maori and European, it may be possible to arrive at certain conclusions on both the results of the research itself and the pattern of the research endeavour.

To assist in the consideration of the research, Table 3.1 summarises the studies reporting sufficiently detailed information on tests employed, sample characteristics and results. Where authors have indicated the presence of a statistically significant difference in performance, this has been indicated along with the level of significance and direction. Where statistical tests were not applied but sufficient data were reported, two-tailed 't' tests were calculated using the appropriate separate variance or pooled variance method as presented in Popham (1967). In addition, a simple classification of the tests employed into 'verbal' and 'non-verbal' has been made on the basis of predominant content and method of administration.

Inspection of Table 3.1 shows that the empirical evidence does not consistently favour the proposition that Maoris perform less well on measures of ability than Europeans. The assumption that Maoris

TABLE 3.1 Summary of Empirical Studies Comparing European and Maori Performances on Measures of Intellectual Ability

Researcher(s)	Date	Test	Sample Sizes		Sample Age and/or Characteristics	Test Performances				Test type verbal or non-verbal	Significance and Direction		
			European	Maori		European		Maori					
					\bar{X}	S.D.	\bar{X}	S.D.					
McKenzie	1930	Haggerty Test of Educable Capacity - Delta 2. (Max. Score, 176). 1927 Data.	NR	NR	Standard 3	51.6	—	35.9	—	Verbal	—		
			"	"	Standard 4	68.1	—	53.2	—		—		
			"	"	Standard 5	80.5	—	70.9	—		—		
			"	"	Standard 6	97.4	—	86.0	—		—		
Downes ¹	1948	Raven Standard Progressive Matrices	1,624	214	Male Adults, Age range 21-45 years	34.70	9.20	25.70	10.40	Non-verbal	.01 E		
		Bennett Mechanical Aptitude	1,624	214	as above	23.35	9.16	18.94	7.93	Non-verbal(?)	.01 E		
		S.P. 25	1,624	214	as above	49.70	22.60	32.00	18.20	Verbal	.01 E		
Calvert, Barbara	1950	Otis Higher	30+	30+	Secondary School Students	(35.5)		(15.5)		Verbal	—		
					Age 13 years							(34.0)	(20.0)
					14 years							(40.0)	(21.0)
					15 years							(49.0)	(22.0)
		A.C.E.R. Non-Verbal	(as above)	Age 13 years	—	(28.0)	Non-verbal	—					
				14 years	(39.5)	(32.0)							
				15 years	(41.5)	(33.0)							
				16 years	(43.0)	(33.5)							
		Raven Standard Progressive Matrices	(as above)	Age 13 years	(46.0)	(37.5)	Non-verbal	—					
				14 years	(47.0)	(37.0)							
				15 years	(48.5)	(41.0)							
				16 years	(52.0)	(41.0)							
Logan ¹	1952	Otis Higher (IQ)	30	50	Teachers College Students	114.00	4.80	100.00	8.50	Verbal	.01 E		
		Raven Standard Progressive Matrices	30	50	as above	51.00	1.30	47.00	5.90	Non-verbal	.01 E		
		Mill - Hill Vocabulary	30	50	as above	55.00	5.60	50.00	6.50	Verbal	.01 E		
Baird	1952	Otis Intermediate (IQ)	18	13	Standard 3 (rural)	104.30	—	91.90	—	Verbal	—		

TABLE 3.1 (Contd.)

Researcher(s)	Date	Test	Sample Sizes		Sample Age and/or Characteristics	Test Performances				Test Type verbal or non-verbal	Significance and Direction
			European	Maori		European		Maori			
					\bar{X}	S.D.	\bar{X}	S.D.			
Walters	1953	Wechsler-Bellevue (IQ)	50	50	Male Prisoners	101.80	13.30	84.40	13.03	Verbal	.01 E
Adcock et al	1954	Wechsler-Bellevue (IQ)	32	32	Teachers College Students	116.40	—	109.10	—	Verbal	—
			—	53	Unselected Maori sample - aged 12-41 years	—	—	93.56	—		
Simpson	1955	Otis Higher (IQ)	94	96	Secondary School Students	(99)		(85)		Verbal	—
		A.C.E.R. Non-verbal test (IQ)	94	96	as above	(103)		(80)		Non-verbal	—
Smith, Leone	1957	Cattell's Culture Fair Intelligence Scale	79	79	Age range, 11-0 to 13-11 urban	29.05	5.75	27.54	5.84	Non-verbal	NS
			28	9	semi-rural	32.50	5.97	27.56	4.05		NS
			36	—	rural	21.25	7.26	—	—		—
		A.C.E.R. Intermediate C and D Tests	77	77	urban	25.30	10.20	21.23	9.83	Verbal	NS
			—	9	semi-rural	—	—	23.89	13.41		—
			36	—	rural	12.53	6.76	—	—		—
		Thurstone's Primary Mental Abilities Test	77	74	urban	94.45	32.02	98.66	30.61	Verbal	NS
			—	7	semi-rural	—	—	89.57	21.42		—
35	—	rural	81.43	35.74	—	—	—	—			
Ritchie	1957	Wechsler-Bellevue (Children's Scale) (IQ)	22	18	Secondary School Students	96.80	11.06	87.20	4.94	Verbal	.05 E
Walters	1958	Thurstone's Primary Mental Abilities Test (IQ)	116	48	School pupils, 11 to 16 yrs. urban	102.69	16.74	96.65	14.97	Verbal	NS
			—	36	Market-garden settlement	—	—	87.22	13.05		—
			—	131	rural	—	—	101.25	15.62		—
		'Non-Language Test'	126	49	urban	74.82	13.58	65.67	13.03	Non-verbal	.01 E
			—	37	Market-garden settlement	—	—	56.49	12.79		—
—	132	rural	—	—	64.45	11.78	—	—			

TABLE 3.1 (Contd.)

Researcher(s)	Date	Test	Sample Sizes		Sample Age and/or Characteristics	Test Performance				Test Type verbal or non-verbal	Significance and Direction
			European	Maori		European \bar{X}	European S.D.	Maori \bar{X}	Maori S.D.		
McClew	1958	Otis Intermediate (IQ)	27	27	Primary School Pupils	95.70	10.60	93.80	11.10	Verbal	NS
		Jenkins Non-verbal Test (IQ)	27	27	as above	98.60	11.80	100.90	10.70	Non-verbal	NS
		Thurstone's Primary Mental Abilities Test	27	27	as above	85.60	12.60	85.30	11.40	Verbal	NS
Ausubel	1961	Otis Higher (IQs)	96	96	Urban Secondary School Students European mean Age: 14 - 7 Maori mean Age: 15 - 1	94.3	---	90.3	---	Verbal	.05 E
			96	96	Rural Secondary School Students European mean Age: 15 - 0 Maori mean Age: 15 - 5	92.8	---	84.9	---		.01 E
Ashcroft	1963	Otis Higher (IQ)	---	NR	Secondary School Students	---	---	92.0	---	Verbal	
		Goodenough Drawing Technique	---	NR	as above	---	---	89.0	---		
Rogers ¹	1964	Raven Coloured Progressive Matrices Test	European & Maori	Maori	Age Group					Non-verbal	
239	30	6 year olds	12.26	5.42	10.00	2.89	.01 E				
303	35	6½ year olds	15.05	5.12	12.48	4.10	.01 E				
298	43	7 year olds	17.67	5.79	15.58	4.62	NS				
283	35	7½ year olds	19.20	5.87	14.97	6.68	.01 E				
273	28	8 year olds	21.57	5.56	20.25	3.63	NS				
313	40	8½ year olds	22.55	6.28	17.70	6.11	.01 E				
242	25	9 year olds	24.30	6.30	23.20	6.23	NS				
300	29	9½ year olds	25.60	6.37	20.90	5.39	.01 E				
238	41	10 year olds	26.98	5.60	24.56	4.65	.01 E				
269	24	10½ year olds	28.38	6.56	26.71	5.31	NS				
245	37	11 year olds	28.92	6.29	24.38	4.44	.01 E				

TABLE 3.1 (Contd.)

Researcher(s)	Date	Test	Sample Sizes		Sample Age and/or Characteristics	Test Performance				Test Type verbal or non-verbal	Significance and Direction
			European	Maori		European \bar{X}	European S.D.	Maori \bar{X}	Maori S.D.		
Lovegrove	1966	Test Battery consisting of:- i) Otis Intermediate ii) Lorge-Thorndike Verbal and Non-Verbal Tests - Level 3.	238	238	Primary School Pupils Standard 3 to Form II	122.11	43.71	107.26	39.38	Verbal	.01 E
McCreary	1966	Otis Intermediate (IQ)	—	57	Primary School Pupils Standard 3 to Form II	—	—	87.40	9.90	Verbal	
du Chateau	1967	Otis Higher (IQ)	719	236	Secondary School Students Form III	97.00	NR	84.00	NR	Verbal	—
Gustafson	1967	Jenkins Non-verbal Test (IQ)	106	42	Secondary School Students Form III Age range: 12 - 5 to 15 - 2.	113.00	NR	93.50	NR	Non-verbal	—
Archer et al	1969	'Motis' (Maori Otis) IQ	113	113	Secondary School Students Form III	67.26	—	103.53	—	Verbal	.01 M
Zimmerman	1971	Queensland Test	14	29	Male Borstal Detainees	43.21	—	42.14	—	Non-verbal	NS
Codd	1972	Otis Higher (IQ)	135	55	Female Secondary School Students Form III	106.2	—	96.1	—	Verbal	—
		Raven Standard Progressive Matrices	135	55	as above	110.3	—	98.1	—	Non-verbal	—
		Otis Higher (IQ)	40	40	as above	100.6	10.70	99.7	10.08	Verbal	NS
			(Matched pairs from above sample)								
St. George, Alison	1972	Pacific Infants Performance Scale	38	12	Pre-school and Primary School Pupils 4 - 6 to 4 - 11	17.25	6.50	14.50	5.82	Non-verbal	NS
			57	26	5 - 0 to 5 - 5	19.95	6.16	16.23	5.48		.01 E
			57	30	5 - 6 to 5 - 11	22.93	5.04	20.93	5.76		NS
			40	35	6 - 0 to 6 - 5	23.78	5.34	24.60	5.38		NS
			50	48	6 - 6 to 6 - 11	24.52	6.70	23.40	5.87		NS

TABLE 3.1 (Contd.)

Researcher(s)	Date	Test	Sample Sizes		Sample Age and/or Characteristics	Test Performance				Test Type verbal or non-verbal	Significance and Direction
			European	Maori		European \bar{X}	European S.D.	Maori \bar{X}	Maori S.D.		
Klippel, Margot	1973	Peabody Picture Vocabulary Test	20	20	Primary School Pupils 5 year olds	50.30	10.69	38.70	11.42	Verbal	.01 E
		Pacific Infants Performance Scale				14.45	4.17	17.05	5.32	Non-verbal	NS
		Memory				28.85	4.93	26.85	3.31	Verbal	NS
		Raven Coloured Progressive Matrices Test	as above			13.75	2.55	16.10	4.97	Non-verbal	NS
		Stanford-Binet				113.90	17.75	110.60	17.60	Verbal	NS
		'Piagetian' Measures: Conservation of Length				3.95	1.47	4.90	1.89	Piagetian Measures: Verbal	NS
		Conservation of Number				4.60	1.20	5.60	2.58		NS
		Seriation				3.75	4.78	2.30	2.93		NS
		Multiple Seriation				14.10	5.25	13.90	6.98		NS
		Linear & Circular Order				6.95	1.63	6.35	3.21		NS
		All and Some				4.65	1.77	3.30	2.47		NS
		Class Inclusion	as above			16.95	3.85	17.60	5.77	as above	NS
		Intersection				3.30	2.10	2.55	2.18		NS
Haptic Recognition			67.55	7.70	73.45	12.09		NS			
Brooks	1973	Knox Cubes	55	55	Pre-school subjects 4 year olds	6.76	2.48	6.44	2.20	Non-verbal	NS
		Design Construction				4.85	2.93	3.67	2.27	Non-verbal	.05 E
		Figural Reasoning				5.40	1.64	5.07	1.59	Non-verbal	NS
		Sorting	as above			7.20	3.41	8.49	3.14	Non-verbal	.05 M
		Series Continuation				24.20	5.59	22.80	5.04	Non-verbal	NS
		Form Assembly				4.89	1.90	4.11	2.31	Non-verbal	NS
		Picture Vocabulary				18.78	5.13	14.42	4.59	Verbal	NS
		Digit Repetition	as above			17.00	6.92	17.22	7.15	Verbal	NS
Sentence Repetition			14.84	6.24	12.85	5.19	Verbal	.05 E			

Notes:- 1 : Indicates that level of significance and direction calculated by present author from data in original source.

() : Indicates reported data are medians.

— : Indicates data either not obtained or not reported.

E(or)M : Indicates statistically significant difference favours European (E) or Maori (M) sample.

NS : Indicates that statistically no significant difference obtained.

always perform unfavourably on such tests would appear to have arisen from repeated and partial readings of the actual data. In addition it has often been assumed that the difference in performance would be reduced if non-verbal tests were used. It can be seen from Table 3.1 that this has not always been the case. The question of differences in the levels of performance between Maori and European on measures of intellectual ability clearly remains open as does the issue of the role of verbal versus non-verbal measures.

If the outcomes of the comparative studies are viewed in chronological sequence there is some evidence to suggest that fewer significant differences in performance are being found of late. Certainly the recent studies of A. St. George, Klippel and Brooks have not produced considerable or consistent differences in performance between young Maori and European children on a wide variety of measures of intellectual abilities. It would perhaps not be unreasonable to entertain the idea that the data reflect social, educational and acculturative processes at work upon both the European and Maori populations. In addition it could be argued that such processes are felt with greater force in Maori society given its minority and historically 'disadvantaged' position in New Zealand since European settlement. Two contemporaneous changes come to mind that could have a bearing on the data.

First, whereas the Maori subjects of today's comparative studies of intellectual abilities have parents who will have received full primary and two to three years secondary schooling, if not more, this was certainly not the case twenty to thirty years ago. The likely consequences for the children of today would be greater facility in English and a greater knowledge of and familiarity with all manner of objects and events associated with the school. Second, broader social changes involving the urbanisation of the Maori population, its economic, social and political advancement have involved Maori people in the

acquisition of new knowledge, the adoption of new behaviours and quite possibly the evolution of some attitudes and values in part congruent with the dominant European society.

The result is perhaps that both Maori and European children today share much more in common in terms of the requisite knowledge and behaviours demanded by the measures of intellectual ability utilized by contemporary researchers than those employed by their earlier counterparts.

While this argument suggests that there may now be a high degree of communality between Maori and European in terms of intellectual ability and organisation that is brought to bear on any situation or task, it does not necessarily assume that the resulting actions must or should be the same. In terms of certain orientations, attitudes and patterns of interpersonal interaction, Maoris continue to display important differences from Europeans (see Bray, 1971; Graves and Graves, 1974; Thomas, 1975). The important point is that many Maoris may prefer to be different in those ways which are said to be the essence of one's "Maoriness" (Rangihau, 1975). The resultant differences stem, however, from the conscious exercise of intellectual abilities and reason, not its absence!

Turning from the results to the overall pattern of research, a number of further points should also be noted.

First, the majority of studies have been with group measures of intelligence. Apart from the small studies with the Wechsler-Bellevue Scales in the 1950's and the more recent studies with pre-school and young primary children (i.e. A. St. George, Klippel and Brooks), comparative research has relied almost exclusively on group measures. Few investigations have explicitly set out to assess the suitability of an important ability test for use with the two major ethnic groups of New Zealand.

Second, many of the reviewed studies have been restricted in terms of age ranges and actual numbers of subjects. As a result the investigation of the technical features of particular instruments for both European and Maori subjects has been limited. The general literature on intelligence test development and use has recommended the separate investigation of test reliability, test validity and test item characteristics for minority populations for some time now (Deutsch, Fishman, Kogan, North and Whiteman, 1964; American Psychological Association (A.P.A.) 1966; 1974). Such calls appear to have passed largely unheeded in New Zealand.

Third, Ritchie's (n.d.) hope for the monitoring and encouragement of comparative research in this field in New Zealand does not appear to have been fully realised. The review of studies above may serve as an overdue step in this direction.

Together, the above considerations provided the basis for this investigation. Simply put, the issues to emerge were these -

- i) Comparative studies were not unequivocal in their findings concerning possible differences in the intellectual abilities of Maori and European. The question therefore warrants further investigation. The evidence as it stands certainly raises questions about the assumption of (a) consistently poorer performances by Maoris on measures of intellectual ability and (b) that such differences would necessarily diminish on non-verbal scales.
- ii) The empirical evidence from recent research with young European and Maori subjects has shown few significant differences on a wide range of measures of intellectual abilities. The question arises as to whether this trend holds for older subjects.
- iii) It was noted that group measures of intellectual ability had been used in the majority of studies. Since psychologists have recourse to individual measures of intellectual ability, and since few if any studies have concerned themselves with such scales at the middle-primary to secondary school level, the investigation of a suitable scale across this

level seemed desirable.

iv) It was also noted that sampling restrictions had in the past limited the generation of pertinent test data separately for European and Maori subjects. Increasing weight is being given to the need for such evidence to support the use of measures of intellectual ability. It was therefore a matter of some concern to obtain sufficient data to facilitate separate ethnic group analyses of both test and item characteristics.

CHAPTER 4

RESEARCH PROPOSALS AND METHOD

Development of Proposals:

The review of comparative research on the intellectual abilities of Europeans and Maoris undertaken in Chapter 3 concluded by noting firstly that the findings were not unequivocal on the question of possible differences. Two assumptions could be challenged when the pattern of research is viewed in toto. The assumption that Maoris consistently perform less well than Europeans on measures of intellectual ability would seem unwarranted and the assumption that differences would diminish on non-verbal scales was not always borne out.

Second, the review noted that the more recent research which had focused on young European and Maori children found few significant differences on a wide range of measures.

Taking these two points together it would appear that the question of possible European and Maori differences in intellectual abilities remains a legitimate one since the evidence, if properly read, provides no single and definitive answer. However, since it appears that little in the way of difference is evident in the recent studies with young children, the question arises as to whether or not this trend is evident with older European and Maori children. It would seem appropriate therefore, to continue to research possible differences in European and Maori intellectual abilities and to do so with children in late childhood and early adolescence.

A third point to emerge from the review was that the research had, in the main, been undertaken with group measures. This was particularly the case once subjects were at the middle primary to secondary school levels, and it could be argued that research with individual ability scales in this age range is indeed necessary.

The final point concerned limited supporting psychometric data available to justify the use of many measures of intellectual ability

with both European and Maori subjects.

It was proposed therefore that a programme of research be pursued which would first facilitate the further investigation of the comparative intellectual abilities of Maori and European subjects. Second, that the research be undertaken with subjects from the middle primary to secondary school levels so that the data would complement that recently available from studies with subjects from the pre-school and junior primary levels. Third, that this research be concerned primarily with an individual measure of intellectual ability and fourth, that the research be undertaken on a scale sufficient to provide data appropriate for the study of the major psychometric properties of the particular test.

Support for these general proposals also arose from some other related considerations. It was noted above that discussion and research into possible causes of Maori educational underachievement has given attention to the issue of the intellectual capabilities of Maori pupils. It has generally been argued that there is no difference between Maori and European in terms of general intellectual capacity although the use of such a term has been somewhat loose and the evidence for or against meagre. Still, the general question of possible differences between European and Maori pupils in terms of intellectual abilities has been of sufficient moment for many commentators and researchers to at least entertain the proposition. Most have summarily dismissed any suggestion of such differences.

It has also been argued that Maori-European differences on measures of intellectual ability and educational achievement stem not from ethnic and associated cultural differences, but from more pervasive socio-economic class factors operating throughout New Zealand society. However, when a degree of control has been placed upon socio-economic factors, Maori-European differences have still been found on important variables, including measures of intellectual ability. The evidence suggests that Maori and European subjects do not always perform equally

well on measures of intellectual ability and it cannot always be concluded that the reason for the difference lies in socio-economic factors.

It was argued above that ethnicity or ethnic grouping is a valid distinction in the New Zealand context and it would appear that the comparative investigation of intellectual ability in terms of such a distinction should still be pursued if dimensions of similarity and difference are to be understood. It may be recalled that Harker (1971) in proposing an interaction model of the 'causes' of Maori educational underachievement included a factor relating to intellectual functioning in its own right. Rather than arguing against the comparative study of Maori and European intellectual abilities Harker's model requires such research to be undertaken.

The confusion that persists in this field can be seen in arguments that have on the one hand intimated that the question had been satisfactorily resolved in favour of no ethnic group differences, while on the other hand made pleas for the development of measures of intellectual ability suited to particular ethnic groups. The Commission on Education (1962), for instance, held the view that -

"..... the evidence of psychologists (did not) suggest the likelihood of any general difference in intellectual potential between the Maori and the European" (p.403).

Nevertheless, the Commission went on to say that -

"The devising of tests that will accurately measure intelligence for different races is no easy task, and tests valid for the Maori have certainly not yet been worked out in New Zealand. The purpose of such tests, if they are to be devised, should be to have available for the Maori the kind of assistance that such tests give to schools and teachers in the classification of European pupils rather than to provide a basis for avid and doubtfully valid comparisons between the races" (p.403).

The choice of the age range of subjects in the present studies was guided by two factors. First, the knowledge that Alison St. George and Margot Klippel were already undertaking research in this area with pre-

school and young primary children. Second, a concern that if the results of the studies encouraged the use of the investigated instrument then the data should be relevant to points in the educational careers of pupils where critical decisions are likely to be made.

Although New Zealand has in many respects created an open and egalitarian educational system, streaming and selection persists in both obvious and subtle forms. Evidence of this is summarised by Harker (1975). Two points at which streaming and selection decisions are likely to be made are at the transition from primary to intermediate school and again at the transition from intermediate to secondary school. In many instances the group Otis Intermediate and Otis Higher Tests of Mental Ability (N.Z.C.E.R., 1969) are used to assist these selection decisions. The availability of an individual measure of ability suitable for subjects of different ethnic and cultural backgrounds with supporting New Zealand data would seem most useful in some situations. Such situations might arise when there were doubts concerning the appropriateness of the content of measures like the Otis Tests with particular pupils and where it was thought that the use of an individual scale may provide a more extensive appraisal of intellectual ability. Thus, research with middle primary to secondary school level Maori and European pupils with a suitable individual measure of ability did seem worthwhile.

Whether or not a new measure of intellectual ability should be developed or an existing scale adopted was a key question. The Commission on Education (1962) had not been alone in calls for the development of measures of ability that would hopefully be more suitable for Maori pupils. It will be recalled that Winiata (1944) set out to do this by adapting the Fiji Test of General Ability. Some years later Riske (1950) in joining the debate on the 'Maori I.Q.' stated that -

"Nobody knows the relationship between Maori and Pakeha intelligence We ought to know, and work ought to be done on the problem thoroughly and scientifically. Tests equally familiar to Maori and ordinary New Zealand European children are impossible to devise; but tests equally unfamiliar could be" (p.285).

It is interesting to note that Riske saw a solution in terms of 'equally unfamiliar' test materials and not tests specifically suitable to Maori subjects which appears to have been the intent of others.

Hunn, whose report on the Maori Affairs Department appeared just prior to that of the Commission on Education, concluded that "Maori education problems offer scope for research. The devising of intelligence tests suitable for Maoris is one such problem" (Hunn, 1961, p.5).

This appeal did not provoke any sharp or strong opposition in the Maori or wider community in New Zealand which was the case with other aspects of Hunn's report (see Maori Synod of the Presbyterian Church, 1961). The lack of controversy did not, however, foster research in the area. Seven years after Hunn's report Kenworthy, Martindale and Sadaraka (1968) in reviewing progress on the implementation of Hunn's recommendations were only able to note that some work had been undertaken, but no details were able to be given.

The review of studies concerned with Maori and European intellectual abilities noted research undertaken subsequent to the Hunn Report and Commission on Education. It is apparent that no project was concerned exclusively with the development of a measure of intellectual ability suitable for Maori subjects. The closest a project of this nature came to implementation would appear to lie in the proposals of Fieldhouse in 1967. He planned to investigate the intellectual development of five and six year old Maori children and develop an assessment instrument. The research was never undertaken however.²

² Personal Communication, Professor A.E. Fieldhouse, 30.5.72.

Historically then, efforts to develop instruments indigenous to New Zealand and specially suited to Maori subjects do not appear to have met with a great deal of success. It has been much more common for research to be based on already published tests of general ability. This procedure can be defended on a number of counts. First, that it is preferable to begin with instruments having the benefit of some theoretical base in the study of intelligence and the support of empirical research. Second, it is possible that economies may be made in that the development costs would be less in the study of the applicability of an existing instrument than in the development and production of an entirely new instrument. Third, some advantages may at times accrue from the use of established and existing measures since other data are usually available for comparative purposes. Fourth, it is perhaps legitimate to ask 'how useful have been the results of efforts to establish unique measures of ability?' McNemar (1964) has written directly (and amusingly) on this question in relation to the measurement of intelligence and the theoretical disputes between the 'general' factor and 'multiple' factor schools of thought. McNemar considered the differences as well as the common ground between these schools of thought. This common ground that McNemar found amongst the theoretical proposals and methods of assessment led him to enunciate the "first cardinal Principle of Psychological Progress" namely "Give new names to old things" (p.872, McNemar's italics). A warning that can be taken from McNemar's paper is that newness, novelty or difference in the formulation of measures of ability does not necessarily mean that any real theoretical or psychometric advances have been made.

As a result the decision was made to adopt an existing individual measure of ability suitable for the age range of subjects envisaged and for which there was recent evidence available on successful application in differing cultural contexts. In this way the research could commence

from an established theoretical and empirical base, economies could be made by avoiding some development costs, opportunities to consider data in relation to that already available would be present and hopefully McNemar's 'first cardinal Principle' avoided.

The choice of the Queensland Test (McElwain and Kearney, 1970), (described in Chapter 5) stemmed from two sets of interrelated factors. First, consideration had to be given to evidence from the use of tests of intelligence in cross-cultural contexts. The research and experiences of those engaged in psychological measurement across cultures has led to the evolution of a number of principles and procedures in the study of intelligence (see Chapter 1). Second, it seemed desirable for this research to be linked with that already being done elsewhere in the South West Pacific region.

The Queensland Test contains many features advantageous to cross-cultural application and it has been applied in a number of Pacific contexts. It is an individual scale and one appropriate for the age range envisaged. It was therefore adopted as the major research instrument in this study.

The Proposals

In the light of the above considerations, four research proposals were arrived at. They were -

- a) To investigate the performances of New Zealand school children aged between 8 and 15 years on the Queensland Test, attending particularly to the comparative performances of European and Maori subjects.
- b) To investigate the psychometric properties of the Queensland Test by obtaining data relevant to the reliability, validity and item characteristics of the scale for both European and Maori subjects.
- c) To develop a set of norms to aid the use of the Queensland Test in New Zealand, and
- d) To compare and contrast the New Zealand Queensland Test data with

that obtained elsewhere from both empirical and theoretical perspectives.

Subjects

A sample of 700 subjects comprising 397 European and 303 Maori school children, both male and female, aged between 8 years 6 months and 14 years 5 months was employed in this research. The subjects came from primary and intermediate schools in both urban and rural settings and were from varying socio-economic circumstances. The sample was of a modified random nature. The conduct of the research was dependent upon obtaining adequate samples of Maori and European subjects. The question of ethnic identification was, therefore, of considerable importance and is considered in some detail. This is followed by a description of the sample in terms of a number of additional variables, notes on sampling procedures and comments on its representativeness.

Ethnic Identification

Obtaining satisfactory samples of Maori and European subjects is critical to this investigation. Therefore discussion and debate on the question of ethnic identification in New Zealand is looked at in some depth.

Pool (1963) in his paper 'When is a Maori a "Maori"?' noted that identification as Maori, or conversely European, could be related to biological, social, cultural or personal factors. After examining the role of such factors in the New Zealand context, Pool was of the opinion that self-identification on the basis of cultural affinity was probably the most accurate index of social reality.

More recently Fitzgerald (1972) arrived at very much the same conclusion. He has asserted that -

"With Maoris the biological dimension ('race consciousness') is sometimes alluded to, but real evidence for a group identity based solely; or even fundamentally, on the fact of colour is hard to find. Clearly, to be Maori is

primarily a cultural classification" (p.53).

Similarly, the same argument might be advanced for cultural affinity being the most realistic criterion for being considered European.

From time to time, however, other procedures for determining ethnic group identification have been suggested. There is no denying that physiognomy and pigmentation play a part. Ephra Garrett (1973) makes this point when she points out that -

"Usually, the noticeably brown-skinned Polynesian-featured Maori is a Maori of necessity because he is ascribed this status by his fellow New Zealanders both Maori and Pakeha" (p.30).

This is a social reality that may over-ride assertions of cultural affiliation one way or the other.

Official definitions of being Maori have been much more closely linked with direct ancestry and the determination of degrees of biological relationship. Recent changes to such definitions in such fundamentally important areas as government legislation (see Maori Affairs Amendment Act, 1974) are possibly indicative of moves towards a wider acceptance of a cultural affiliation view and a recognition of the unworkability of the "biological formula" in a population whose Polynesian proportion (including New Zealand Maoris) is increasing in both numbers and biological diversity.

Between the extremes of cultural affiliation and the biological-come-blood definitions, attempts have been made to devise scales to tap ethnic identification. The Ritchies (Ritchie, J.E. 1957b, 1958, 1963 and Ritchie, Jane 1958) and Williams (1960) made the first systematic attempt to develop a measure of "Maoriness". A Maori Background Data Schedule was devised which sought to scale degrees of acculturation. The Schedule was based upon a model involving -

"..... an acculturation gradient moving from aboriginal Maori Culture at one end of the continuum to contemporary New Zealand Culture at the other" (Ritchie and Beaglehole, 1958, p.150).

The items of the scale covered degree of Maori ancestry, affiliation to a Maori religious faith, use of the Maori language, knowledge of aspects of Maori culture and level of participation in activities associated with Maori society.

A relevant but largely neglected contribution was Williams' (1960) attempt to factor analyse responses of a sample of Maori subjects to this Schedule. He found two relatively independent clusters of items in the Schedule. From an inspection of the content of the item clusters he tentatively identified the first factor as 'Maori by enculturation' and the second as 'Maori by choice'. Items in the first factor suggested action and participation in things Maori in a continuous sense, whereas items in the second factor suggested an expression and interest in things Maori via knowledge about these things. Williams tried to draw out his 'enculturation-choice' distinction by saying of a person that -

"If he does Maori things then knowledge is superfluous but if he does not act like a Maori but still wishes to retain his Maoriness then he must compensate for this inactivity by studying and knowing about the Maori" (p.41).

Further research on the Schedule or on the 'enculturation' or 'choice' elements suggested in Maori identity did not eventuate. Possibly the impetus was lost in the debate that was to surround the Schedule itself.

Joan Metge (1958) criticised the Maori Background Data Schedule in terms of both its conceptual framework and the nature of the items. Her criticism of the acculturation model involved points such as

- a) that the Schedule encouraged interpretations in variations of Maoriness as being on a developmental time sequence for each person rather than focusing on the inter-relationships amongst the people of a community,
- b) that the polar "types" were treated very much as static entities,
- c) that changes in Maori culture and society were seen as movement towards the predominant European culture,
- d) that the

continuum or gradient concept implied that European culture was higher, inviting interpretations of "progress" or "upward mobility" and finally e) that actual movement across levels of the continuum had not been clearly demonstrated.

Criticisms of items hinged on the applicability of these to the Maori population of New Zealand at large as distinct from the subjects of the Rakau studies. Failure to score highly on items concerning the use of Maori foods or consulting a tohunga would weigh against urban Maoris, but the results may just reflect "urbanness" rather than a real lack of "Maoriness".

In her critique, Metge suggested a number of areas of Maori culture that she considered worthy of inclusion in a scale of "Maoriness" and in reporting subsequent studies she listed some twelve characteristics considered to be 'characteristic and proper to' Maoris. These were -

- "(a) loyalty to other Maoris;
 - (b) attachment to the land and community of one's ancestors;
 - (c) the recognition and observance of obligations towards kinsfolk;
 - (d) generosity, sociability and co-operativeness, especially towards other Maoris and if necessary at the cost of personal economic gain (i.e. the practice of "Maori aroha");
 - (e) enjoyment of group activity;
 - (f) faithful attendance at gatherings on the marae;
 - (g) adherence to Maori ceremonial forms, especially at hui;
 - (h) interest in whakapapa and Maori history;
 - (i) the use of the Maori language, at least ceremonially;
 - (j) a deliberate happy-go-lucky attitude to time and money;
 - (k) refusal to worry over the future or plan too far ahead; and
 - (l) a taste for "Maori Kai", foods gathered from the sea, forest and wilderness and/or cooked in a hangi (earth oven"
- (Metge, 1964, p.94).

Metge cautioned that variation and difference in the practice of Maori ways is to be expected and this implies no loss of "Maoriness".

Ritchie (1968) in a paper published a decade after the Rakau studies has indicated a very similar view noting the wide spectrum of possible variation in and importance of features likely to contribute to the determination of ethnic identity. Speaking of Maoris he states that -

"There may be some who, though they declare themselves to be Maori have in fact made in all other respects a complete ideological transition to Pakeha culture incorporating basic Pakeha values to the exclusion of Maori ones. There are many who cannot speak more than a word of Maori, or show any of the more obvious signs of Maoriness and yet who continue to act in a detectable and definitely Maori manner, not because they have chosen to do so, but because they grew up with models around them who acted in this way, because, from the beginning, their family ideology was implicitly or explicitly Maori whether they recognised this or not" (p.289).

In this same paper Ritchie also lays greater stress on the dynamic as opposed to the static nature of the polar forms of each culture implied in the earlier Maori Background Data Schedule. This can be seen in his comment that "Maori culture will persist and change. It is an abiding and developing facet of New Zealand life" (p.289, my italics). In retrospect, it is perhaps sobering to note that Metge, who had criticised Ritchie for attempting to scale or identify the components of "Maoriness" had as an alternative only advanced her own list of the essential elements of "Maoriness". The step beyond the description to the measurement of "Maoriness" has proved to be something of a hurdle for the social scientists of New Zealand.

Clearly any attempt to scale "Maoriness" must take into account a number of criteria, each of which may display considerable variation. This has compounded the measurement problems. Harker (1971b) was confronted with the question of determining "Maoriness" in his study of socio-economic and cultural factors influencing Maori academic attainment. He had examined the Rakau studies and the subsequent critiques and as a result evolved a scale known as the "Group Index of Maoriness". Stemming largely from the criteria suggested by Metge it included ownership of land, degree of financial support derived from such land, parental ethnicity, language in the home as a child, facility in the Maori language, whether home community was predominantly Maori or Pakeha, attendance of family hui, ethnicity of everyday friends and self-

identification. Weightings of 1 to 5 were attributed to each criterion. Recognising the likely wide variation in responses to the items, Harker did not attempt to use his scale to compare individuals but rather to compare the groups involved in his research. He also warned against the drawing of developmental relationships between the research groups noting that the index was intended to be 'purely descriptive' of two Maori research groups.

Harker investigated which of the criteria in his index best approximated the cumulative index total from the whole scale for two Maori groups which differed markedly in the type of tertiary education being pursued. The best single indicator of "Maoriness" was found to be 'facility in the Maori language'. However, Harker points out that difficulties would be encountered in the use of items covering this area in research involving a questionnaire since the meanings of such terms as "fluent" or "competent" would be difficult to define. He argues that differences in the perceptions of respondents over fluency would introduce extreme variability. Harker then considered the next best indicator of Maoriness which was 'self-identification'. He concluded that -

"the single most valid way to establish the extent of the feeling of Maoriness would be to ask individuals what they considered themselves to be. Not only is this the most valid index, it is also the simplest to use and analyse" (p.38).

The literature and research on this issue appears then to favour the use of a criterion of self-identification when making distinctions between the ethnic and cultural groups of Maori and European in the New Zealand population. While other concepts and criteria of Maoriness are recognised, their variety appears to have prevented the development of a suitable and agreed upon scale. This being the case it has been seen by many as preferable for persons, aware of their life, concerns and circumstances to make any decision on ethnic group membership for themselves in terms of a sense of identity, belonging or affiliation.

It should be noted that Geraldine McDonald (1976) in a very recent paper has commented on the use of the categories Maori and Pakeha in educational research; She finds fault with the way in which researchers have employed differing procedures in identifying samples and points to confusions arising over the treatment of persons who may wish to identify as both 'Maori and Pakeha'. McDonald supports the use of self-identification as the most socially valid procedure in determining ethnic and cultural identification and adds the important corollary that the form, relevance and meaning of any category system must be explored from the perspective of the groups concerned.

It is apparent from this discussion that within the New Zealand context the question of ethnic identification has been concerned primarily with obtaining Maori or other ethnic minority group samples. It would appear that subjects who make a non-Maori or non-ethnic minority choice are assumed to be European. It may also be the case that the opportunity to exercise an ethnic identification choice is given less frequently to subjects who appear to be European to the researcher.

Subjects in this study were categorised as being either Maori or European solely in terms of self-identification in response to a direct question put by the researcher. All subjects were simply asked "Do you consider yourself to be Maori or European?" Their response was noted and all data thereafter treated accordingly. Very few subjects had difficulty with this question. Occasionally subjects of mixed-marriage equivocated, but on reflection were able to make a choice in terms of the community and people with whom they were most involved. That children from the age levels of this study are able to make such a choice is supported by Vaughan's (1964a, 1964b) studies on the development of ethnic awareness and Donna Awatere's (1972) replication.

Some schools in providing class lists of pupils indicated whether a pupil was Maori or European. Inquiry showed that the criteria varied. Cultural affinity, knowledge of home and family and "looks" all played a

part. In a small number of cases the school's perception of a subject's ethnic affiliation differed from his or her choice. As indicated, the subject's choice determined inclusion in either the Maori or European group. The numbers and percentage of the sample in each ethnic group are set out in Table 4.1.

Additional Variables

The sample of Maori and European subjects consisted of both males and females selected within pre-determined age bands. The subjects came from both urban and rural settings and differing socio-economic circumstances. Descriptions of the sample on these variables follow.

Gender

Both male and female subjects were included in the sample. The numbers and percentage distribution by sex of subject for the Total Sample and for each ethnic group sample is given in Table 4.2.

TABLE 4.1 Distribution of Total Sample by Ethnic Group

European		Maori	
N	%	N	%
397	56.7	303	43.3

TABLE 4.2 Distribution by Gender of the Separate and Combined Ethnic Group Samples

		European		Maori		Total Sample	
MALE	N	209		152		361	
	%		52.6		50.1		51.6
FEMALE	N	188		151		339	
	%		47.4		49.9		48.4

Age

All subjects were between the ages of 8 years 6 months and 14 years 5 months. This age range covered six age groups of 12 months. Table 4.3 reports numbers and percentages of European, Maori and Total sample subjects, male and female, in each age group.

All ages were calculated from the verified birth date held on school records to the date of testing with the Queensland Test.

Urban/Rural Location

The sample was drawn from eighteen schools. Eleven were located in urban situations with concentrated populations greater than 1000 and the remaining seven in rural areas. The numbers of European, Maori and Total sample subjects from schools designated as urban or rural were calculated and the percentage distribution of the samples are given in Table 4.4.

The figure of a 74 percent urban sample overall is a little less than the national figure of 82 percent for 1971 noted in Johnston (1974). The Maori sample was 69 percent urban which is equivalent to the 70 percent figure reported in the same paper. The use of the relatively crude classification adopted did show that the sample approximated national data on the urban/rural distribution of the population.

Socio-Economic Status

The continuing debate over the role of socio-economic factors as distinct from ethnic and cultural factors in relation to the educational achievement differences between European and Maori has been reviewed above.

It was thought desirable to have some data relating to the socio-economic status composition of the sample. This would be useful for descriptive purposes even if the issue of applying a perhaps inappropriate scale to the Maori sample might argue against other analyses.

TABLE 4.3 Distribution by Age and Gender of the
Separate and Combined Ethnic Group Samples

Age Group	European				Maori				Total Sample			
	Male		Female		Male		Female		European		Maori	
	N	%	N	%	N	%	N	%	N	%	N	%
8:6 - 9:5	19	2.7	17	2.4	19	2.7	18	2.6	36	5.1	37	5.3
9:6 - 10:5	28	4.0	29	4.1	15	2.1	18	2.6	57	8.1	33	4.7
10:6 - 11:5	73	10.4	52	7.4	33	4.7	36	5.1	125	17.8	69	9.8
11:6 - 12:5	32	4.6	31	4.4	29	4.1	30	4.3	63	9.0	59	8.4
12:6 - 13:5	38	5.4	40	5.7	34	4.9	26	3.7	78	11.1	60	8.6
13:6 - 14:5	19	2.7	19	2.7	22	3.1	23	3.3	38	5.4	45	6.4

TABLE 4.4 Percentage Distribution by Urban/Rural
Location of the Separate and Combined Ethnic Group Samples

	European %	Maori %	Total Sample
Urban	79	69	74
Rural	21	31	26

The scale adopted was that by Elley and Irving (1972), a 6-point socio-economic status scale based on 'father's occupation' with equal weightings being given to educational level required and income derived from the occupation. The scale point 1 represents the highest socio-economic status level and 6 the lowest. Data on father's occupation were obtained for all subjects from the school records and this was verified with each subject as far as possible. Table 4.5 presents the percentage distribution of the socio-economic status levels for the European and Maori samples and the Total Sample. The distribution obtained by Elley

and Irving for a national New Zealand sample is reported for comparison.

It can be seen from Table 4.5 that the European sample best approximates the distribution reported by Elley and Irving. The high

TABLE 4.5 Percentage Distribution of Socio-economic Status Levels of the Separate and Combined Ethnic Group Samples and Comparative Data from a National Sample

S.E.S. Level	European	Maori	Total Sample	NZ Sample (Elley & Irving 1972)
1	6.3	0.7	3.9	5.8
2	19.9	5.0	13.4	19.3
3	8.3	2.0	5.6	13.3
4	34.3	24.4	30.0	28.2
5	21.4	43.6	31.0	21.3
6	8.1	18.2	12.4	12.1
Not known	1.8	6.3	3.7	

proportion of the Maori sample represented in the lower socio-economic status levels on the scale (5 and 6) reflects the well documented disadvantage of the Maori population on numerous indices of social and economic well being vis a vis the population at large (see Walsh, 1973). An inquiry was made as to whether or not a separate Maori socio-economic status level distribution was, in fact, available from the data upon which the Elley and Irving scale was based. Unfortunately no ethnic group breakdown of the data had been made.³

The Total sample socio-economic status level distribution departs somewhat from the pattern of the Elley and Irving figures. This is the result of the skewed distribution obtained by the Maori sample.

In addition, the realities of the research situation meant that a 'not known' category was required. This category was employed for subjects of the research sample where the father was deceased, where the

³ Personal Communication, Dr W.B. Elley, 17. 3.76.

subjects were State Wards, where varying de facto relationships precluded inquiries into father's occupation and where some Maori subjects as mokopuna were being cared for by a grandparent (usually grandmother) or grandparents.

Sample Selection

The eighteen schools from which subjects were selected were all in the northern half of the North Island of New Zealand. All were in the Auckland and South Auckland Education Board districts. The particular schools selected contained 15 to 25 percent Maori pupils according to estimates prepared by the Education Boards. Selection of these schools was necessary in order to obtain suitable numbers of Maori subjects within the constraints of time and resources. There were a number of schools in the two Education Board districts with considerably higher proportions of Maori pupils but these were not approached because of a) their atypicality and b) their saturation by other researchers.

In this research the sample of 700 subjects was obtained in two ways. A total of 234 subjects came from nine class groups from the Standard 3 to Form II class levels. It was planned that sets of test data, class performance data and teacher assessment data would be obtained in addition to Queensland Test performances for reliability and validity studies. The Queensland Test was administered to five class groups selected from two schools in March and April 1970. The classes were followed for six months to collect the additional data. This attempt to collect validation data failed. Firstly, class performance records were not maintained consistently by all teachers. Secondly, three class groups experienced class re-organisations after three months, leaving samples of 14 subjects per class with their original teachers. In addition, some test data for these classes was not available in raw score form. Consequently the Queensland Test data from these five class groups

(N = 107) were retained for the age group analyses only.

In March 1971 and under improved conditions, four further class groups (N = 127) were selected from two schools for studies into aspects of Queensland Test reliability and validity. For the validity studies test data, class performance and teacher assessments were obtained and the class groups were retested after a six-month interval with the Queensland Test in order to estimate test-retest reliability. The Queensland Test data for the 127 subjects collected in March were also incorporated into the age group analyses. Details on the four class groups are given in Table 4.6.

TABLE 4.6 Details on Numbers, Ethnic Composition, Class Level and Age of Class Groups Employed in the Queensland Test Reliability and Validity Studies

Class Group	Total N	European N	Maori N	Class Level	Average Age
1	27	17	10	Std. 3	9y 9m
2	35	27	8	Std. 4	10y 6m
3	34	18	16	Form I	11y 7m
4	31	18	13	Form II	12y 5m

The remaining 466 subjects were obtained by a stratified-random sampling procedure. The ages of potential subjects were obtained and accordingly allocated to the appropriate age group. Random numbers were then used to select subjects from the subject pool. Sufficient subjects were selected for each ethnic group sample at each age level for all samples to exceed 30.

From Table 4.3 it can be seen that the European age groups 9:6 to 10:5 and 10:6 to 11:5 contain almost twice as many subjects as the corresponding Maori age groups. This resulted from the preponderance of European subjects in the class group samples. Also, of the nine class group samples tested two were at the Standard 3 level and five at the Standard 4 level, thus increasing the numbers of subjects aged between

nine and a half and eleven and a half. The allocation of class group subjects to age groups did result in an uneven distribution of subjects. However, it was thought desirable to bring as much data as possible to bear on the major aspects of this research, therefore data were not discarded in order to equalise subject numbers.

A number of ethnic group or age group shifts also occurred as a result of subjects a) making an ethnic group self-identification contrary to that initially indicated by the school and b) falling into an older group by the date of testing. These factors also contributed to the unequal distribution of subjects in the ethnic and age group samples.

The collection of data from subjects in the stratified-random sample took place from October to December of 1971 and February and March of 1972.

Representativeness

Since it is proposed that a set of norms for the Queensland Test be developed, it is appropriate to comment on the representativeness or otherwise of the sample. The degree to which the sample(s) can be considered representative of New Zealand subjects of the same age range has a bearing upon the usefulness of any norms.

The proportions of Europeans and Maori subjects in the Total Sample do not correspond to the proportions in the New Zealand population at large aged between 8 and 14 years. Official figures based on the 1971 Census indicate that some 88 percent of the New Zealand population in this age range were non-Maori and 12 percent Maori (Department of Statistics, 1974). It is therefore apparent that the 56.7% European and 43.3% Maori sample employed in this research is at variance with the national data. This under-representation of European subjects and over-representation of Maori subjects is the result of attempting to obtain sufficiently large ethnic group samples to enable separate analyses of

various psychometric properties of the Queensland Test.

Males and females were equally well represented in the sample as a whole and in the separate European and Maori samples. There is a slightly higher percentage of male subjects which is also the case in national data (N.Z.O.Y.B., 1971).

The distribution of subjects over the six age groups is uneven. The samples covering the ages 10:6 to 13:5 are considerably larger than at the remaining age levels. Krejcie and Morgan (1970) have presented a table for determining sample sizes necessary in order to be representative of a given population. Data on the estimated age distribution of the New Zealand total population, New Zealand Maori population and New Zealand population excluding Maoris, based on the 1971 Census (Department of Statistics, 1974) was employed in determining whether or not the samples could be considered representative.

In terms of the population size at each age level from 8 to 14 years the numbers of subjects included in the age level samples cannot be considered representative when assessed against the sample size determination table of Krejcie and Morgan. This is so for both separate and combined European and Maori samples at each of six age levels.

If the samples are combined across age, the table of Krejcie and Morgan indicates that a sample of 397 European subjects would be representative of the estimated population (excluding Maoris) between the ages of 8 to 14 years. The sample of 303 Maoris is, however, smaller than the required sample indicated by Krejcie and Morgan's table to be representative of the estimated Maori population in this age range.

Since it is envisaged that the norms would be derived in relation to the six age levels, it must be pointed out that on this basis the samples at each age level are not representative and consequently any norms must be considered as tentative.

Data on the urban or rural location of the European and Maori

samples and on their socio-economic status distribution suggest that the samples do not differ markedly from what is known about the wider population.

The geographical distribution of the samples is, of course, not representative. All subjects came from the northern half of the North Island.

The above comments suggest, then, that the samples of European and Maori subjects do conform acceptably with the known population parameters on gender, urban-rural location and socio-economic status distributions. However, the age group samples are of insufficient size to be considered representative and geographical representation of subjects in the samples is limited. These latter restrictions on the sampling should not be overlooked in the use of the norms that were derived for the Queensland Test from this New Zealand data.

Procedures

The implementation of the proposals involved the administration of the Queensland Test to 700 European and Maori subjects. A sub-sample of 127 subjects were involved in the reliability and validity studies. Details on the selection of subjects and the composition of the samples were given above. The Queensland Test, its development and application in other research are described in Chapter 5. It remains to describe the procedures and conditions relating to the collection of the Queensland Test data and supplementary data collected in relation to the validity studies.

The Queensland Test was administered to all subjects by one of four trained examiners - two female and two male (one being the author). All were European and graduates in psychology. All testing of subjects in the test-retest reliability studies was undertaken by the author.

The Queensland Test was administered in accordance with the

procedures and conditions laid down in the Handbook (McElwain and Kearney, 1970). The one departure of significance relates to the use of verbal instructions. The Queensland Test has been described as an instrument for use under conditions of 'reduced communication' and can be administered by mime or imitation.

It was found in practice administrations of the Queensland Test that subjects expected that some verbal exchanges would take place and the situation of a total embargo on verbal communication was somewhat contrived and disturbing. The situation in New Zealand is that for the great majority of Maori children English, or a Maori-English dialect (see Hawkins, 1972), is their primary language. English is, of course, the primary language of most New Zealand European children. The impediments to examiner-subject verbal communication did not exist in the forms experienced in other studies with the Queensland Test.

It was decided therefore, to introduce the Queensland Test with the short and simple verbal instruction, 'Watch what I do and you do the same'. This accompanied the introductory practice items to each subtest. This instruction was found to be sufficient to direct the subject to the task and its requirements. No major difficulties were encountered in the administration of the Queensland Test in this manner. Occasional reinforcers such as 'good' or 'fine' were permitted. It should be noted that McElwain and Kearney (1970) do allow for some limited verbal communication. They point out that most languages and cultural groups contain words or gestures of approval and state that "it is sometimes helpful to use some supportive words or gestures to encourage the subject" (p.35). Similar limited verbal communications have also been found helpful when using the Queensland Test in the Cook Islands (St. George, R. and St. George, A., 1974) and in Zambia (Serpell, 1974).

All testing took place during normal school hours. Testing conditions in the schools varied somewhat depending on the resources of

the school. Spare classrooms, the school hall, the sick-bay, the library and even the cleaner's room were used. While not totally unsatisfactory, conditions for research of this nature involving data collection on an individual basis could not be described as ideal in New Zealand schools. Seating, illumination and ventilation were always adequate, but ambient noise was a problem in a few instances.

Supplementary Data

In an attempt to obtain evidence on the validity of the Queensland Test in the New Zealand school context supplementary data were obtained from a number of class groups. These data took the form of results from group measures of ability, group measures of language skills, class performance data in subjects of instruction and teacher rankings. Details on the supplementary data are given below.

(a) The Pacific Reasoning Series Test - Form A (Ord, 1968a).

This is a paper and pencil group measure of general ability comprising 60 items of the reasoning series continuation type. The items require the subject to complete a figure, number, word or letter series. This test was administered with a 30 minute time limit. The test does not require high level English skills and has been used extensively in Papua New Guinea with subjects having minimal skill in the English language (see Preston, St. George and St. George, 1974a). This test was administered by the author according to the instructions in the Manual to the class groups in a morning 'test session'.

(b) The Otis Test of Mental Ability - Intermediate Examination - Form A (N.Z.C.E.R., 1969).

This is a paper and pencil group measure of general ability consisting of 75 items of a predominantly verbal nature. Items include synonyms, antonyms, general knowledge (e.g. meaning of proverbs), verbal and numerical reasoning. This test takes 30 minutes to administer. The Otis tests have been and are still widely used throughout New Zealand schools for classificatory purposes. Again this test was administered by the author according to the instructions given in the manual and in the morning 'test session'.

- (c) The Progressive Achievement Tests: Reading Vocabulary and Reading Comprehension. (Elley and Reid, 1969).

Both are paper and pencil multi-level measures. The Reading Vocabulary Test is designed to measure the understanding of common words and the Reading Comprehension Test the difficulty level of material a subject can read with adequate comprehension. The Reading Vocabulary Test requires 30 minutes to administer and the Reading Comprehension Test 40 minutes. Since these tests are multi-level scales, subjects at different class levels do not all do the same items or same number of items. Thus, the data reported from these measures for each class group are not always from exactly the same set of items. These tests are used extensively in New Zealand primary and intermediate schools with the bulk of this testing being undertaken by school personnel during February and March of each year. The data reported in this study were obtained from the test forms held at the schools from the administration of the tests to the class groups at the beginning of the school year.

- (d) Class Performance Data on Language Skills, Mathematics and Nature and Social Studies.

The teachers of the class groups all maintained records of evaluations

of pupil performance in the 'Register of Progress and Achievement' (Department of Education, n.d.). Evaluations took the form of both ratings on 5-point scales and percentage results from class tests or assignments. Only the latter data were employed in this study. Language Skills data were obtained by averaging with equal weights the results available from the assessments of oral language, written language, reading, spelling and writing. The Mathematics data were the averaged results from arithmetic and 'new mathematics' assessments with equal weights. The separate evaluations of Nature Study and Social Studies were combined, again with equal weights, to form a single Nature and Social Studies assessment. All these data had a base of 100.

(e) Teacher Ranking of 'General Ability'

The teachers of each class group were required to rank-order their classes in terms of perceived 'General Ability'. Tied ranks were permitted but not encouraged.

The data for the class group were collected in the following sequence -

- i) The Queensland Test was administered to each class group member.
- ii) When all administrations of the Queensland Test were completed for the class group a morning 'test session' was organised with the Pacific Reasoning Series Test and then the Otis-Intermediate being given.
- iii) Reading Vocabulary and Reading Comprehension Test data were obtained from the school's test forms at this time.
- iv) Accumulated class performance data and the teacher rankings were obtained for Class Groups 1 to 4 during July and August 1971 prior to re-testing with the Queensland Test. At re-testing the class groups had been together six months.

All analyses undertaken in relation to both the class group data and all other Queensland Test data were performed on the data in raw score form.

The results of the research as they relate to the investigation of the comparative performances of European and Maori on the Queensland Test, its psychometric properties and the development of New Zealand norms are reported in Chapter 6.

CHAPTER 5

THE QUEENSLAND TEST

Development

The origins of the Queensland Test extend back over some twenty years to the work of McElwain and Griffiths (1957) in Papua New Guinea, to the subsequent studies by Ord (1959, 1968b, 1970) in the same country through to the studies by Kearney (1962) in Australia.

In the mid-1950's McElwain and Griffiths undertook studies into the feasibility of employing psychological tests in the selection of Melanesian recruits into the Pacific Islands Regiment (P.I.R.) of the Territory of Papua New Guinea. They were able to isolate a number of item types which showed some success in separating 'learners' from 'non-learners' in the P.I.R. cadet training context and recommended further research.

The report by McElwain and Griffiths arrived at four criteria for the inclusion of item types. These were that: (a) the items were known to be a good test of general intellectual capacity for European populations, (b) the items types were to be largely free of language skills, reading, writing, listening or speaking requirements, (c) the items should not require bulky equipment, and (d) the items should not involve fine manipulative skills.

On the basis of their research McElwain and Griffiths drew up a provisional list of thirteen potentially suitable item types. These were:

- "(a) Plan of search (SB.37.L.XIII.1)⁴
- (b) Pictorial likenesses and differences (SB.37.L.VI.5)
- (c) Mutilated pictures (SB.37.L.VI.3)
- (d) Picture absurdities (SB.37.L.VII.1) (first three only)
- (e) Copying a diamond (SB.37.L.VII.3)

⁴ SB.37.L.XIII.1 stands for Stanford-Binet Test, 1937 Revision, Form L, Age 13, item 1.

- (f) Paper cutting (SB.37.L.XIII.3)
- (g) Memory for designs (SB.37.L.IX.3)
- (h) Goodenough's Draw-a-Man test
- (i) Alexander's Passalong test
- (j) Porteus Maze test
- (k) Item type 2 from the A.C.E.R. Junior Non-Verbal test
- (l) Army Test GB (Test G from the Stephenson GVK series used by the British Army)
- (m) The Enticknap extension of the Binet picture vocabulary test" (from Ord, 1970, p.15).

McElwain and Griffiths concluded that "a battery of tests could now be made up that should be fairly useful for P.I.R. recruiting and that it should take the form of an individual performance battery" (from Ord, 1970, p.15). The unenviable task of developing this test battery in the isolation and difficulties presented by the Papua New Guinea context fell to Ord.

For a decade from the late 1950's Ord worked towards the development and refinement of such an individual performance battery, guiding it through a number of revisions as the P.I.R. Test and into its final form as the New Guinea Performance Scales.

Ord's studies commenced with research into the applicability of some item types suggested by McElwain and Griffiths as well as into item types he had on his own initiative already begun to study in Papua New Guinea. The first and provisional test battery was made up of the following twelve subtests:

1. The Cube Imitation Test
2. The Passalong Test
3. Pictorial Likeness and Differences
4. Picture Absurdities
5. Copying a Diamond
6. Plan of Search
7. Memory for Design
8. Copying a Bead Chain from Memory
9. Draw-a-Man Test
10. Paper Cutting
11. Army Test GB
12. A.C.E.R. Non-verbal Test - Part 2

Ord administered the above tests to Papua New Guinea subjects in the course of Police and Army recruitment programmes and the data, along with the experience of administering particular tests, led to refinements in the test battery, the deletion of some tests and the

introduction of new and novel tests.

The first major revision of the tests Ord had been working with became known as the P.I.R. Test (Form I) and consisted of the following subtests; Cube Imitation, Bead Threading from Memory, Passalong, Form Assembly (Space Relations) Test and Observation Test. While each of these subtests are more fully described by Ord (1970) in his detailed account of this work in Papua New Guinea a short description is appropriate at this point.

The Cube Imitation Test was an adaptation of Cattell's (1953) revision of the Knox Cube Imitation Test (1914). Subjects are required to copy a sequence of block taps presented by the examiner. The block taps are made with a detached block upon a standard frame upon which four blocks are mounted at equal intervals. A test of the same type is to be found in the Arthur Point Performance Scale (1943).

The Bead Threading from Memory was a development of items from the Stanford-Binet Tests of Intelligence (Terman and Merrill, 1937). With this item type subjects are required to reproduce a previously seen threaded sequence of beads on a string with loose beads.

The Passalong Test was originally developed by Alexander (1932) and is a type of sliding block puzzle in many ways similar to some children's games. Here the subject must rearrange a pattern of tiles in a tray to a terminal pattern presented by the examiner. In endeavouring to achieve the terminal pattern the subject is only permitted to slide the tiles in the tray. This type of test has been studied by Semeonoff and Trist (1958).

The Form Assembly Test was developed by Ord from an A.C.E.R. space relations type test. In tests of this nature the subject is commonly presented with a square or a diagram of a square with a part or parts missing. The subject must select from diagrams or an array of actual pieces the piece or pieces which would complete the square. Ord employed a 'concrete' version in his Form Assembly Test.

The Observation Test was original to Ord and inspired by a game described in Rudyard Kipling's (1908) novel Kim. Observing that Papua New Guinean subjects evidenced "continual cautious alertness" (Ord, 1970, p.23) while being tested Ord developed this test where subjects observed and memorised objects varying in material, size, shape and colour and then identified these from newly added objects.

Along with the development of the P.I.R. Test, Ord was also working on the modification of Koh's (1923) Cube Construction Test. This type of test which was more commonly known as Koh's Block Design had been incorporated into the Wechsler Scales and had been used with indigenous subjects in Africa (Biesheuvel, 1952; Oliver, 1933) and elsewhere (e.g. Bhatia (1955) in Northern India, Warburton (1951) amongst the Gurkhas).

In the initial studies with this item type Ord found Papua New Guinean subjects to have considerable difficulty in constructing the required pattern with the blocks. In Ord's words -

"The major problem seemed to be one of reproducing patterns illustrated in two dimensions by using three dimensional material. Subjects repeatedly tried to make patterns using the tops and sides of blocks at the same time, and it proved too difficult to communicate the need to use one side only. It was therefore decided that the third dimension, in fact, was unnecessary for the correct completion of the task and could be removed. This was done simply by using relatively flat plastic tiles" (Ord, 1970, p.64).

In this way Ord was able to advance the development of this item type to the stage of producing a separate test consisting entirely of such items - namely the Pacific Design Construction Test (Ord, 1968c). Interestingly Morin (1962) had made the same modification to the Koh's Block Design test when developing his 'G52' test for use in North Africa. Neither the work of Morin nor Ord appears however, to have influenced the publishers of the Wechsler Scales towards this modification even though Biesheuvel (1969) in editing an International

Biological Programme Handbook on the measurement of psychological performance recommends the use of Ord's modified scale.

The 1960's were a period of considerable research and experimentation by Ord and his associates in Papua New Guinea with the P.I.R. Test, the Pacific Design Construction Test and a number of related scales. New items and item types were being studied, methods of test item presentation investigated and slowly a body of evidence on test reliability and validity established. The many and varied aspects of this work are well documented in Ord's (1970) book Mental Tests for Pre-literates.

The early 1960's brought Professor D.W. McElwain and his staff at Queensland University back into close association with the test development studies of Ord and set in motion events that led to the development of the Queensland Test. Kearney (1962) had already undertaken limited studies with the P.I.R.-IV Test with Australian Aboriginal and Papua New Guinean subjects and had suggested further item-type research with Aboriginal subjects. Kearney's studies with the P.I.R.-IV Test also highlighted the need for certain modifications and improvements.

A problem of some importance with the P.I.R. Test was the considerable bulk and weight of the Form Assembly subtest since each item consisted of a single form board. McElwain and his staff therefore set about minimising the material required for this subtest without radically altering its content and purpose. Their solution consisted of a single plastic form board on which was mounted a raised grey 4 inch by 4 inch square and seven fixed black shapes mounted on grey bases. The items are presented by placing a black template on the grey square. These templates correspond to the dimensions of the grey square except that a part is missing. The black square can be completed by the addition of one, two or three of the seven black shapes.

Since these are fixed the subject must manipulate the shapes mentally in attempting to complete the black square. Ord compared the results obtained from administering the original 'Multiple Form' and the new 'Single Form' method of the Form Assembly subtest when administered to Papua New Guinea subjects finding very small differences between the mean scores (see Ord, 1970, p.72). As a result he recommended the use of either form of the Form Assembly subtest.

In 1962 Ord's P.I.R. Test and Pacific Design Construction Test formed the basis for a series of independent test development studies undertaken by Kearney (1966b) at the University of Queensland into the construction of a general ability measure for use with Aboriginal Australians. While Kearney worked on item types and procedures that were to lead to the Queensland Test in its final form Ord continued the development of the P.I.R. Test in Papua New Guinea (see Ord, 1967a, 1970, 1972). With the alternative forms of the Form Assembly subtest and the addition of the Pacific Design Construction Test in its entirety as a further subtest the P.I.R. Test became the New Guinea Performance Scales (Ord, 1967b). Changing circumstances were however to bring the work on the New Guinea Performance Scales in Papua New Guinea to a close but the ideas, methods and materials were to flourish elsewhere.

Kearney's (1966b) studies of the item types found in the P.I.R. Test and the Pacific Design Construction Test with Papuan, Aboriginal Australian and European Australian subjects led to the adoption of some item types, the modification of others and the exclusion of yet others in the development of the new test. Full details on the development of this new test are to be found in Kearney's (1966b) doctoral dissertation. It is important however to note the nature of and reasons for the changes Kearney instituted.

In the course of developing the Queensland Test Kearney investigated eight tests:

- (1) Knox Cube Imitation
- (2) Beads
- (3) Alexander Passalong
- (4) Form Assembly
- (5) Pattern Matching (Design Construction)
- (6) Revised Continuation Test
- (7) Revised Observation Test
- (8) Revised Lucifer Test

These tests other than the Revised Continuation Test and the Revised Lucifer Test were described above when discussing the work of Ord.

The Revised Continuation Test was developed from the work of Hector (1960) in Africa. Hector's paper and pencil test required subjects to continue a pattern, part of which had been completed by drawing connecting lines between two rows of dots spaced at equal intervals.

The Lucifer Test was the unpublished work of Van Den Hout, a psychologist working in Dutch New Guinea. The test consisted of series continuation type items constructed from matches in order to 'concretize' the task. Subjects were shown a card with matches fixed in four cells in a series continuation arrangement. In a fifth cell the subject had to place a loose match or matches in the appropriate position to continue the series.

Kearney revised both Hector's Continuation Test and the Lucifer Test. However, the paper and pencil nature of the first test and problems in establishing sufficient items of varying difficulty with the same underlying logic with the second led to their abandonment.

Kearney also revised the Observation Test but was forced to discontinue work on it because of difficulties in its administration and manufacture. In his revision Kearney had replaced some of the objects displayed in the Observation Test so that all were non-representational and had set about standardizing the placement of the objects used in the test. This was achieved by fixing the objects to both the 'observation' and 'test' boards. Thus, the subject was first presented with an 'observation' board so that the attached objects

could be viewed for a set time. This was then removed and replaced by a 'test' board on which both the original objects and some new objects were attached. The subjects task was to identify the latter. These worthwhile changes did not however overcome difficulties in communicating the requirements of the task to pre-literate subjects.

The five remaining item types were to eventually form the basis of the Queensland Test after a programme of extensive research.

The Knox Cube Imitation Test had been used in a number of ability scales prior to the Queensland Test. Kearney (1966b) has documented the forms and revisions of this test. He noted that apart from one revision incorporated into the Arthur Point Performance Scale (1943) the initial tap was always on the first block. For subjects who perceived this initial starting point the difficulty of each item was reduced by one tap. Kearney's revision of this test randomised the starting point of the block tapping after the first item. Conventions concerning tapping a block once and once only in the four block sequences and the avoidance of tapping a block twice in succession in longer sequences were also applied. In contrast with other versions of this test Kearney introduced a second practice item of five block taps on the completion of all the four-tap items. This was done to break any response set to tap only four blocks and to facilitate the transition to longer and more difficult items.

It has already been noted that Bead Threading items of the Stanford-Binet Tests of Intelligence had been used in the Papua New Guinea studies. Kearney was to retain this item type but made substantial alterations to the items and method of administration that evidenced great ingenuity on his part. In working with the Bead Threading subtest in the P.I.R. Test Kearney concluded that three factors were influencing performance.

First, dexterity was likely to influence test performance. A clumsy subject or one who had difficulty in threading beads would

lengthen the time required to complete the task and experience the item as more difficult since delayed recall would be more necessary.

Second, bead threading is a sequential task. If an error is made the subject may recognise it immediately, later, or not at all. Kearney noted that if a subject recognised an error immediately he/she simply had to alter the last bead on the string. But, if the error was recognised "later in the pattern all the beads following the incorrect one would need to be removed and replaced after a correction had been made" (Kearney, 1966b, pp.102-103). The outcome would increase the delay in recall time resulting in poorer memorisation of the remaining unthreaded beads.

Third, Kearney in undertaking preliminary studies with Papuan subjects observed that the symmetry of many bead patterns was detected and utilised but that Aboriginal Australian subjects failed to employ this strategy. For subjects who perceived the symmetry the difficulty of the task was reduced since only half the pattern and the centre bead had to be remembered. Arguing that the ability to detect symmetry appeared to be culturally induced and not necessarily a function of general cognitive ability, Kearney took steps to remove symmetry from the Beads subtest items. A further compelling reason for the removal of symmetry was that if retained, bead sequences could only be increased by the addition of an even number of beads at a time.

Although the problem of symmetry could be overcome while retaining a bead threading task the problems of dexterity and ease of error corrections could not. The solution lay in the substantial modification of the method of administration. The threading of the beads was discarded altogether. In its place a frame with two parallel grooves was devised. In this the bead sequences were displayed by the examiner and then constructed by the subject from loose beads provided. A full description of the materials and procedures of this new Beads subtest is given below.

The Passalong Test was included in substantially the same form as it had appeared in the P.I.R. Test, Ord in his use of this test had enlarged the stimulus cards so that they matched the size of the Passalong item trays and so that the tiles in the stimulus pictures corresponded in size to those used by the subject. These modifications were retained.

The Passalong Test in the P.I.R. Test had, however, been constructed from plywood. This had led to inaccuracies in the tile shape which influenced their ease in sliding. In addition the plywood frames were found to warp in tropical conditions causing further difficulties in moving the tiles. Both problems were overcome by constructing all test pieces from solid plastic in the Queensland Test version of the Passalong Test.

The Form Assembly Test incorporated into the Queensland Test was of the single form board type which had been developed by McElwain and associates to overcome the bulk and weight problem encountered with Ord's original multiple form board method.

The last item type to become part of the Queensland Test was based on the Kohs Blocks Design Test as modified to tiles by Ord (1968c) in the Pacific Design Construction Test. The only change to the form of this test by Kearney was to adopt a uniform border for the frames in which the designs are constructed by the subject. The frames now corresponded exactly to the stimulus picture dimensions. Previously Ord had varying width borders on the frames depending on the space required for the design construction. The title 'Pattern Matching' was given to this subtest in favour of block design or design construction.

In the course of the revisions to the five item types and their reformulation into the subtests of the Queensland Test Kearney instituted changes to the item orders and introduced some new items on the basis of his studies. These are all documented in detail in his dissertation.

Kearney also investigated the effects of various scoring procedures and testing discontinuation rules. The evidence favoured the adoption of a scoring procedure based on a simple binary system with all items being scored 1 if completed successfully, 0 if wrong. Partial score credits for partial solutions or speed of performance were studied but not finally introduced. In three subtests a testing discontinuation rule of three successive failures was adopted and two subtests employ rules based on the completion of sub-sets of items. The details are given below, in the description of the Queensland Test.

In discussing the development of the Queensland Test Kearney lays great stress on the lengths he went to so as to reduce ambiguity in the items and confusion on the part of the subject. The attention to detail and exactness in the construction and production of the Queensland Test testifies to this effort. The reason for the attention to detail was that information irrelevant to the task (solving a particular item) only causes difficulty to subjects.

In Kearney's (1966b) words -

"The irrelevant aspects of the task that the subject defines for himself as being relevant make the item of a different level of difficulty. He includes an extra component in the process of finding a solution. This often escapes the notice of the test constructor and is seen as a reflection of subjects' level of ability rather than as an increased level of difficulty of the task" (pp.120-122).

Thus, precision has been accorded a high priority in the design and manufacture of the Queensland Test so that all irrelevant aspects might be eliminated. The Queensland Test (McElwain and Kearney, 1970) in its final and published form will now be described.

Subtest Descriptions

The Queensland Test (henceforth QT) is an individually administered performance test of general cognitive ability. It consists of five subtests administered in a non-cyclic omnibus form in the following order -

- (a) Knox Cube Imitation Subtest
- (b) Beads Subtest
- (c) Passalong Subtest
- (d) Form Assembly Subtest
- (e) Pattern Matching Subtest

The QT consists of 60 items scored on a right-wrong basis. It is essentially unspeeeded and takes some 45 to 65 minutes to administer. Administration does not require the use of verbal instructions on the part of the examiner nor verbal responses on the part of the subject. McElwain and Kearney (1970) suggest that the test "could be described as one having 'a wide communication channel'. It is not much hindered by language or verbal barriers" (p.11). Mime plays a critical role in the administration, with the examiner 'inviting' the subject to imitate a demonstrated manipulation of the test materials to achieve an overt item goal. McElwain and Kearney (1970) comment that "The induction of 'set to imitate' is usually easily effected" (p.11).

With the exception of the Knox Cube Imitation and Beads subtests the item goal is physically present to the subject during his/her attempt at each item. In the case of the Knox Cube Imitation the subject must repeat exactly a set of block taps demonstrated by the examiner and in the Beads subtest the bead sequence is covered after a fixed exposure time and the subject must construct the sequence from memory with beads provided.

The materials employed in the QT are all non-representational. No sketches or photographs of objects are used and the materials employed do not have a common use or meaning. The materials have been constructed to a high degree of accuracy and are unambiguous in that cues extraneous to the task at hand or likely to confuse have been eliminated as far as possible.

The items in each subtest are arranged in increasing order of difficulty on the basis of the Australian studies and as Kearney (1966b) notes -

"The subject must rely on the experience received in attempting the solution of the preceding items to be able to solve the present item. This involves the subject's learning the methods of solution as they are evolved item by item" (pp.86-87).

Practice in an item type is given prior to the administration of the subtest proper. Practice items have also been included in various subtests when a new requirement is to be introduced to the subject. The practice items are repeated until the subject solves them correctly. This procedure was adopted on the grounds that evidence that the subject comprehends the nature of the task is essential prior to the administration of the actual test items.

Kearney (1966b) has argued that if a subject fails to solve an item correctly it is important that the solution to the item be demonstrated before proceeding to subsequent items. He gives two reasons. First that the subject be given the opportunity to learn the method of solution and second to demonstrate that the item can be solved by normal methods. With the Beads, Passalong and Pattern Matching subtests a subject can readily be directed to any error in his/her solution to items. Here the subject or examiner will correct any such error after the subject has attempted a solution.

While the QT is essentially unspeeeded the Passalong and Pattern Matching subtest items must be solved within fixed time limits. The time limits are very generous and were imposed so that the total test time was kept within practical limits. In the Beads subtest the bead sequences are exposed to the subject for a fixed time but there is no time limit on the solution of the items.

The QT is a performance scale in a quite strict sense being devoid of any obvious formal educational content, however it is recognised by McElwain and Kearney (1970) that the test is "relatively although not completely free from the intrusive effects of culturally determined skills which facilitate performance" (p.11).

The development of the QT has involved the collection of data

relating to reliability, validity and the generation of decile norms for groups of Australian children. Rather than attempting at this point to summarise the data which are available from the Queensland Test Handbook, reference will be made to such data as and when necessary in relation to the currently reported research. It is appropriate to note however that the reported reliability, validity and item difficulty were sufficiently encouraging to the author to support the adoption of the QT as the major instrument in this research project.

The technical details of each QT subtest are now summarized. This material is based on the descriptions by Kearney (1966b) and McElwain and Kearney (1970).

(a) Knox Cube Imitation Subtest

This subtest consists of four black $\frac{1}{8}$ inch cubes attached at equal intervals along an 8 inch white base frame. There are two detached black $\frac{1}{8}$ inch cubes for tapping. The test materials are shown in Figure 5.1.

The examiner gives the subject one of the detached cubes and with the other taps out a set item sequence on the frame. The examiner then invites the subject to tap the same sequence. There are 15 scored items and 2 practice items in the subtest. The subtest task is introduced with the first practice item which the subject must perform correctly prior to commencing the subtest. The second practice item occurs when five or more cube taps are required as opposed to the four taps in the previous items. The number of taps increases from four to seven as the items become more difficult. Table 5.1 presents the Knox Cube Imitation Subtest items. The numbers indicate the sequence in which the cubes are tapped by the examiner in presenting the items when the frame is viewed from the examiner's perspective and the subject seated opposite. The numbers are indicative of each cube as if they were actually numbered 1, 2, 3 and 4 respectively from left to right

FIGURE 5.1 Knox Cube Imitation Subtest Materials

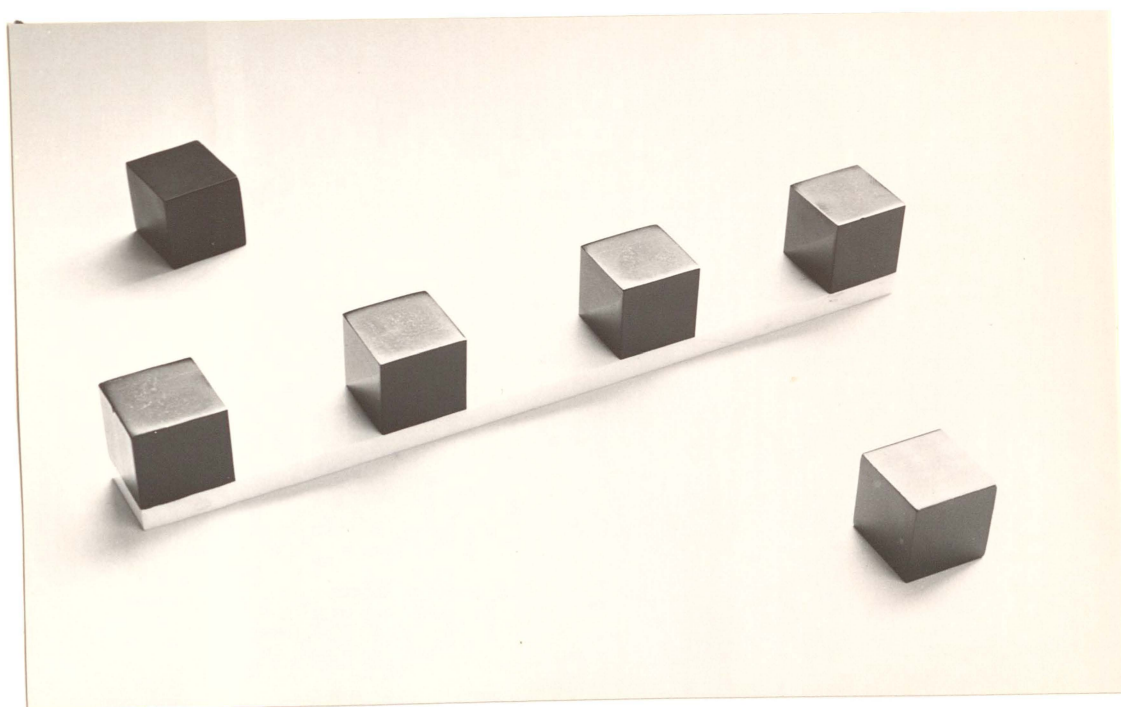


TABLE 5.1 The Block Tapping Sequences Employed in the
Knox Cube Imitation Subtest of the QT

<u>Item</u>	<u>Sequence</u>						
Practice Item 1	1	2	3	4			
Item 1	1	4	3	2			
2	4	1	2	3			
3	2	1	4	3			
4	3	1	4	2			
5	2	4	1	3			
6	3	2	1	4			
Practice Item 2	1	2	3	4	1		
Item 7	1	3	1	2	4		
8	4	3	2	3	1		
9	3	2	4	3	1		
10	2	1	3	4	2		
11	1	2	3	2	4	1	
12	2	1	4	2	1	3	
13	3	1	2	4	1	3	
14	1	4	2	3	2	4	1
15	3	1	4	2	4	1	3

from the examiner's position. Testing is discontinued after the failure of three consecutive items. Henceforth this subtest will just be referred to as the Knox Subtest.

(b) Beads Subtest

A frame with two parallel grooves is employed in administering this subtest. The frame is placed between the subject and the examiner and the examiner proceeds to construct a set bead sequence in the nearest groove. This is done hidden from the subject's view by the plastic cover provided. The bead sequence is then shown to the subject for a set time. It is then covered and the subject must construct the same sequence from memory in the other groove with the loose beads provided.

The frame has a fixed stopper at one end which is always oriented to the examiner's left. A second stopper can be moved along the grooves so that for any item the same length of groove used by the examiner is available to the subject. Since this stopper rests against the examiner's bead sequence the subject's groove is equivalent to the bead sequence in length thus implicitly defining the number of beads required to solve the item. By this procedure any advantages accruing to a subject who is able to count the number of beads required are limited.

The beads employed are of three shapes; spherical, cylindrical and cubical. The test materials are shown in Figures 5.2 and 5.3. The Beads subtest consists of 10 items ranging from 4 to 9 beads in the item sequences. The two practice items are 3 bead sequences. The exposure time for the subject to view the bead sequences increases from 10 to 25 seconds as the items lengthen. The items and exposure times are set out in Table 5.2. Testing is discontinued after three consecutive failures.

FIGURE 5.2 Beads Subtest Materials

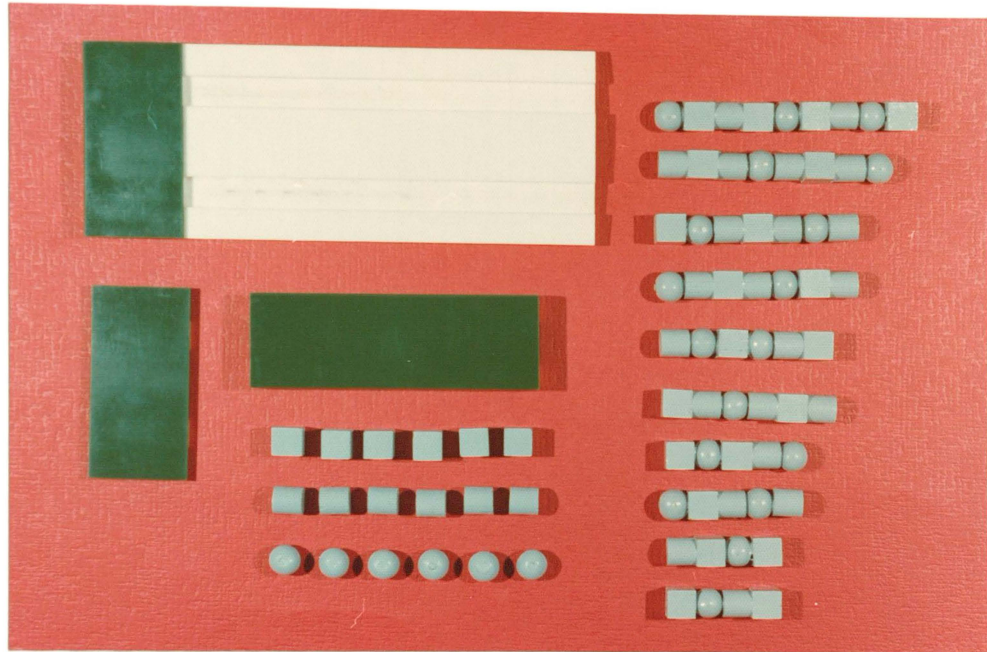


FIGURE 5.3 Item 6 of the Beads Subtest as Presented to Subject

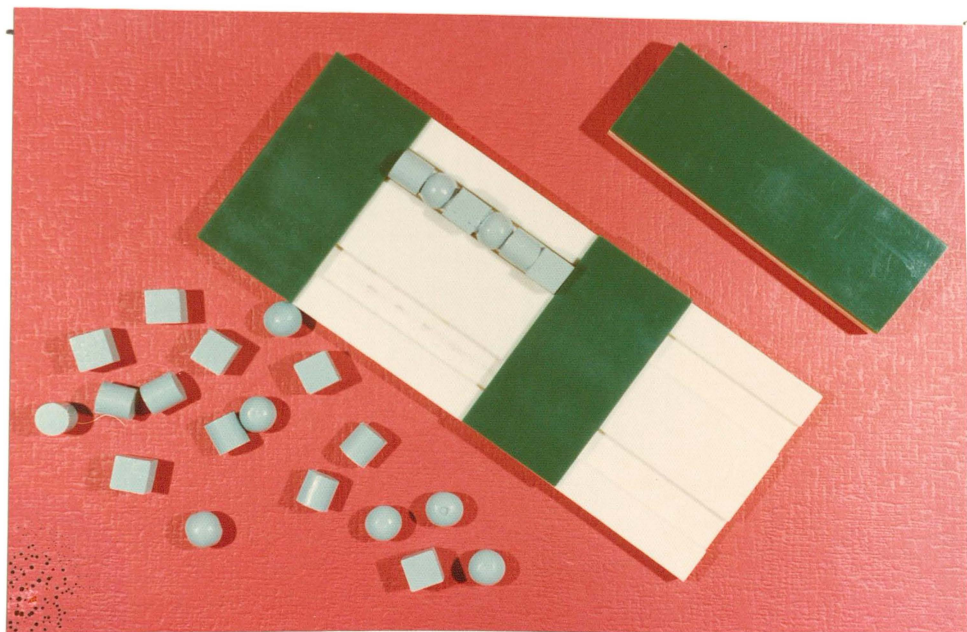


TABLE 5.2 The Bead Sequences Employed in the
Beads Subtest of the QT

<u>Item</u>	<u>Bead Arrangement</u>								<u>Exposure Time</u>	
Practice Item 1	=	0	+						Unlimited	
2	+	=	0						Unlimited	
Item 1	+	0	=	+					10 seconds	
2	=	+	0	+					10 seconds	
3	0	+	=	0	=				10 seconds	
4	+	0	+	=	0				10 seconds	
5	+	=	0	=	+	=			15 seconds	
6	=	0	+	0	=	+			15 seconds	
7	0	=	+	=	0	+	=		20 seconds	
8	+	0	=	+	=	0	=		20 seconds	
9	=	+	=	0	=	+	=	0	25 seconds	
10	0	+	=	+	0	+	=	0	+	25 seconds

0 represents a sphere

+

represents a cube

= represents a cylinder

(c) Passalong Subtest

This subtest requires the subject to slide blue and red tiles in a frame from a fixed starting position to an arrangement which is printed on a stimulus card. The card is the same size as the frame and the depicted tiles exactly the same size and colour as those in the frame. The subject must simply move the tiles in the frame so that they occupy exactly the same location as represented on the card. This the subject may only do by sliding the tiles in the frame. The task is described by Kearney (1966b) as "clear and unambiguous" (p.109). The subtest consists of seven items, each of which must be completed in a liberal time limit to obtain credit. The starting positions of the practice item and the seven subtest items are shown in Figures 5.4 to 5.12. The respective time limits are also noted. The subject is presented with the frame and tiles as illustrated and a card depicting the required rearrangement of the tiles. This is shown for Item 4 in Figure 5.13.

The testing discontinuation rules are such that the subject must pass one of the first pair of items (1 and 2) in order to continue to the next pair (3 and 4) and one of this pair must be passed to continue to the next pair (5 and 6). If one of the last pair is solved correctly then Item 7 in one of its alternative forms (7A or 7B) is attempted. The choice of administering Item 7A or 7B depends on the direction the subject moves the red tile in Item 6. This is taken as an indication of the direction-set favoured by the subject. The alternative administered is that which is more easily solved if the subject employs the same direction-set.

(d) Form Assembly Subtest

This subtest requires the use of a white 14 by 6 inch plastic base board on which seven black plastic shapes (triangles, squares, etc.) are mounted. The black plastic shapes are raised on the base board on

FIGURE 5.4 Passalong Subtest - Practice Item

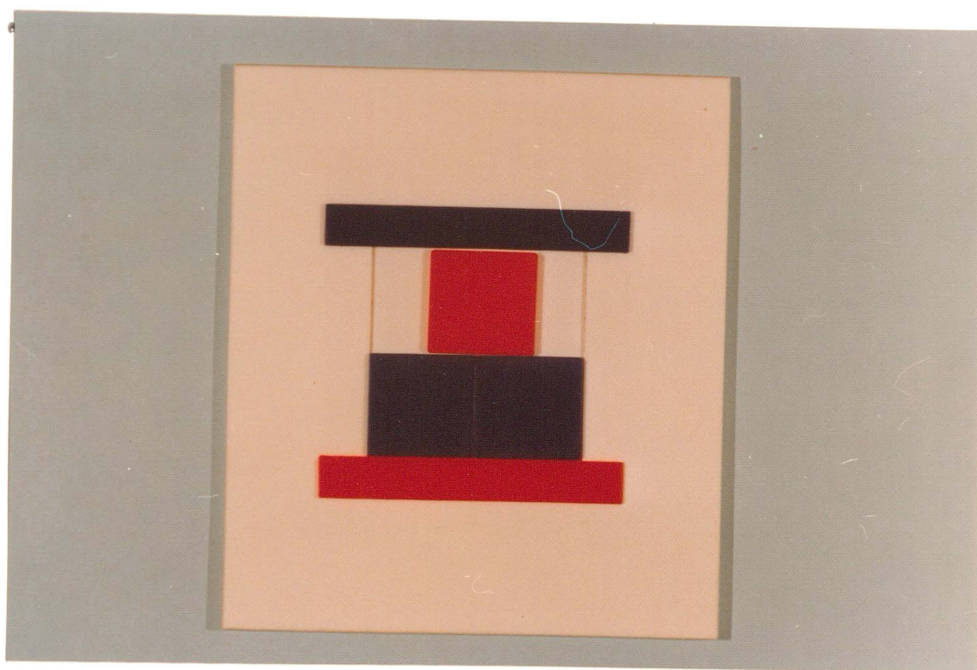


FIGURE 5.5 Passalong Subtest - Item 1
(Time limit: 60 seconds)

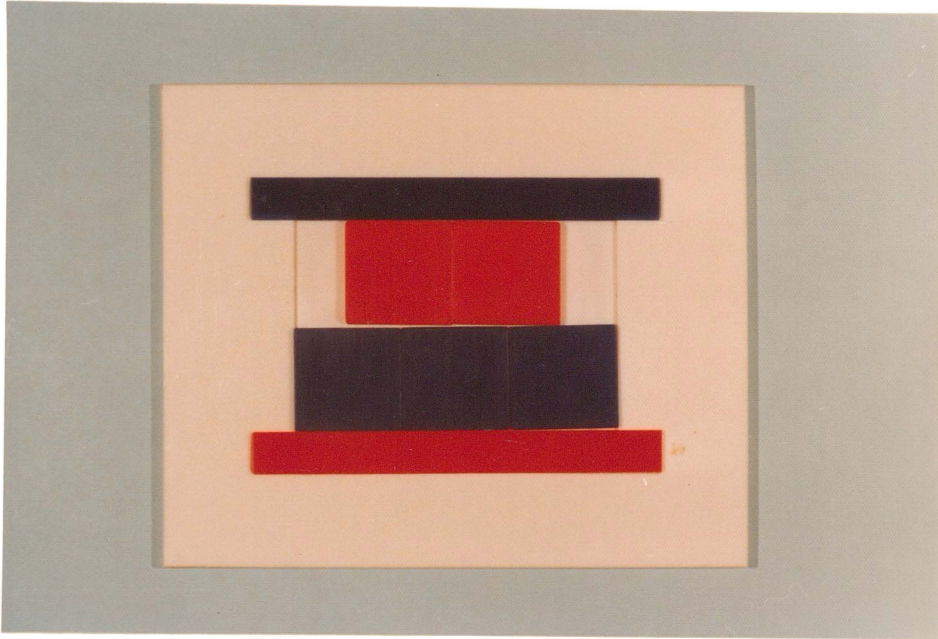


FIGURE 5.6 Passalong Subtest - Item 2
(Time limit: 90 seconds)

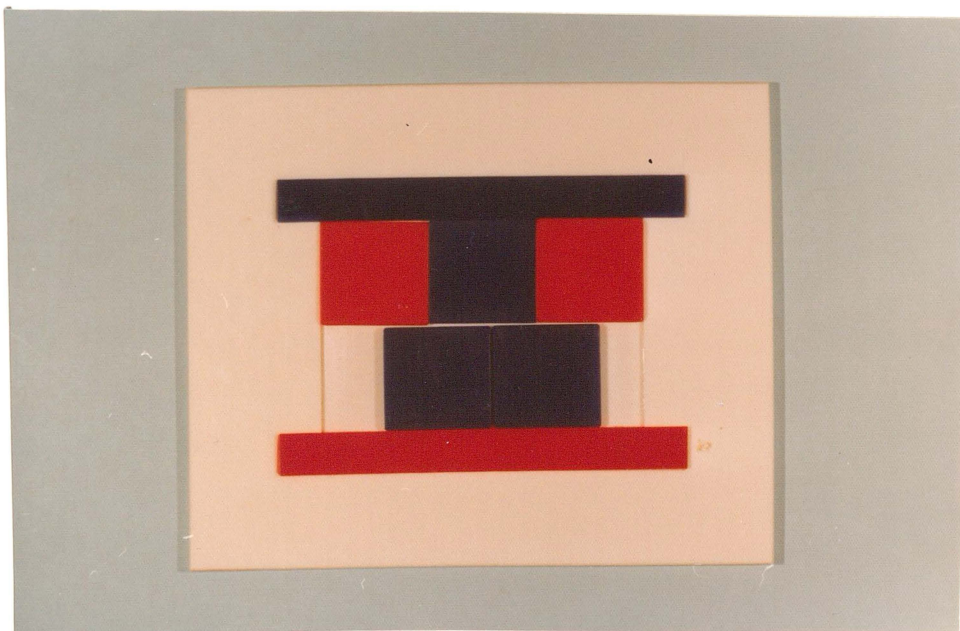


FIGURE 5.7 Passalong Subtest - Item 3
(Time limit: 90 seconds)

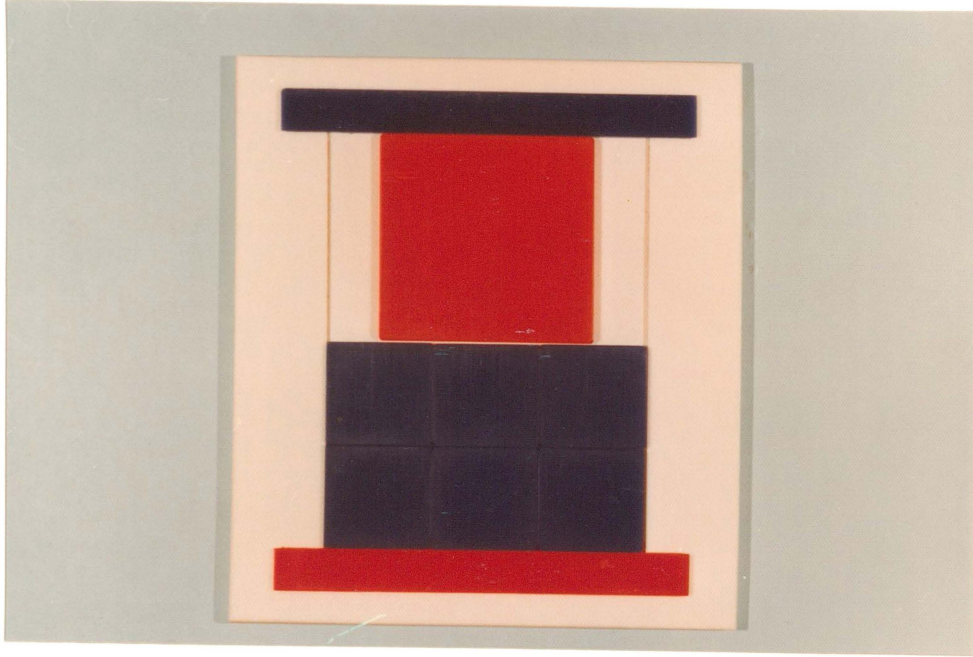


FIGURE 5.8 Passalong Subtest - Item 4
(Time limit: 90 seconds)

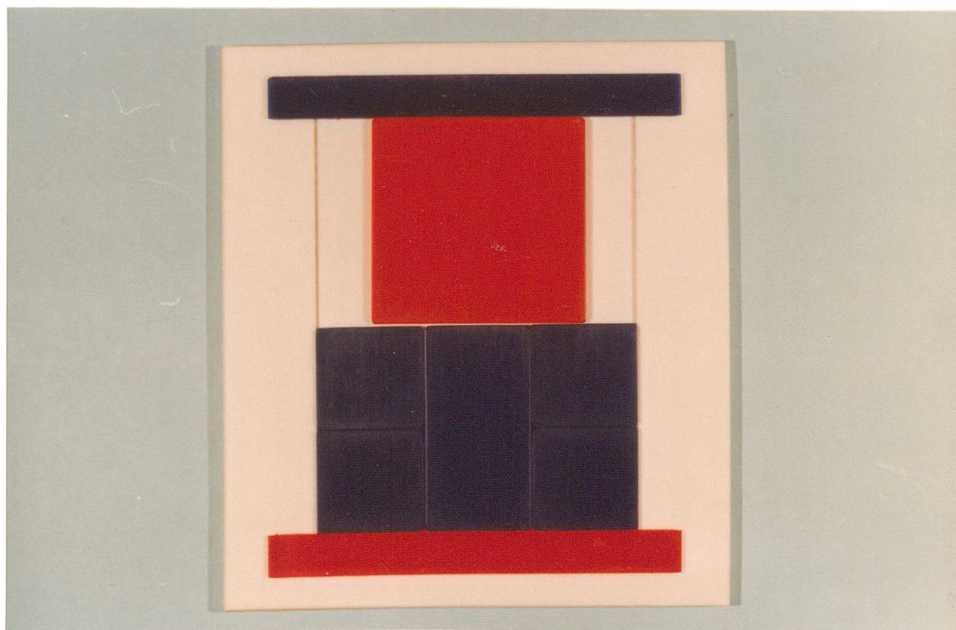


FIGURE 5.9 Passalong Subtest - Item 5
(Time limit: 90 seconds)

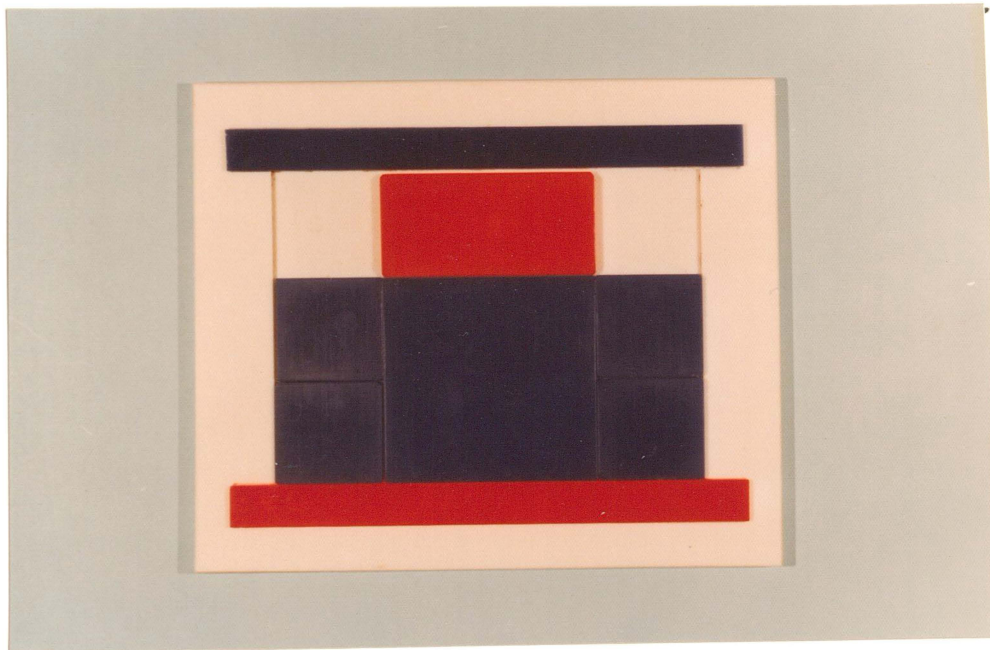


FIGURE 5.10 Passalong Subtest - Item 6
(Time limit: 120 seconds)

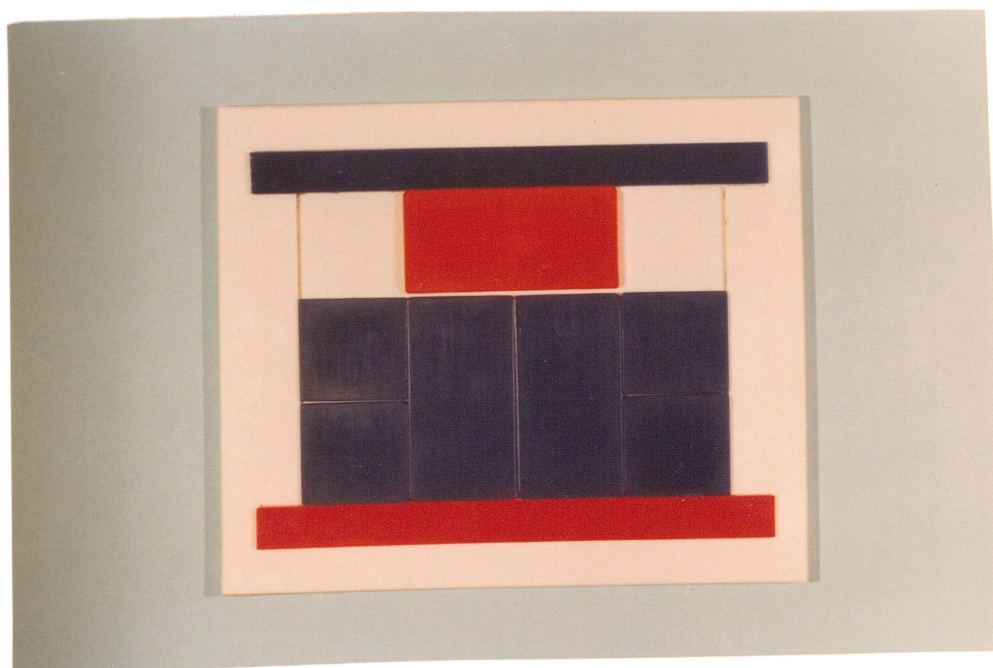


FIGURE 5.11 Passalong Subtest - Item 7A
(Time limit: 120 seconds)

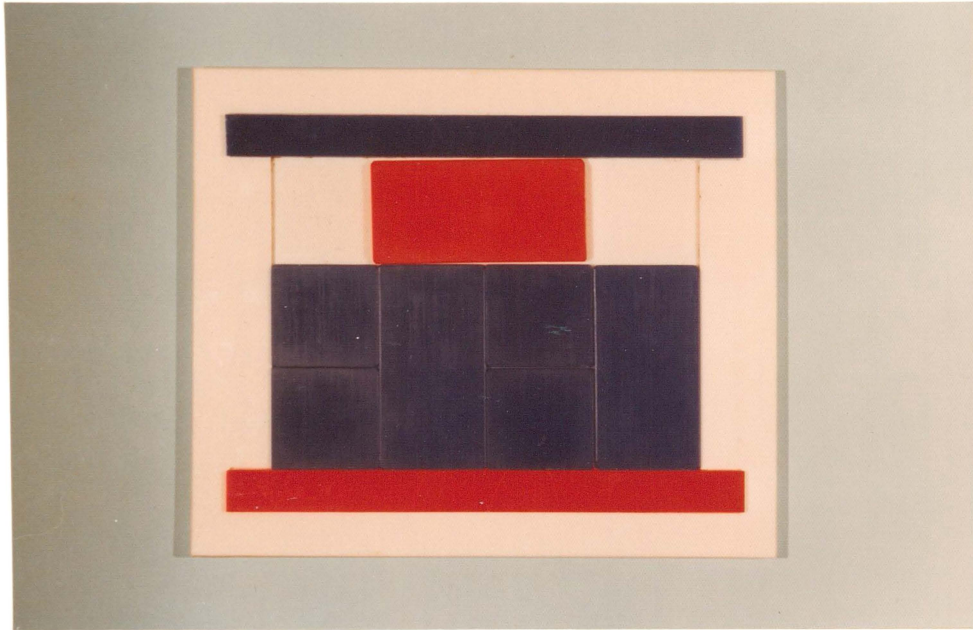


FIGURE 5.12 Passalong Subtest - Item 7B
(Time limit: 120 seconds)

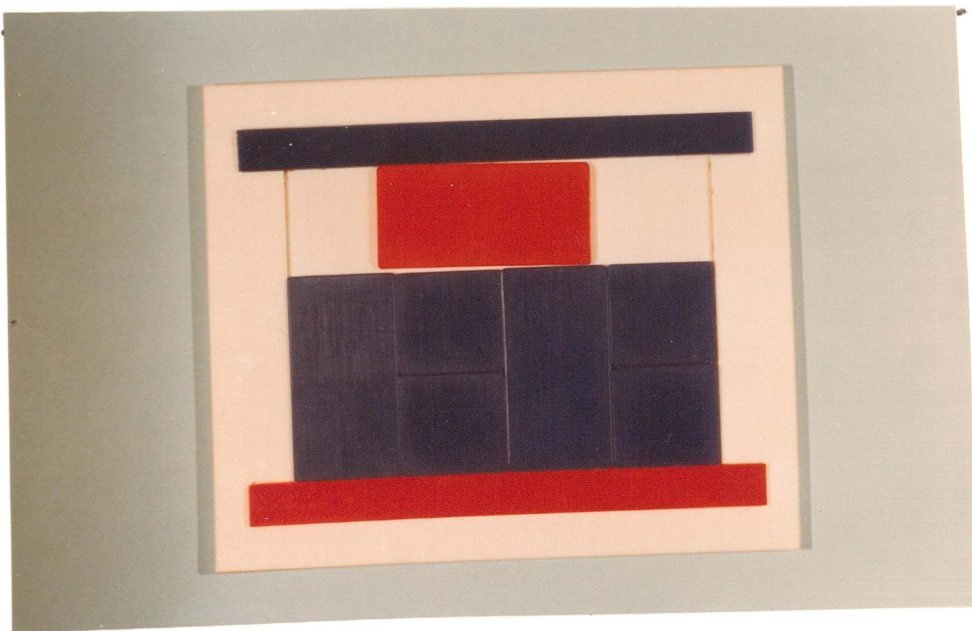
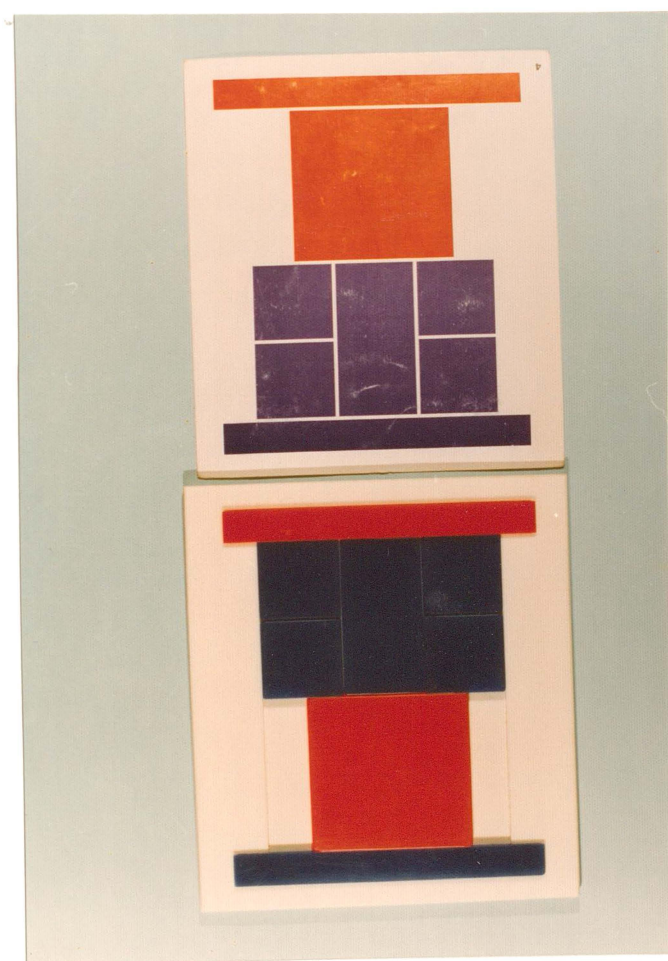


FIGURE 5.13 Item 4 of Passalong Subtest and Stimulus
Card as Presented to Subject



slim grey plastic pieces of identical shapes. A grey plastic 4 by 4 inch square is mounted on one side of the base board. In each item a black plastic stencil is placed on the grey square. The black plastic stencils are incomplete 4 by 4 inch squares thus an area of grey is left uncovered. The uncovered grey area can be filled exactly by using one, two or three of the seven black shapes to form a square. The subject's task is to identify the piece or pieces required to cover the grey area and form the black square.

. . The items systematically require one, two and three piece solutions with pieces arranged separately and then adjacent to one another. Rotation of the pieces becomes necessary as item difficulty increases. Kearney (1966b) provides the following clear description of the four types of items in the subtest -

"The first or 'one-piece solution' items require only one of the seven set pieces to complete the design. There are three such items (Items 1 - 3). The next type requires two pieces to complete the design. These are the 'two one-piece solution' items. The areas to be covered in solution are not adjacent and the items can be solved as if they required one-piece solutions. These are Items 4 - 6. The next type of item also requires two of the set pieces in its solution but they must be adjacent. These 'two-piece solution' items (Items 7 - 10) are different from the 'two one-piece solution' items in that the area to be covered is not identical with any single set piece but with a combination of two of them. The 'three piece solution' items (Items 11 - 13) have only one area to be covered and the three set pieces must be adjacent to each other to cover this" (pp.114-115).

Two practice items of the 'one-piece solution' type introduce the subtest. 'Two-piece solution' and a 'three-piece solution' practice items are administered to introduce these more difficult items. Loose copies of two of the black plastic shapes are used to demonstrate the solutions to all the practice items. The first practice item does not require any rotation of the solution piece but the second practice item requires the subject to rotate the solution piece through 90° to properly cover the missing area.

The subject need only point to the pieces to indicate the solution while the examiner indicates through gesture the number of pieces required. Figure 5.14 shows the form board, Figure 5.15 the form board as seen by the subject for Item 5 and Figure 5.16 the 4 practice items and the 13 black plastic item stencils. Table 5.3 lists the sequencing of practice items, the type of solution required by each item and the solution pieces by the alphabetic code used on the QT test form (see Appendix A).

The subject must pass one item in each block of item types (i.e. 'one-piece solutions', 'two one-piece solutions', 'two piece solutions') to be able to continue with the next block of items. Thus, one must pass one of Items 1 - 3 to attempt Items 4 - 6 and pass one of Items 4 to 6 to attempt Items 7 to 10. Similarly, one of Items 7 to 10 must be passed in order to proceed to the final set of items.

Notwithstanding the above discontinuation rules Item 13 of the Form Assembly Subtest is always administered. This unusual convention was adopted because -

"It was found that some children, particularly Aboriginal children, often get this item correct after failures with several apparently easier problems. What appears to happen is that they observe that the solution requires three acute angles and that only designs B, E and F have acute angles European adults find the item difficult" (McElwain and Kearney, 1970, p.48).

No time limits are imposed in this subtest.

(e) Pattern Matching Subtest

This subtest requires the subject to construct a pattern in a square plastic frame with red, white, and red and white plastic tiles. The required pattern is presented to the subject on a card. The card's external dimensions correspond to those of the frame and the pattern and area in the frame for its construction are of identical dimensions. The

FIGURE 5.14 Form Assembly Subtest Formboard as viewed by Subject

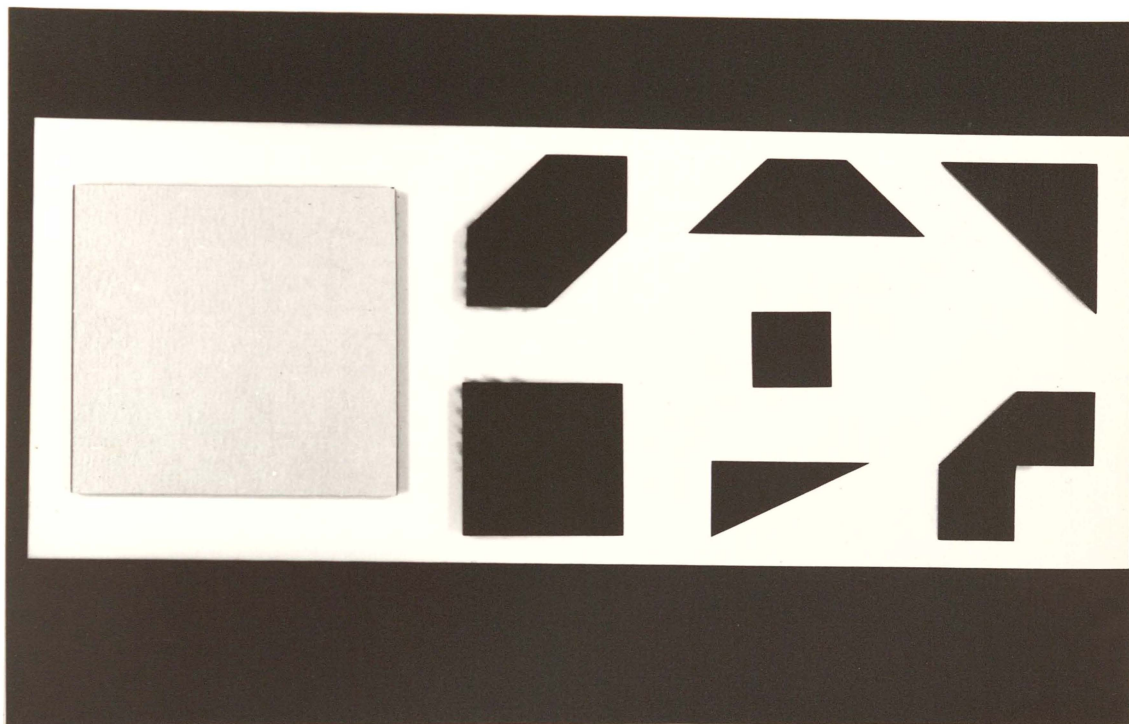


FIGURE 5.15 Item 5 of the Form Assembly Subtest as viewed by Subject

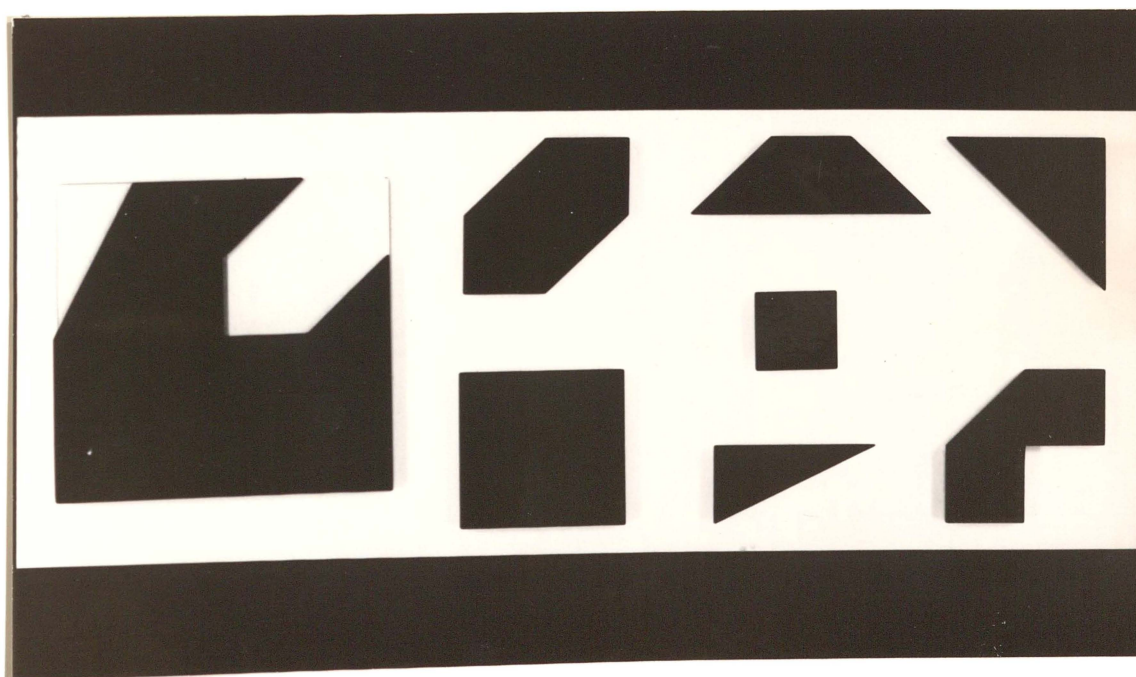


FIGURE 5.16 Form Assembly Subtest Practice Items (4) and
Test Items (13) as viewed by Tester

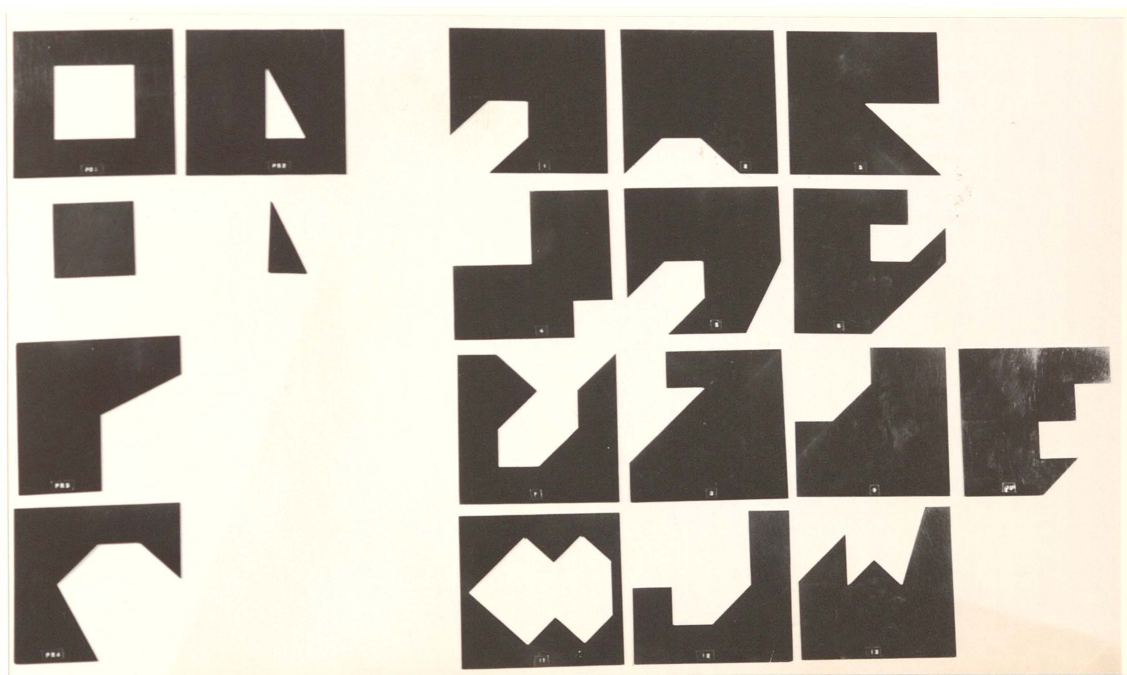


TABLE 5.3 Order of Practice and Test Items of the
Form Assembly Subtest of the QT

<u>Item</u>		<u>Piece(s) needed for Solution</u>
Practice Item	1	C
	2	B
One-Piece Solution Items	1	G
	2	F
	3	F
Two One-Piece Solution Items	4	C + D
	5	B + G
	6	A + E
Practice Item	3	BC
Two-Piece Solution Items	7	FG
	8	DE
	9	AD
	10	AE
Practice Item	4	BCF
Three-Piece Solution Items	11	ADG
	12	ACD
	13	BEF

⁵ Letters refer to solution pieces as coded on Queensland Test Record Form (See Appendix A).

FIGURE 5.17 Pattern Matching Subtest Practice Item (1) and Test Items (15) as viewed by Tester

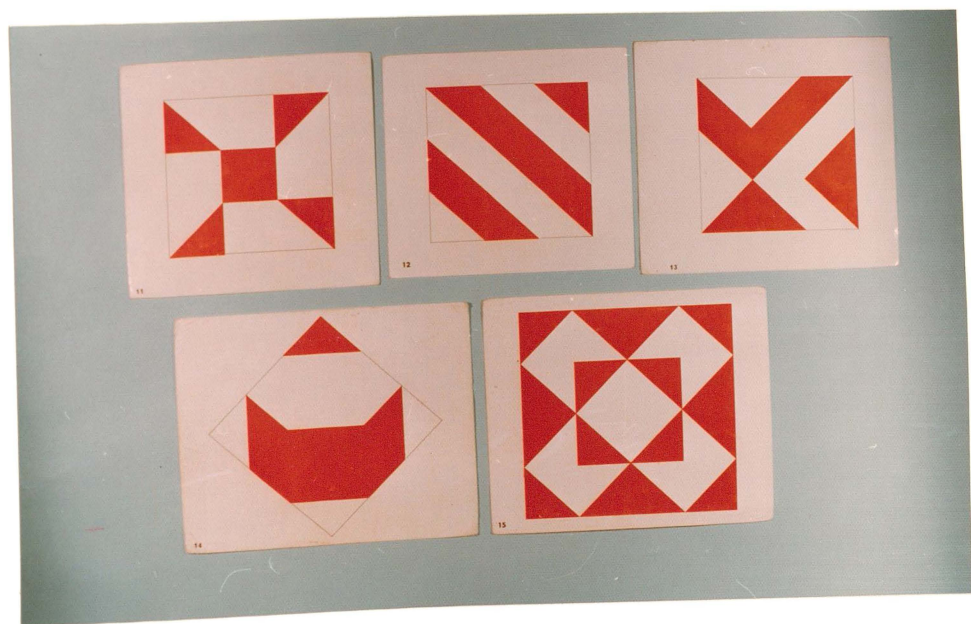
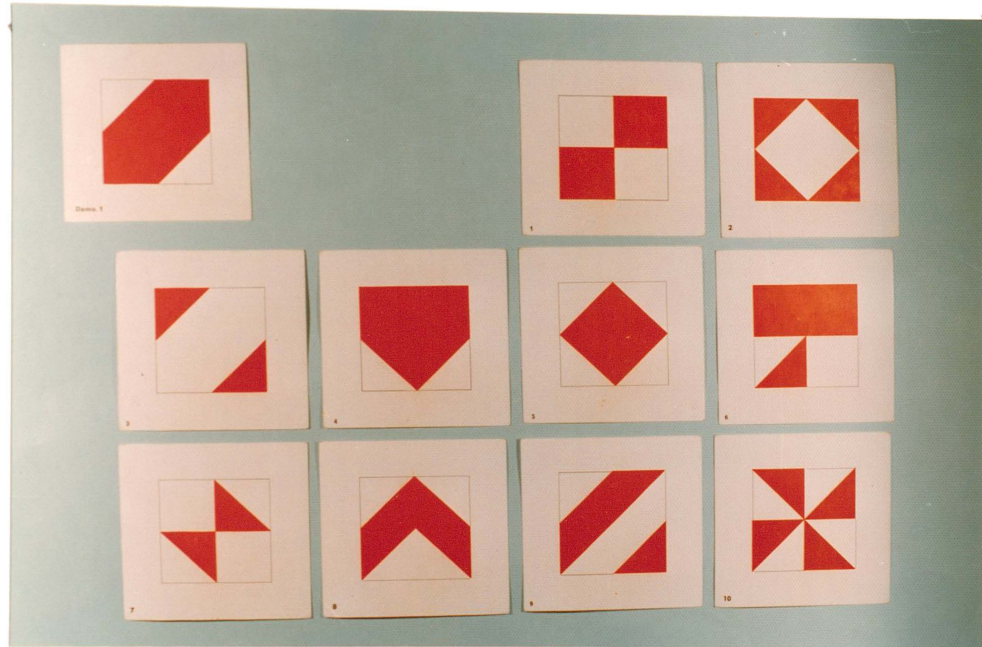
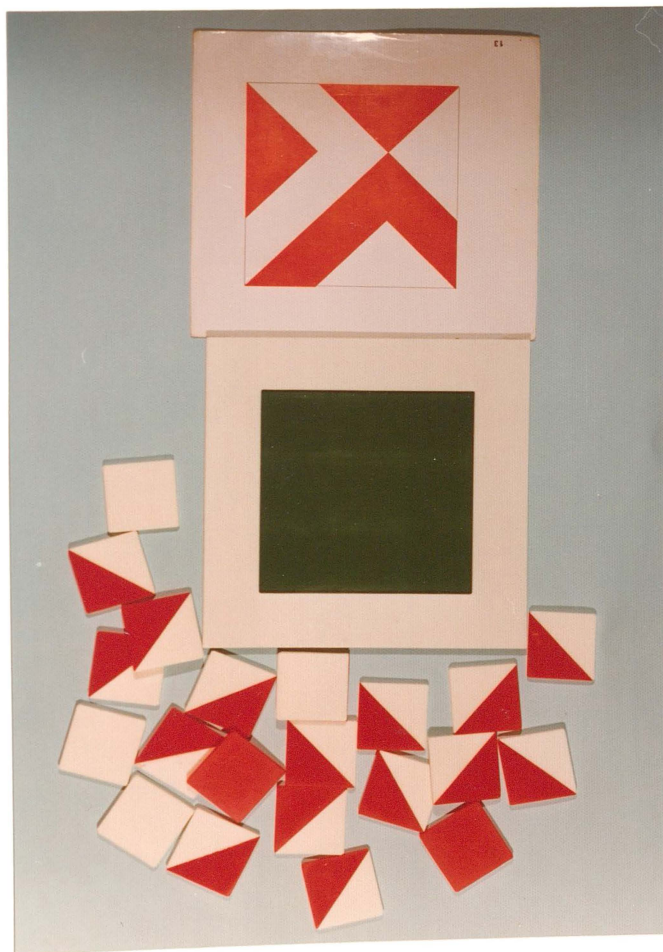


FIGURE 5.18 Item 12 of the Pattern Matching Subtest
as Presented to Subject



use of the frame is seen as "clearly defining the boundaries of the item and giving the task a focal point" (Kearney, 1966b, p.116). The subject is provided with six white, six red and six tiles split diagonally into red and white halves. All tiles are $1\frac{5}{16}$ inch square. Four plastic frames are required by the subtest. The first holds 4 tiles, the next two 9 tiles and the fourth 16 tiles. The subtest consists of 15 items. Items 1 to 10 require four tiles, Items 11 to 14 nine tiles and Item 15 sixteen tiles. Figure 5.17 shows the Pattern Matching Subtest item cards and Figure 5.18 shows Item 13 as seen by the subject.

The subtest commences with one practice item which the subject must solve correctly before Item 1 is administered. Generous time limits are imposed on each item. McElwain and Kearney (1970) comment that "These are not thought to give a 'speed' component to the performance but to provide limits outside which it is not practicable to extend the testing time" (p.30). The time limits are set out in Table 5.4. A discontinuation rule of three successive failures is employed with this subtest.

Some General Notes

Detailed administration procedures for each subtest are set out in the Queensland Test Handbook (McElwain and Kearney, 1970, see pp. 39-49). A Queensland Test Record Form is to be found in Appendix A. The Record Form itself repeats details on scoring and the discontinuation rules.

The total score on the QT is obtained by the simple addition of the five subtest scores, the items of which are scored on a right (+1) or wrong (0) basis. The maximum possible score is 60.

The QT authors view the scale as a 'test' and not a 'battery of tests'. They say that -

TABLE 5.4 The Number of Tiles Required and Time Limits for Each Item of the Pattern Matching Subtest of the QT

<u>Item</u>	<u>Number of Tiles Required</u>	<u>Time Limits</u>
Practice Item	4	Unlimited
Item 1	4	2 minutes
2	4	2 minutes
3	4	2 minutes
4	4	2 minutes
5	4	2 minutes
6	4	2 minutes
7	4	2 minutes
8	4	2 minutes
9	4	2 minutes
10	4	2 minutes
11	9	3 minutes
12	9	3 minutes
13	9	3 minutes
14	9	3 minutes
15	16	4 minutes

"Present experience suggests that all sub-tests be given and that great caution be exercised in drawing inferences from single sub-tests or from differences or ratios between the sub-tests" (McElwain and Kearney, 1970, p.13).

The authors stress the need to administer the five subtests in the order suggested and insist that the Knox Cube Imitation Subtest be given first since it quickly engenders the 'set to imitate'. McElwain and Kearney argue that while this subtest is the least reliable and valid its omission would increase the difficulty of administering the other subtests. They also view the preservation of the item orders in the subtests as important noting that the items of "The last three sub-tests - Passalong, Form Assembly and Pattern Matching - all have 'built-in progressive learning'" (McElwain and Kearney, 1970, p.13).

McElwain and Kearney have described a number of situations in which the use of the QT would be appropriate. First, since it can be administered without verbal communication between examiner and subject it has an obvious application in situations where communication barriers exist. The barriers may be those of culture and language. Alternatively, the barriers may have been brought about by deafness. Second, it is suggested that the QT might be applied in selection situations where level of educational attainment is largely fortuitous and third, in situations where teachers require "an indication of 'basic capacity' for an educationally backward pupil" (McElwain and Kearney, 1970, p.11).

The QT authors recognised also situations where the QT might not prove to be the most suitable test. They suggested that measures containing education factors correlated with 'basic cognitive capacity' are probably better than the QT with normal European subjects. Handicaps such as poor vision or very poor manual functioning on the part of a subject would make the QT inappropriate. In addition, they do not think the QT should be used with subjects below the age of seven

years. "The QT becomes less reliable at this level. Results obtained for children in age ranges lower than 7 should be treated cautiously" (McElwain and Kearney, 1970, p.12).

Comments on Interpretation

The discussion thus far has focused on the development and form of the QT rather than on theoretical questions concerning what the test might measure. Terms like 'basic cognitive capacity' have been employed above when done so by the test authors. However, theoretical questions are of some moment and are now considered more fully.

The development of the QT does appear to have been influenced by at least two important considerations. First, the adoption of a pragmatic approach to the solution of the particular psychometric problems being encountered and second, a general preference for the 'British school of thought' on the concept of intelligence with its emphasis on a paramount general factor.

The pragmatism, spurred on by the need to respond to a set of practical problems, can be seen in the way McElwain and Griffiths (1957), Ord (1959) and Kearney (1966b) all commenced from the base laid by other test constructors in choosing item types that had been used with some measure of success in other situations where the problems of measuring intellectual abilities across cultures had arisen. Thus, the subtests of the QT are not new but are the current endpoints in a long chain of developments guided as much by day to day experience as by any lengthy debate over theoretical points or properties.

Theoretical discussion as such has arisen out of consideration of the QT's content and the results of its application rather than from any theoretical scheme proposed prior to the construction of the test. Put simply, item types thought likely to work and said to measure 'general cognitive capacity' were tried, those found to be psychometrically satisfactory and able to be relatively easily administered to

subjects of markedly different cultural backgrounds were retained. The item types incorporated into the QT have a history of use in measures of general intelligence in the 'g' sense and this interpretation has been retained.

While there may be a priori grounds for accepting a general intelligence - 'g' type - interpretation of results on the QT, there are also the various validity studies to be considered.

Ord (1970) speaking of his P.I.R. Test stated that "As the first four sub-tests have been demonstrated to measure g to different extents when applied to European populations, it seems likely that all sub-tests are likely to be measuring a g factor"(p.156). Ord has detailed various Papua New Guinea factor analytic studies involving the item types of the P.I.R. Test and the New Guinea Performance Scales. There is evidence that both spatial and memory factors might be expected as well as the predominant 'g' depending on the samples and range of measures involved. Prior to Ord's report, Kearney (1962) factor analysed a limited amount of data from the P.I.R.-IV Test when used with both Australian Aboriginal and Papua New Guinean subjects. All subtests loaded highly on the first factor which was interpreted as a general intelligence factor.

Kearney (1966a) continuing his studies with Ord's measures made a factor analytic study of the New Guinea Performance Scale with Papuan subjects from the Orokaiva District. He obtained a general factor which loaded on five of the seven subtests. The second factor was interpreted as a "spatial-closure-Gestalt type of factor" (p.22) and the small third factor was tentatively described by Kearney as being associated with manipulative ability. Ord (1970) disputed the interpretation on this third factor on the grounds that other 'manipulative' subtests should have been implicated. However, an alternative interpretation was not suggested.

subjects of markedly different cultural backgrounds were retained. The item types incorporated into the QT have a history of use in measures of general intelligence in the 'g' sense and this interpretation has been retained.

While there may be a priori grounds for accepting a general intelligence - 'g' type - interpretation of results on the QT, there are also the various validity studies to be considered.

Ord (1970) speaking of his P.I.R. Test stated that "As the first four sub-tests have been demonstrated to measure g to different extents when applied to European populations, it seems likely that all sub-tests are likely to be measuring a g factor"(p.156). Ord has detailed various Papua New Guinea factor analytic studies involving the item types of the P.I.R. Test and the New Guinea Performance Scales. There is evidence that both spatial and memory factors might be expected as well as the predominant 'g' depending on the samples and range of measures involved. Prior to Ord's report, Kearney (1962) factor analysed a limited amount of data from the P.I.R.-IV Test when used with both Australian Aboriginal and Papua New Guinean subjects. All subtests loaded highly on the first factor which was interpreted as a general intelligence factor.

Kearney (1966a) continuing his studies with Ord's measures made a factor analytic study of the New Guinea Performance Scale with Papuan subjects from the Orokaiva District. He obtained a general factor which loaded on five of the seven subtests. The second factor was interpreted as a "spatial-closure-Gestalt type of factor" (p.22) and the small third factor was tentatively described by Kearney as being associated with manipulative ability. Ord (1970) disputed the interpretation on this third factor on the grounds that other 'manipulative' subtests should have been implicated. However, an alternative interpretation was not suggested.

Richardson (1970) has factor analysed a small set of data from the QT and group intelligence and educational attainment measures with Papua New Guinean Armed Service recruits. He found the QT subtests to load into a second factor which was largely independent of the group measures. He viewed the group tests as measuring 'acculturation' influences while the QT results reflected a general intelligence factor divorced from such influences.

The most important studies and comments bearing upon the interpretation of the QT are to be found however, in Kearney's research (1966b) and in the Queensland Test Handbook by McElwain and Kearney (1970).

Kearney (1966b) reports the results of the application of factor analytic procedures to QT data from a number of Australian Aboriginal and European samples. Two-factor solutions were sought with the factor loadings rotated to maximize on age on the first factor. It was postulated that "The first factor could represent processes closely correlated with age, e.g. maturation, developmental influences" (Kearney, 1966a, pp. 268-269). The second factor if free of maturational effects might thus represent general educational and environmental experiences.

Rotated factor loadings on the first factor were found to be moderate to high on all five subtests and age. Kearney (1966b) has offered the following interpretation -

"The first rotated factor is general and tentatively can be identified as being genetically determined in much the same way as Vernon (1961) sees g as being 'largely dependant on some psycho-physiological and innate property of the higher nervous system'" (p.269).

The second rotated factor loaded most highly for each sample on the Form Assembly Subtest followed by the Pattern Matching Subtest. As in the interpretation of the Orokaiva data the spatial and manipulative components of these subtests were noted. Kearney suggests that it is a 'k' type factor depending on formal or informal learning and experience with spatial shapes and their manipulation, either concretely or abstractly. This factor is seen as being much more environmental in origin.

Factor analytic evidence bearing upon QT interpretation also comes from the application of a principal components analysis to the tetrachoric correlations between items. Four components were extracted. The first which accounted for almost half the variance with loadings on all items was seen as being "of a general nature" (Kearney, 1966b, p.234). The second component loaded primarily on the Knox items and the third on the Beads items. A small fourth component gave positive loadings on most Form Assembly items and negative loadings on most Passalong items.

Kearney (1966b) viewed the differences between the subtest and item analyses as resulting primarily from differences in the form of the data and the correlational techniques employed. The differences in the second and third order factors were not seen to seriously challenge the overall 'g' - or general intelligence factor interpretation of QT results.

McElwain and Kearney (1970) report a number of criterion-related validity studies with the QT involving both Aboriginal and European subjects. The criterion measures have included teacher ratings of 'observed general ability', place in class, highest school grade, measures of educational attainment and even a productivity measure based on ownership of cash crop trees. The studies have been either correlational in nature or employed a contrasted group method. The positive correlations with various criteria and the contrasted group differences have been seen as supporting the view that the QT is primarily measuring a broad general ability factor. Similarly the positive correlations between the QT and a number of other general ability measures have also been taken as support for this interpretation.

Weight has also been given to the theoretical formulation that general ability or intelligence is subject to maturational factors and consequently test scores should increase monotonically with age up to

some limit. Thus linear regression equations of QT results on age were determined for a number of samples. For all of the samples, except one from the armed services and therefore adults, the slope of the regression line was ascending. McElwain and Kearney (1970) have reported in some detail these regression analyses. It is sufficient to note that they do conform to the formulation that general ability is subject to maturational factors.

A little more on what the QT might measure can also be gleaned from remarks concerning its possible use. Earlier it was noted that McElwain and Kearney (1970) stressed the fact that they regarded the QT as a single measure rather than a battery. They cautioned against the use of subtest scores as measures of differentiated skills or intellectual abilities. Although, when briefly discussing such matters as cultural differences in cognitive styles they do say when speaking of Australian Aboriginal and European QT comparisons that, "We have a few observations on differences in cognitive style and these will be reported when the sub-tests are discussed" (p.3). Without adequate cautions such remarks could however, easily lead to unwarranted emphasis being placed on subtest differences.

While spin-off into cognitive styles research from the use of the QT might be both possible and welcomed, it would seem that in the first instance an instrument conceived and nurtured in the tradition of 'g' will remain best understood in those terms. Consequently its use in these New Zealand studies has been based on the view that the QT is first and foremost a measure of general intellectual ability.

Notes on Research Applications

The QT and related instruments have seen considerable use over the last ten years, especially in the South West Pacific.

Ord (1970, 1972a) has already given accounts of the extensive application of the P.I.R. Test and the New Guinea Performance Scales in Papua New Guinea. His accounts detail the development of these tests, give data on reliability and validity and present some normative data for select Papua New Guinean samples. While these instruments were important forerunners to the QT, the course of events has been such that they now see little use in their own right. Certainly this has been the case in Papua New Guinea of late where much greater emphasis has been placed on more specialised group measures (see St. George and Preston, 1976). The Pacific Design Construction Test which Ord (1970) had already recommended as a 'short-form alternative' to the New Guinea Performance Scales has however, been used as a selection instrument in both public and private employment sectors in Papua New Guinea (see Preston, St. George and St. George, 1974b). It may well continue to be found useful in these areas.

MacArthur (1974) included three New Guinea Performance Scale subtests in a factor analytic study of over thirty primarily cognitive measures with samples of Central Canadian Eskimo and Nsenga African adolescents from Zambia. MacArthur was particularly interested in the "constructs assessed by the Single Form Assembly, Passalong, and Design Construction subtests....." (p.51). In considering the oblique first order factors obtained and the factor loadings of the three subtests for the two sample groups he noted that -

"..... while in a very general way they seem to be loaded with a sort of spatial-field independence correlated with other cognitive factors, these tests are clearly assessing rather different constructs in the two cultures" (p.52).

When the results of the application of factor analytic procedures within the culture groups but across an age range were considered, MacArthur noted that the three subtests were assessing a factor that was generally held in common with the majority of the other cognitive measures in the battery. He concluded that the data -

"..... lends support to the NGPS author's conjecture that for purposes of general assessment of mental ability these three tests have useful validities in cultures other than the New Guinea and Australian Aboriginal cultures in which they were developed" (p.54)

Kearney (1966a) in developing the QT applied the item types to Papuan subjects from the Orokaiva district. Dividing his sample into groups according to villages, Kearney found subjects from the village of Beporo to achieve a significantly higher mean score on the measures of cognitive ability than all other groups. Conversely the subjects from the village of Ongoho achieved the lowest mean scores. An explanation in terms of the different histories of contact was offered. Beporo village had a long history of contact including the Gona Mission of some sixty years standing which had brought with it elementary schooling. The Ongoho sample, on the other hand, had received few opportunities for schooling.

Kearney also related the test results to the ownership of cash crop trees. This was used as an index of productivity. He found that subjects of higher cognitive ability owned more cash crop trees. Kearney clearly recognised the myriad of other factors that would bear on any such relationship and was content just to demonstrate that cognitive ability could not be entirely disregarded. The study also showed that the measures could be used in culturally diverse situations such as Papua New Guinea to obtain data of both theoretical and practical interest.

Kearney (1966b) in his research with the final form of the QT in Australia has applied the measure to large numbers of European and Aboriginal subjects. In the main the subjects have been children and as far as the Aboriginal subjects are concerned have come from 'low' and 'medium' contact situations (described in Kearney, 1966b, pp.151-166). Since features of these data will be discussed throughout this dissertation no attempt is made to summarize it at this point. It is

appropriate however, to note the general conclusions of McElwain and Kearney (1973) in relation to the comparative studies undertaken with the QT. They comment that -

"The results of this testing may be summed up by saying that the Aboriginal groups are inferior to Europeans, and in approximately the same degree as they have lacked contact with European groups It seems clear that test results are dependent to a considerable degree upon contact or some variable related to contact" (p.47).

Mention has already been made of Richardson's (1970) use of the QT as part of a test battery administered to Papua New Guinean recruits. Mean subtest and total score results of 107 subjects were reported. For 55 of the subjects additional group test, educational level and age data was available. The intercorrelations were factor analysed. The first factor loaded positively on the six group tests and educational level but negatively on age. This factor Richardson noted "appears to involve the ability to succeed at group tests with a cultural loading. It is therefore tentatively labelled an acculturation factor" (p.8). The second factor loaded on three QT subtests - Knox, Beads and Pattern Matching - and two group tests. It was interpreted as a general intelligence factor largely independent of cultural influences. The third factor loaded on the Knox, Beads and Passalong subtests, a group measure of speed and accuracy, and educational level. Richardson tentatively thought it might be a speed-type factor. The fourth factor was also obscure loading on Form Assembly and a word knowledge test. Richardson (1970) concluded that the QT was a measure of general intelligence that is perhaps relatively free of acculturation factors but not entirely free of educational effects.

On the basis of the report of this research the above interpretations do seem to be open to question. The decision to label the first factor 'acculturation' does seem hard to justify when the measures appear to be of the 'v:ed' nature talked about by Vernon (1969) and when no direct attempt to measure acculturation was actually made. Further-

more, that educational level should be associated with some subtests of the QT should not cause alarm. Kearney (1966a) had already alluded to educational effects in the QT data from Orokaiva studies. Perhaps the difficulty in obtaining more agreed interpretations lay with the rather loose use of the terms 'acculturation' and 'education effects'.

The QT has in recent years been employed in research in a number of Pacific countries other than Papua New Guinea and Australia. McElwain (n.d.) in 1965 undertook a small study into the issue of selection into secondary school in the Colony of the Gilbert and Ellice Islands. His unpublished report presents data from the administration of the QT to a small number of primary school children. The results for the same subjects from the local Secondary School Entrance Examination were also available. McElwain noted a strong degree of agreement between those subjects obtaining high QT scores and high Entrance Examination results. On the other hand, low QT scores tended to predict low Entrance Examination results. Recommendations on various educational selection strategies were made to the Resident Commissioner of the Colony.

In Fiji both Bennett and Chandra have employed the QT as a research instrument. Bennett (n.d.) initially reported on the application of the QT to very small samples of rural and urban Fijian children. There was little difference between the two samples in terms of the QT Total Score but a difference of over 5 points on the Pattern Matching Subtest favouring the urban sample. The important point of this small study was, however, to demonstrate that the QT could be effectively administered in the Fijian context.

Bennett and Chandra (1974) continuing their research with the QT reported on its use with 94 twelve-year old children broken down into three groups - urban Fijian, rural Fijian and rural Gilbertese. In terms of QT Total Score the urban Fijian group performed at a higher level than the rural Gilbertese followed by the rural Fijian. Bennett

and Chandra were particularly interested in subtle differences in performance. They argued from Berry's (Berry, 1971; Berry and Annis, 1974) ecology-culture-behaviour model that the urban Fijian group as a transitional group and the rural Gilbertese group as members of traditional-atoll and low food accumulating societies would perform better on tasks requiring spatial skills in comparison with the rural Fijian group. The latter group was described as being of a traditional but high food accumulating type society. Though it is not entirely clear from their paper, Bennett and Chandra (1974) appear to be arguing that the ecological demands of the first two societies would produce greater field independence and that this would be reflected in those subtests of the QT thought to be largely spatial in nature - namely Form Assembly and Pattern Matching. It transpired that the Pattern Matching Subtest did show the hypothesised difference but not the Form Assembly and that an unexpected difference occurred on the Beads Subtest. The results were read as a measure of support for the hypotheses and possible explanations were offered for the discrepant results.

Chandra (1975) has continued research in Fiji with the QT but with less direct emphasis on fitting QT subtest results into Berry's model. To the data from the urban and rural Fijian samples Chandra was able to add that from two urban Fijian-Indian samples and one rural Fijian-Indian sample. The mean age for all samples was between 11 and 13 years. Comparisons were made between the groups on the five subtests and QT Total Score. Results from the QT Total Score comparisons showed no statistically significant differences between the rural samples. The urban samples all performed better than the rural. All but one urban-rural comparison reaching statistical significance. Chandra views the urban-rural differences as resulting from environmental factors and urges their more detailed study if causation is to be fully understood.

A second section of Chandra's (1975) paper deals with a

comparison of samples of Fijian and Fijian-Indian children in three age groups on the Pattern Matching subtest. Increments in performance across the age range were observed as expected. When comparisons were made between gender within each ethnic group, males were found to score significantly higher. The comparisons between the ethnic group samples at each age level did not however, evidence any significant differences. Chandra has argued that the Pattern Matching subtest results and those from the QT as a whole "lend some support to the notion that the test is 'culture fair'" (p.190). It should be noted however, that McElwain and Kearney (1970) do not exactly make such claims for the QT. They state that it was not their intention to pursue the probably impossible task of constructing a 'culture-free' test, preferring to think of the QT as being "markedly 'culture reduced' when compared with most other tests in general use" (p.5). This may not imply the same as 'culture fairness' as used by Chandra in this instant.

The data to be reported in this dissertation stem from the first major use of the QT in the Polynesian region of the South Pacific. While this research was in progress, a small study in New Zealand (Zimmerman, 1971) and another in the Cook Islands (St. George, R. and St. George, Alison, 1974) involving the use of the QT were reported. The major findings are noted in this review of QT use.

Zimmerman (1971) applied the QT to a limited sampling of Maori and European borstal detainees. There were no significant differences between the groups. Very little other data pertinent to the QT was reported by Zimmerman.

St. George and St. George (1974) administered the QT to 90 Cook Islands children in 3 age groups. Mean subtest performances and inter-correlation analyses were reported as well as data on QT Total Score performances. For the present, discussion will focus primarily on the QT Total Score data.

When QT performances were tested by analysis of variance (henceforth ANOVA) procedures for the effects of age and gender, it was found that both had significant effects on performance. Inspection of the data clearly showed that QT scores increased with age and that age for age males outperformed females. When each subtest was analysed by ANOVA procedures it was found that all but the Knox Subtest evidenced age effects but that gender effects were found on only the Form Assembly and Pattern Matching Subtests. This result on the Pattern Matching Subtest is similar to that obtained by Chandra (1975) in Fiji.

In attempting to 'explain' the gender differences in performance St. George and St. George (1974) speculated upon differences in male and female socialisation and role in Cook Islands society. They noted that pressures for educational advancement and achievement fall primarily on males. It was suggested that Witkin's (1967) concept of psychological differentiation and the reported male/female differences in field-dependence versus field-independence could be further investigated in Polynesia.

School class group studies were also undertaken by St. George and St. George (1974). The class groups incorporated most but not all subjects from the age groups. Correlations between the QT and six group tests of a primarily 'figural reasoning' nature were derived. High positive correlations were found between the QT and a reasoning series test and space relations test. These results were seen as consistent with the evidence and claims concerning the constructs assessed by the QT.

Stepping outside of the Pacific region we find that Serpell (1974) employed the QT in a study in Zambia. Serpell investigated the relationships between conventional measures of intelligence, including the QT, measures based on games engaged in by Zambian

children (i.e. construction of wire or plasticine models) and estimates of intelligence from the subjects' community based on indigenous and traditional concepts. Serpell found the QT to correlate well with the conventional and games measures and level of education but not with community estimates of intelligence. It was also found that the other conventional and games measures failed to correlate with the community estimates. Serpell argued that the content of the community estimates contained major non-cognitive components such as 'co-operation' and 'obedience'. That is, socially valued qualities that contribute to the social cohesion of closely-knit communities. He argued that these qualities were absent in both conventional measures and games tasks. He also raised important questions as to the social meaning of, say, education if it is premised on a conception of intelligence or attainment totally divergent to the community it is supposed to serve.

In summary then, it can be seen that the QT and related instruments have seen considerable and varied use. The settings have ranged from the Pacific, to Africa and to North America. Within the Pacific there has also been considerable variety in the situations in which the QT has been used. Certainly the cultural backgrounds of subjects have been very diverse indeed. The research itself has also been undertaken with a variety of aims in mind. The studies of Kearney (1966b) and Ord (1970) focused on the development of the QT and related instruments. Those of Bennett (n.d.), Chandra (1975) and St. George and St. George (1974) were largely concerned with the feasibility of applying the QT in new and different cultural contexts. Then again the studies by Richardson (1970), Bennett and Chandra (1974) or Serpell (1974) can be seen as drawing the QT into wider issues concerning the possible constructs underlying the instrument as conditions and cultures vary. Some of these issues will need to be taken up again later in the light of the New Zealand QT research reported in the next chapter.

CHAPTER 6

RESULTS

The results of this research are presented in five parts -

PART 1 reports summary statistics for QT Total Score and QT subtest performances of the two ethnic groups, two gender groups and six age groups. Mean scores, standard deviations, score distributions and inter-correlation analyses are reported and major features of the data discussed.

PART 2 reports separate and combined ethnic group data on QT reliability.

PART 3 reports separate and combined ethnic group data on QT validity.

PART 4 reports data concerning QT items.

PART 5 reports a preliminary set of New Zealand decile rank norms for the QT.

In the application of statistical tests of significance to the data presented the α level has been set at $p < .01$. The research proposals indicate that the intent of these studies is to investigate and report on various psychometric properties of the QT when applied to European and Maori subjects. The studies are exploratory and descriptive rather than experimental. Consequently, a set of preconceived hypotheses specifying anticipated results has not been established for testing in the manner prescribed by a strict experimental 'hypothetico-deductive' framework. Therefore two-tailed (non-directional) tests have been employed.

PART 1 QT SUMMARY STATISTICS, SCORE DISTRIBUTIONS
AND INTERCORRELATION ANALYSES

(a) QT Total Score

The mean QT Total Scores for male and female European and Maori subjects are reported for each of the six age groups in Table 6.1.1. These data were subjected to a three-way analysis of variance (henceforth ANOVA) for unequal cell frequencies by the method presented in Winer (1962) to test for the significance of the effects of gender, age group, ethnic group and their interaction upon QT Total Score performance. The results are summarised in Table 6.1.2. The ANOVA revealed that gender did not have a significant effect on QT Total Score performance but that both age group and ethnic group of subjects did have a significant effect.

Performance on the QT increased with age over the age-range of this study. The increase in the level of performance for the samples across the age range can be observed in Table 6.1.1. The increase is not, however, even across the age range. There is obviously little difference in the performances of subjects aged between 10:6 to 11:5 and 11:6 to 12:5. This is very apparent in Table 6.1.3. In addition, the level of performance drops slightly for the oldest age group.

The ANOVA main effect of ethnic group was found to be significantly related to QT Total Score. This finding is obviously of importance with regards this research and thus warrants further investigation.

In the absence of significant gender or interaction effects male and female data have been combined. Means, standard deviations and differences between means are reported by age group and ethnic group in Table 6.1.3. Inspection of the mean differences at the six age levels shows QT Total Score differences favouring the European samples ranging from 1.14 to 3.54 points and averaging 2.15 points. It was decided that a meaningful statistical evaluation of the apparent QT Total Score

TABLE 6.1.1 Means and Standard Deviations for
QT Total Score by Age, Gender and Ethnic Group

Age Group	European						Maori					
	Male			Female			Male			Female		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	19	36.32	6.89	17	33.18	7.29	19	32.63	6.69	18	30.83	8.15
9:6 - 10:5	28	37.64	6.88	29	38.21	5.62	15	36.27	4.82	18	32.72	4.85
10:6 - 11:5	73	40.55	4.94	52	40.38	5.99	33	37.91	5.37	36	40.61	5.41
11:6 - 12:5	32	40.31	4.25	31	41.52	6.42	29	39.14	4.93	30	39.23	6.47
12:6 - 13:5	38	45.87	4.98	40	43.43	5.34	34	43.82	5.20	26	43.73	4.25
13:6 - 14:5	19	43.68	5.08	19	43.89	6.64	22	42.50	6.12	23	40.91	6.25

TABLE 6.1.2 Summary of ANOVA for QT Total Score

Source	S.S.	df.	M.S.	F	p
Gender (A)	59.08	1	59.08	1.83	N.S.
Age (B)	8311.33	5	1162.27	36.03	< 0.01
Ethnic Group (C)	668.08	1	668.08	20.71	< 0.01
AB	241.02	5	48.20	1.49	N.S.
AC	0.21	1	0.21	0.01	N.S.
BC	113.89	5	22.78	0.71	N.S.
ABC	9.90	5	1.98	0.06	N.S.
Within Cell (Error)	21810.23	676	32.26		
TOTAL	31213.74	699			

TABLE 6.1.3 Means, Standard Deviations and Mean Score
Differences for QT Total Score by Age and Ethnic Group

Age Group	European			Maori			Mean Score Difference E > M
	N	Mean	S.D.	N	Mean	S.D.	
8:6 - 9:5	36	34.83	7.16	37	31.76	7.38	3.07
9:6 - 10:5	57	37.93	6.22	33	34.39	5.13	3.54
10:6 - 11:5	125	40.48	5.38	69	39.32	5.52	1.16
11:6 - 12:5	63	40.90	5.41	59	39.03	6.02	1.87
12:6 - 13:5	78	44.62	5.28	60	43.48	4.77	1.14
13:6 - 14:5	38	43.79	5.83	45	41.69	6.17	2.10

differences could be made through the regression of QT scores upon age for the European and Maori samples and the subsequent testing of the significance of difference between the respective regression lines.

Post hoc multiple comparison methods were considered, but deemed unsuitable. Methods based on the studentized range (e.g. Newman-Keuls) required equal sample sizes. Scheffé's post hoc t presented problems of both a logical and probabilistic nature. The choice of a small set of contrasts from all possible contrasts would require logical justification and the probability of obtaining a significant difference amongst a set of contrasts by chance alone could not be discounted. It was thought preferable to adopt an analysis which would assist in determining the magnitude and significance of any relative and constant difference in performance.

The regression equations for QT Total Score on age and corresponding correlation coefficients for European and Maori samples are reported in Table 6.1.4. The linear regression lines are plotted in Figure 6.1. The functions appear to be almost identical. Ferguson (1971) has described methods for determining the homogeneity of regression coefficients. First, the proposition that the slopes of the regression lines were the same was tested. The test indicated that the proposition that the relationship between QT Total Score and age for European and Maori subjects was the same could not be rejected ($F_{1,696} = 0.14$, N.S.). Thus, the slopes of the regression lines may be deemed parallel.

Second, the proposition that the regression lines fit a single common slope was tested. This test found that a single common slope could not adequately describe that data ($F_{1,696} = 12.39$, $p < .01$). The regression lines while parallel were found to be separated by a small but constant difference of 2.15 QT raw score points.

These findings are in accord with the significant ethnic group effect and non-significant age-ethnic group effect in the ANOVA. In

FIGURE 6.1 Regression of QT Score on Age for European (N = 397)
and Maori Samples (N = 303)

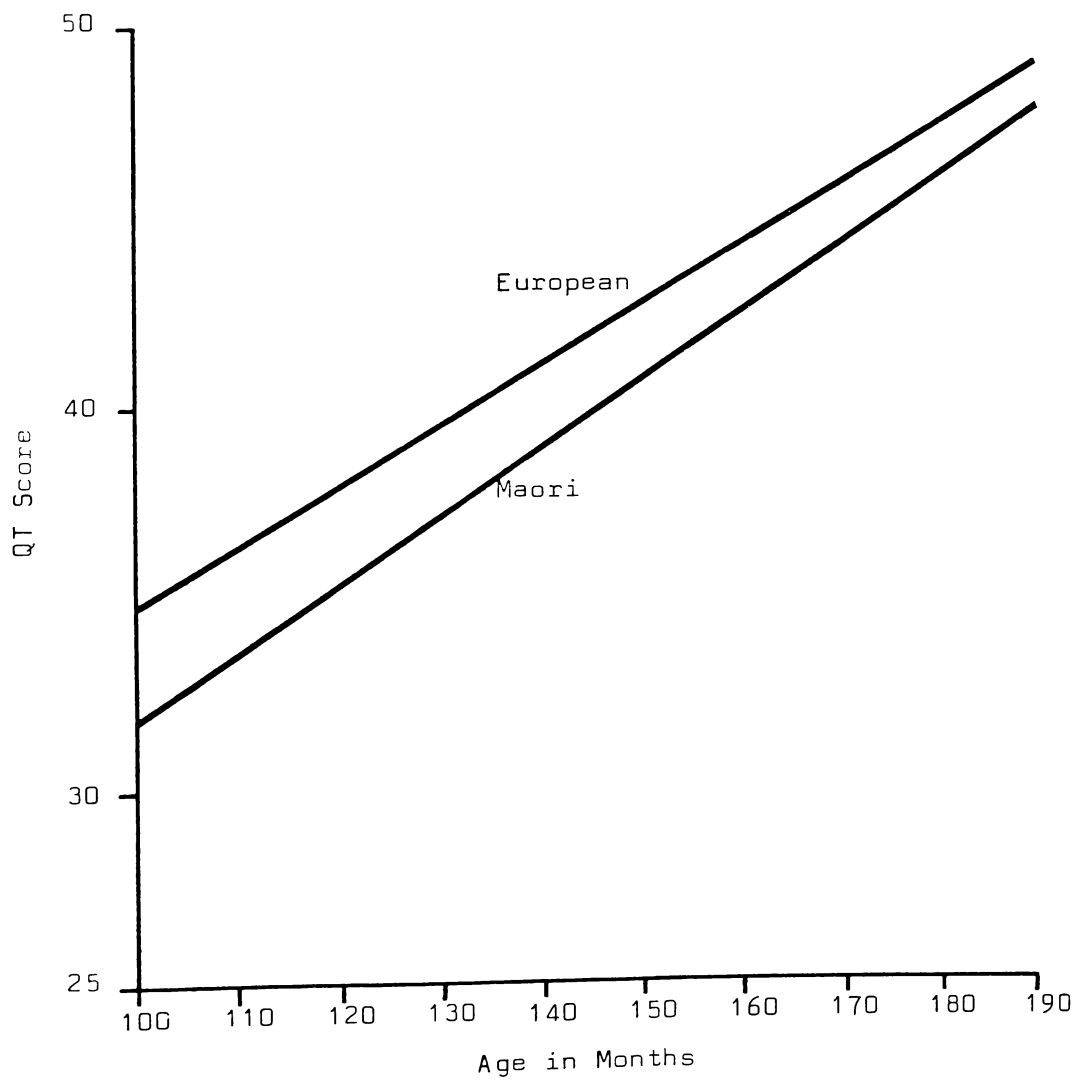


TABLE 6.1.4 Linear Regression Equations and Correlation Coefficients
For QT Total Score on Age for European and Maori Samples

European:

$$y = 0.16x + 19.19$$

$$r = 0.425, p < .01$$

$$\bar{x} = 137.74$$

$$\bar{y} = 40.80$$

$$N = 397$$

Maori:

$$y = 0.18x + 13.79$$

$$r = 0.496, p < .01$$

$$\bar{x} = 139.92$$

$$\bar{y} = 38.98$$

$$N = 303$$

Where y = linear regression equation, r = correlation between QT Total Score and age, \bar{x} = mean age, \bar{y} = mean QT Total Score and N = sample size.

addition, the constant difference estimated in the test for a single common regression equation is identical with the averaged mean score differences for the six groups - namely 2.15 QT raw score points.

These analyses indicate a small and relatively stable difference in QT performance age for age between European and Maori subjects. Questions concerning the practical significance of this ethnic group difference and the extent of its contribution to variation in QT performance, relative to other factors, were examined by multiple regression methods as described in Kerlinger and Pedhazur (1973).

Through ANOVA procedures three possible sources of variance in QT Total Score performance were investigated. These were sex, age and ethnicity. Further partitioning of the data would have reduced cell sizes and created considerably more complex interaction terms. Multiple regression enables an analysis to be made of the collective and separate contributions of an increased number of independent variables to variation in the dependent variable of QT performance. In addition, the relative importance of the variables as 'explanations' of the variance

could be ascertained.

The following independent variables were entered into the multiple regression analysis - age, socio-economic status, gender, urban or rural location and ethnicity. These variables have been described in Chapter 4.

When describing the variables of socio-economic status and urban or rural location it was noted that such data were collected primarily for the purpose of describing the samples of European and Maori subjects. It was argued that the method of determining socio-economic status may not be particularly appropriate in the case of Maori subjects. It was also noted that the urban - rural distinction was made through a simple classification of the communities in which the schools participating in the study were located. Nevertheless, it could still be argued that the variation in QT performance might be related to socio-economic status and urban - rural differences as well as the factors of age, gender and ethnicity which were entered into the ANOVA. Multiple regression methods could assist in providing a test of these contentions.

The matrix of correlation coefficients between QT Total Score and the five independent variables are reported in Table 6.1.5. The multiple regression analysis is reported in Table 6.1.6. In evaluating the contribution and relative importance of the independent variables attention is directed to the columns in Table 6.1.6 titled 'Zero Order Variance Accounted For' and 'Variance Accounted For When Entered Last'. The first reports the squared zero order correlation coefficient which indicates the proportion of variance accounted for in the criterion variable by each of the independent variables without consideration of any interactions with other independent variables. The second reports the amount of variance accounted for by the independent variable after interactions with other independent variables have been semi-partialled out of the regression analysis. It indicates the unique contribution

TABLE 6.1.5 Correlations Between QT Total Score and the Five Independent Variables Entered in the Multiple Regression Analysis (N = 673)¹

	1	2	3	4	5	6
1. QT Total Score						
2. Age	.45					
3. SES	-.15	-.03				
4. Gender	-.03	.02	.05			
5. Urban/Rural	.14	.35	.00	.03		
6. Ethnicity	.15	-.05	-.35	.02	-.08	

All values of .12 and above significant:
p < .01, two-tailed test.

¹ All correlations are based on a sample of 673. Twenty-seven subjects were removed from the analysis because of missing socio-economic status data.

TABLE 6.1.6 Multiple Regression Analysis - QT Total Score

Criterion - Queensland Test Total Score
Multiple Correlation Coefficient (R) = .49
R² = .24
F Ratio = 41.36 (p < .01)

Variable	Beta	F to delete	Zero Order Variance Accounted For	Variance ¹ Accounted For When Entered Last	p
Age	.45	156.11	.20	.18	<.01
SES	-.09	6.21	.02	.01	N.S.
Gender	-.02	0.54	.00	.00	N.S.
Urban/Rural	.00	0.00	.02	.00	N.S.
Ethnicity	.14	15.05	.02	.02	<.01

¹ "Variance accounted for when variable X is entered last can be calculated from the sum of squares of the variable when entered last (SS_{x_i}), divided by the total sum of squares (SS_y). SS_{x_i} can be calculated by multiplying F to delete (F_{del}) from the last step in a stepwise analysis, by the mean square of the residual (σ²).

Hence:
variance accounted for when entered last = $\frac{F_{del} \cdot \sigma^2}{SS_y}$

This is equivalent to R² with X entered last, minus R² without X, from the previous step (Kerlinger and Pedhazur, 1973, pp.290-295)." (from Harker, 1976, p.90).

made by each independent variable after controlling for any indirect effects made through the covariance of one independent variable with other independent variables. Inspection of Table 6.1.6 shows that the variables of socio-economic status, gender and urban or rural location failed to account for any unique variance in QT Total Score at the required significance level. Thus the small but significant zero order correlations between QT Total Score and socio-economic status and QT Total Score and urban or rural location are misleading. When account is taken of other intervening variables the effect is to reduce the variance accounted for directly by each of these variables to statistically non-significant proportions.

The variables of age and ethnicity were found however, to make statistically significant direct contributions to the variance in QT Total Score. Age was found to directly account for eighteen percent of the variance and ethnicity two percent. As a result, eighty percent of the variance in the data can be attributed to differences in ability as assessed by the QT. Although specification error cannot be overlooked in this type of analysis, the results from the multiple regression analysis are in accord with previous findings.

Both age and ethnicity were found to have a significant effect on QT performance while gender of subject does not. In terms of the relative importance of the direct contributions of the effects of age and ethnicity however, it is apparent that while age is of considerable importance the influence of ethnicity is comparatively minor. The regression of QT Total Score on age for the European and Maori samples demonstrates this point. Both slopes were influenced by age in the same manner and the ethnic group difference in the slopes of 2.15 points is between a half to a third of the mean score standard deviations for the European and Maori samples in each age group. These latter data in themselves point to considerable overlap in score distributions.

The data suggests then that while age has a strong influence upon QT performance, ethnicity (being European or Maori), although statistically significant, is of little practical importance in determining QT performance. Likewise, QT performance has been found to be independent of any direct influences from the variables of socio-economic status, gender and urban or rural location.

It might be noted that if an alpha level of $p < .05$ had been acceptable the variable of socio-economic status which directly accounted for one percent of the variance in QT Total Score would have been considered statistically significant. As a check a further multiple regression analysis was undertaken without the redundant variables of gender and urban or rural location. Age accounted for two percent more unique variance while socio-economic status and ethnicity accounted for the same proportions of unique variance as before. The levels of statistical significance attained by the three variables were the same.

(b) QT Subtests

The five subtests of the QT have been examined through ANOVA procedures for the effects of age, gender and ethnicity on performance. Since the QT is viewed by its authors as a single measure rather than a battery, and given that the primary concern of this research is with the QT in that form, analysis and discussion of subtest performances has been deliberately limited.

Knox Subtest

Table 6.1.7 presents the mean Knox subtest performance for male and female, European and Maori subjects in each age group. The data were subjected to a three-way ANOVA and the results of the ANOVA are summarised in Table 6.1.8. The ANOVA shows that neither gender of subject nor ethnic group had a

significant effect upon Knox subtest performance. In keeping with the significant effect of age upon QT Total Score performance the Knox subtest performance is also affected by age. Since the ethnic group comparative data may be of interest it is presented in Table 6.1.9. Combined ethnic group data is also presented in the same Table. The trend of the age group data is for performance to increase with age. The increases are however, neither dramatic nor constant.

TABLE 6.1.7 Means and Standard Deviations for Knox Subtest by Age, Gender and Ethnic Group

Age Group	European				Maori			
	Male		Female		Male		Female	
	N	Mean S.D.	N	Mean S.D.	N	Mean S.D.	N	Mean S.D.
8:6 - 9:5	19	6.89 2.51	17	6.00 2.26	19	6.68 2.19	18	5.89 1.97
9:6 - 10:5	28	7.25 2.05	29	7.41 1.80	15	7.07 1.44	18	6.94 2.21
10:6 - 11:5	73	7.55 2.19	52	7.96 1.98	33	7.79 2.07	36	8.94 2.07
11:6 - 12:5	32	7.31 2.10	31	8.03 1.82	29	7.79 1.97	30	7.93 1.89
12:6 - 13:5	38	8.63 2.11	40	8.03 2.22	34	8.44 2.50	26	8.81 1.58
13:6 - 14:5	19	8.00 2.29	19	7.63 2.31	22	7.59 2.67	23	8.22 2.33

TABLE 6.1.8 Summary of ANOVA for Knox Subtest

Source	S.S.	df.	M.S.	F	p
Gender (A)	0.69	1	0.69	0.15	N.S.
Age (B)	281.94	5	56.39	12.53	< 0.01
Ethnic Group (C)	2.07	1	2.07	0.46	N.S.
AB	38.52	5	7.70	1.71	N.S.
AC	3.98	1	3.98	0.88	N.S.
BC	13.50	5	2.70	0.60	N.S.
ABC	15.14	5	3.03	0.67	N.S.
Within Cell (Error)	3040.60	676	4.50		
TOTAL	3396.44	699			

TABLE 6.1.9 Means and Standard Deviations for Knox Subtest
by Age for Separate and Combined Ethnic Group Samples

Age Group	European			Maori			Total Sample		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	36	6.47	2.41	37	6.32	2.08	73	6.40	2.23
9:6 - 10:5	57	7.33	1.91	33	7.00	1.87	90	7.21	1.89
10:6 - 11:5	125	7.72	2.10	69	8.39	2.14	194	7.96	2.13
11:6 - 12:5	63	7.67	1.93	59	7.86	1.92	122	7.76	1.95
12:6 - 13:5	78	8.32	2.18	60	8.60	2.14	138	8.44	2.16
13:6 - 14:5	38	7.82	2.28	45	7.91	2.49	83	7.87	2.38

Beads Subtest

Table 6.1.10 presents the mean Beads subtest performances for male and female, European and Maori subjects in each age group. The results of a three-way ANOVA on these data are summarised in Table 6.1.11. Neither gender of subject nor ethnic group had a significant effect upon Beads subtest performance. Again, performance was found to be significantly affected by age. Data on the separate and combined ethnic group performances by age group are presented in Table 6.1.12. The data show that performance on the Beads subtest increases with age, but as with the Knox subtest the trend is not even across the age groups.

TABLE 6.1.10 Means and Standard Deviations for Beads Subtest
by Age, Gender and Ethnic Group

Age Group	European						Maori					
	Male			Female			Male			Female		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	19	3.32	2.16	17	3.35	2.21	19	3.32	1.53	18	3.17	2.13
9:6 - 10:5	28	3.61	2.31	29	4.17	2.19	15	3.20	1.97	18	2.83	1.98
10:6 - 11:5	73	4.93	2.34	52	4.63	2.19	33	3.48	2.45	36	4.86	2.66
11:6 - 12:5	32	4.22	2.18	31	5.32	2.68	29	4.00	2.05	30	4.63	2.06
12:6 - 13:5	38	6.42	2.24	40	5.98	2.35	34	5.97	1.93	26	6.42	1.88
13:6 - 14:5	19	5.37	2.71	19	6.68	2.36	22	6.00	1.98	23	5.17	2.42

TABLE 6.1.11 Summary of ANOVA for Beads Subtest

Source	S.S.	df.	M.S.	F	p
Gender (A)	12.21	1	12.21	2.41	N.S.
Age (B)	715.74	5	143.15	28.29	< 0.01
Ethnic Group (C)	25.79	1	25.79	5.10	N.S.
AB	11.69	5	2.34	0.46	N.S.
AC	1.84	1	1.84	0.36	N.S.
BC	13.30	5	2.66	0.53	N.S.
ABC	61.66	5	12.33	2.44	N.S.
Within Cell (Error)	3418.17	676	5.06		
TOTAL	4260.40	699			

TABLE 6.1.12 Means and Standard Deviations for Beads Subtest by Age for Separate and Combined Ethnic Group Samples

Age Group	European			Maori			Total Sample		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	36	3.33	2.15	37	3.24	1.82	73	3.29	1.98
9:6 - 10:5	57	3.89	2.25	33	3.00	1.95	90	3.57	2.18
10:6 - 11:5	125	4.81	2.27	69	4.20	2.64	194	4.59	2.42
11:6 - 12:5	63	4.76	2.48	59	4.32	2.06	122	4.55	2.29
12:6 - 13:5	78	6.19	2.29	60	6.17	1.91	138	6.18	2.12
13:6 - 14:5	38	6.03	2.59	45	5.58	2.23	83	5.78	2.40

Passalong Subtest

Table 6.1.13 presents the mean Passalong subtest performances for male and female, European and Maori subjects in each age group. The data were subjected to a three-way ANOVA and the results are summarised in Table 6.1.14. ANOVA shows that neither gender of subject nor ethnic group had a significant effect upon Passalong subtest performance. Age however, had a significant effect on performance. Data on the separate and combined ethnic group performances by age group are presented in Table 6.1.15.

Inspection of the data shows that performance appears to increase only marginally with age which no doubt reflects the narrow score range on a subtest of seven items.

Two significant interaction effects were also obtained in the analysis of the Passalong subtest data. First, it was found that performance on the subtest was significantly affected by the interacting effects of gender and ethnic group. Table 6.1.16 presents the mean performance data for the male and female samples for each ethnic group. It can be seen that while European males and females obtain higher mean scores than Maori males and females, the male samples for each ethnic group achieved slightly higher mean scores than the female samples.

The second interaction effect on the Passalong subtest concerned the effects of age and ethnic group on performance. Inspection of Table 6.1.15 shows that with the exception of the age group 8:6 to 9:5, European subjects obtained higher mean scores than Maori subjects. The data suggest that while there are no differences in the performances of the younger subjects on this subtest, increasing age enhances slightly the performances of European subjects over Maori subjects.

TABLE 6.1.13 Means and Standard Deviations for Passalong Subtest by Age, Gender and Ethnic Group

Age Group	European						Maori					
	Male			Female			Male			Female		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	19	3.89	1.29	17	4.06	1.43	19	4.16	1.50	18	3.89	1.32
9:6 - 10:5	28	4.68	1.22	29	4.41	1.45	15	3.80	1.26	18	3.94	1.26
10:6 - 11:5	73	4.82	1.34	52	4.98	1.24	33	4.61	1.22	36	4.50	1.54
11:6 - 12:5	32	5.13	1.10	31	5.42	1.15	29	4.76	1.33	30	4.63	1.45
12:6 - 13:5	38	5.79	0.87	40	5.33	0.92	34	5.03	1.19	26	5.08	1.23
13:6 - 14:5	19	5.68	0.95	19	5.16	1.61	22	5.23	1.45	23	4.61	1.41

TABLE 6.1.14 Summary of ANOVA for Passalong Subtest

Source	S.S.	df.	M.S.	F	p
Gender (A)	3.29	1	3.29	2.01	N.S.
Age (B)	136.99	5	27.40	16.71	< 0.01
Ethnic Group (C)	8.75	1	8.75	5.34	N.S.
AB	9.63	5	1.93	1.18	N.S.
AC	21.38	1	21.38	13.04	< 0.01
BC	29.97	5	5.99	3.65	< 0.01
ABC	5.75	5	1.15	0.70	N.S.
Within Cell (Error)	1107.29	676	1.64		
TOTAL	1323.05	699			

TABLE 6.1.15 Means and Standard Deviations for Passalong Subtest by Age for Separate and Combined Ethnic Group Samples

Age Group	European			Maori			Total Sample		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	36	4.00	1.35	37	4.03	1.40	73	4.01	1.37
9:6 - 10:5	57	4.54	1.34	33	3.88	1.24	90	4.30	1.34
10:6 - 11:5	125	4.89	1.30	69	4.55	1.39	194	4.77	1.34
11:6 - 12:5	63	5.27	1.11	59	4.69	1.38	122	4.99	1.28
12:6 - 13:5	78	5.55	0.92	60	5.05	1.20	138	5.33	1.08
13:6 - 14:5	38	5.42	1.33	45	4.91	1.44	83	5.14	1.41

TABLE 6.1.16 Means and Standard Deviations for Passalong Subtest by Gender and Ethnic Group

Gender	European			Maori		
	N	Mean	S.D.	N	Mean	S.D.
Male	209	5.02	1.28	152	4.68	1.36
Female	188	4.97	1.32	151	4.50	1.42

Form Assembly Subtest

Table 6.1.17 presents the mean Form Assembly subtest performances for male and female, European and Maori subjects in each age group. The data were subjected to a three-way ANOVA and the results are summarised in Table 6.1.18. The ANOVA shows that while gender did not have a significant effect on performance both age group and ethnic group did. Data on ethnic group performances by age group are presented in Table 6.1.19. Inspection of this table shows that mean performance increases with age for both ethnic groups for five age levels but that there is a small decline in the level of performance at the oldest age level. It can also be seen from Table 6.1.19 that in each age group European subjects obtained slightly higher mean scores than Maori subjects.

TABLE 6.1.17 Means and Standard Deviations for Form Assembly Subtest by Age, Gender and Ethnic Group

Age Group	European						Maori					
	Male			Female			Male		Female			
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	19	9.11	1.82	17	8.41	1.75	19	8.05	1.78	18	8.17	1.98
9:6 - 10:5	28	8.96	1.77	29	9.00	1.67	15	9.40	1.50	18	8.33	1.24
10:6 - 11:5	73	9.27	1.38	52	9.60	1.19	33	8.91	1.68	36	9.25	1.59
11:6 - 12:5	32	9.66	1.07	31	9.19	1.64	29	9.34	1.49	30	8.97	2.13
12:6 - 13:5	38	10.37	1.40	40	9.83	1.36	34	9.71	1.47	26	9.62	1.47
13:6 - 14:5	19	10.11	1.70	19	9.95	1.72	22	9.59	1.79	23	9.26	1.79

TABLE 6.1.18 Summary of ANOVA for Form Assembly Subtest

Source	S.S.	df.	M.S.	F	p
Gender (A)	8.95	1	8.95	3.64	N.S.
Age (B)	141.49	5	28.30	11.50	< 0.01
Ethnic Group (C)	24.86	1	24.86	10.11	< 0.01
AB	2.86	5	0.57	0.23	N.S.
AC	0.01	1	0.01	0.00	N.S.
BC	5.15	5	1.03	0.42	N.S.
ABC	21.72	5	4.34	1.76	N.S.
Within Cell (Error)	1663.33	676	2.46		
TOTAL	1868.37	699			

TABLE 6.1.19 Means and Standard Deviations for Form Assembly Subtest by Age and Ethnic Group

Age Group	European			Maori		
	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	36	8.86	1.68	37	8.11	1.85
9:6 - 10:5	57	8.98	1.71	33	8.82	1.45
10:6 - 11:5	125	9.41	1.31	69	9.09	1.63
11:6 - 12:5	63	9.44	1.40	59	9.15	1.84
12:6 - 13:5	78	10.09	1.40	60	9.67	1.46
13:6 - 14:5	38	10.03	1.68	45	9.42	1.78

Pattern Matching Subtest

Table 6.1.20 presents the mean Pattern Matching subtest performances for male and female, European and Maori subjects in each age group. These data were subjected to a three-way ANOVA which is summarized in Table 6.1.21. The ANOVA indicates that gender, age group and ethnic group of subjects all had a significant effect on performance. In addition, age and ethnic group were found to interact significantly with Pattern Matching subtest performance.

In relation to the gender effect, inspection of Table 6.1.20 suggests that irrespective of ethnic group, males obtained higher scores than females at each age group level. The effect of age on performance can also be observed in Table 6.1.20 in that the level of performance increased with age for both sexes and for European and Maori subjects.

The ethnic group effect is evidenced through the European subjects at each age level obtaining higher mean scores than their Maori counterparts. This trend can be observed more clearly in Table 6.1.22 where the data have been re-cast disregarding gender. In addition, it can be observed here that the differences in ethnic group mean scores tend to diminish as age level increases. This probably accounts for the significant age group and ethnic group interaction term.

TABLE 6.1.20 Means and Standard Deviations for Pattern Matching Subtest by Age, Gender and Ethnic Group

Age Group	European				Maori			
	Male		Female		Male		Female	
	N	Mean S.D.	N	Mean S.D.	N	Mean S.D.	N	Mean S.D.
8:6 - 9:5	19	13.11 2.79	17	11.35 2.87	19	10.42 3.56	18	9.72 3.16
9:6 - 10:5	28	13.14 1.96	29	13.21 1.66	15	12.80 1.97	18	10.67 2.47
10:6 - 11:5	73	13.97 1.12	52	13.21 2.15	33	13.12 1.71	36	13.06 2.07
11:6 - 12:5	32	13.97 1.28	31	13.55 1.79	29	13.24 1.62	30	13.07 2.63
12:6 - 13:5	38	14.66 0.67	40	14.28 1.22	34	14.18 1.17	26	13.77 1.45
13:6 - 14:5	19	14.53 0.84	19	14.47 0.84	22	14.09 1.66	23	13.65 1.75

TABLE 6.1.21 Summary of ANOVA for Pattern Matching Subtest

Source	S.S.	df.	M.S.	F	p
Gender (A)	55.12	1	55.12	16.07	< 0.01
Age (B)	692.42	5	138.48	40.37	< 0.01
Ethnic Group (C)	143.57	1	143.57	41.86	0.01
AB	22.15	5	4.43	1.29	N.S.
AC	0.40	1	0.40	0.12	N.S.
BC	58.99	5	11.80	3.44	< 0.01
ABC	41.61	5	8.32	2.43	N.S.
Within Cell (Error)	2320.94	676	3.43		
TOTAL	3335.20	699			

TABLE 6.1.22 Means and Standard Deviations for Pattern Matching Subtest by Age and Ethnic Group

Age Group	European			Maori		
	N	Mean	S.D.	N	Mean	S.D.
8:6 - 9:5	36	12.28	2.92	37	10.08	3.34
9:6 - 10:5	57	13.18	1.79	33	11.64	2.47
10:6 - 11:5	125	13.66	1.67	69	13.09	1.89
11:6 - 12:5	63	13.76	1.55	59	13.17	2.17
12:6 - 13:5	78	14.46	1.00	60	14.00	1.30
13:6 - 14:5	38	14.50	0.83	45	13.87	1.70

(c) Summary Figures and Comments

Figures 6.2 to 6.7 illustrate the mean score performances of the separate and combined ethnic groups at the six age levels for QT Total Score and the five QT subtests. The Figures highlight a number of features of the data, some of which were evident from the preceding statistical analyses.

Figure 6.2 shows both the increments in QT Total Score with age and the small but rather stable ethnic group difference. Inspection of the figures for each subtest shows a very similar pattern of performance in relation to age for the two ethnic groups. The slightly higher

FIGURE 6.2 Mean Scores for Each Total Sample, European
and Maori Age Group on the QT

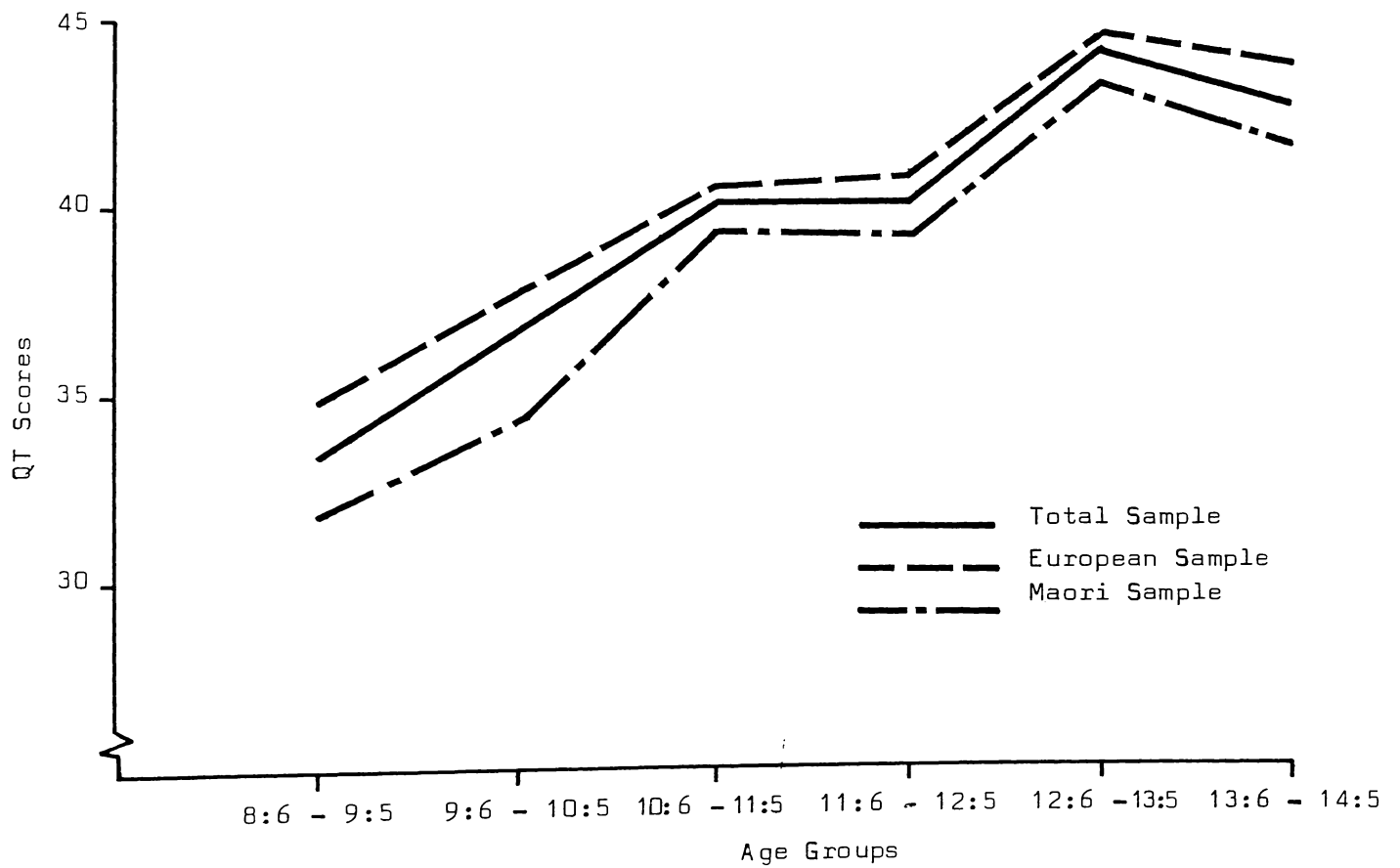


FIGURE 6.3 Mean Scores for Each Total Sample, European and Maori Age Group on the Knox Subtest

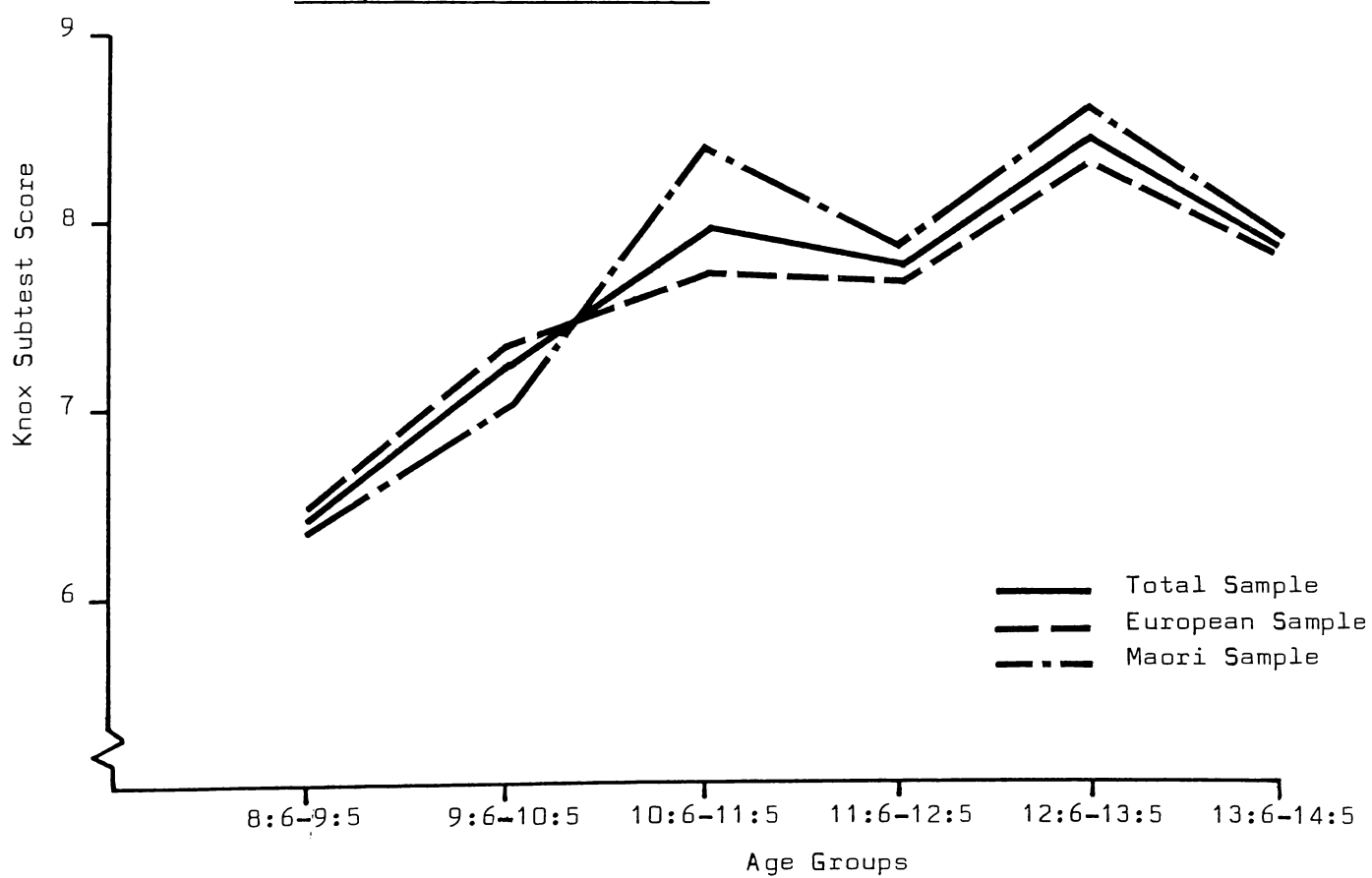


FIGURE 6.4

Mean Scores for Each Total Sample, European and Maori Age Group on the Beads Subtest

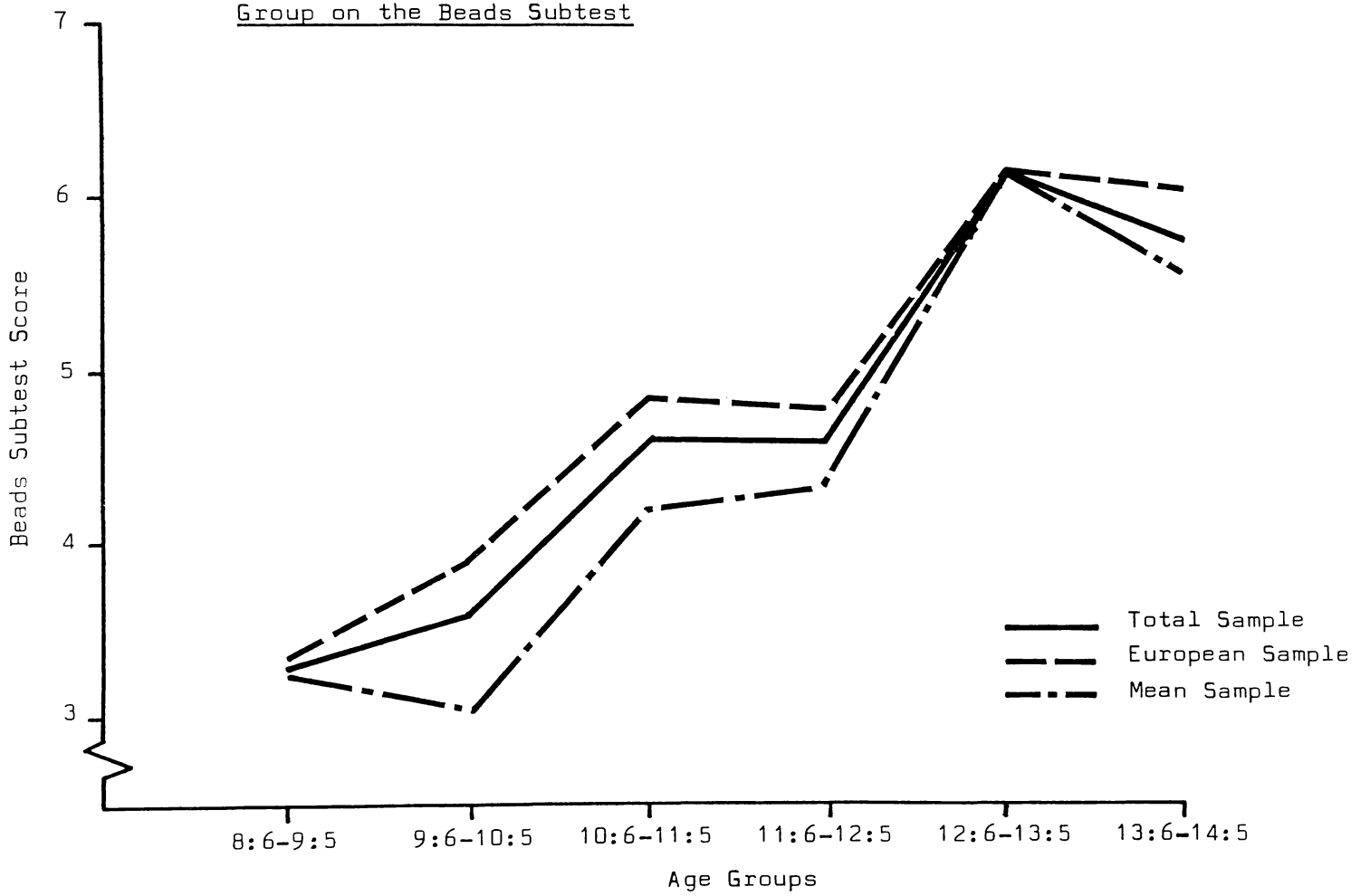


FIGURE 6.5 Mean Scores for Each Total Sample, European and Maori Age Group on the Passalong Subtest

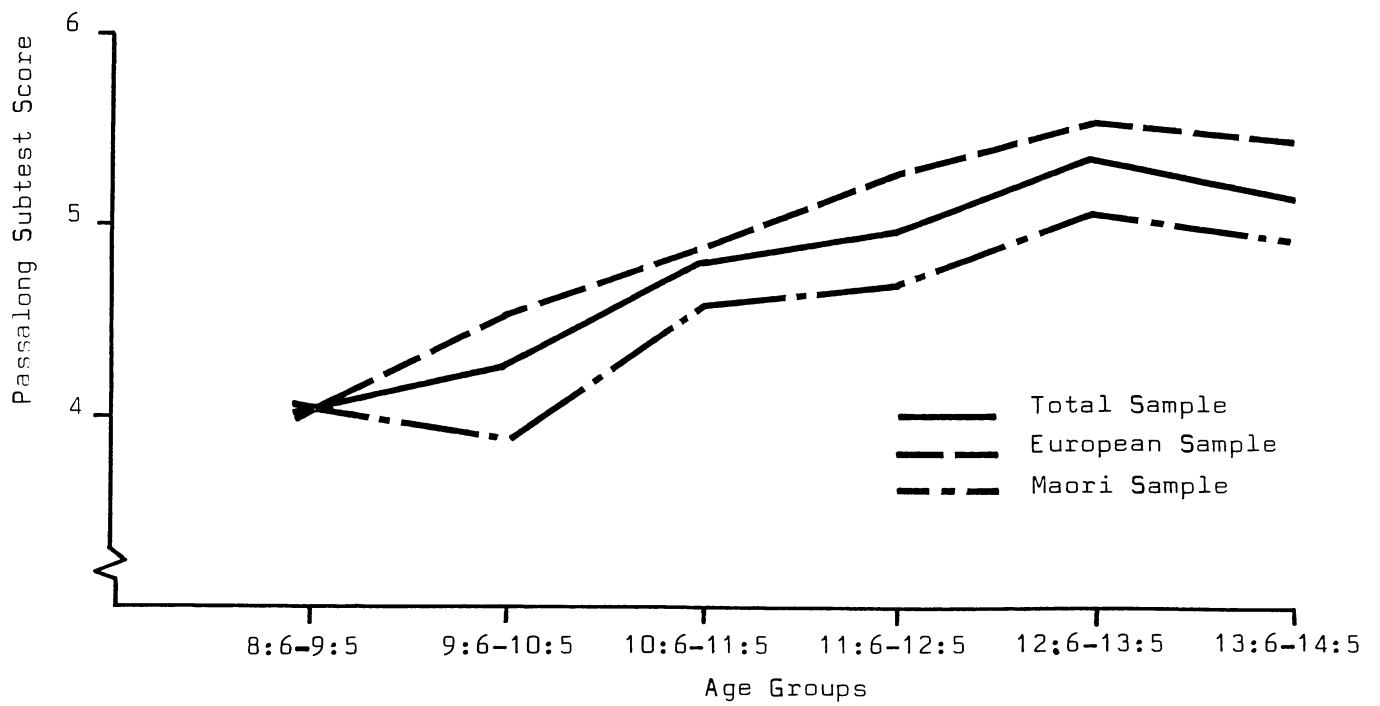


FIGURE 6.6 Mean Scores for Each Total Sample, European and Maori Age Group on the Form Assembly Subtest

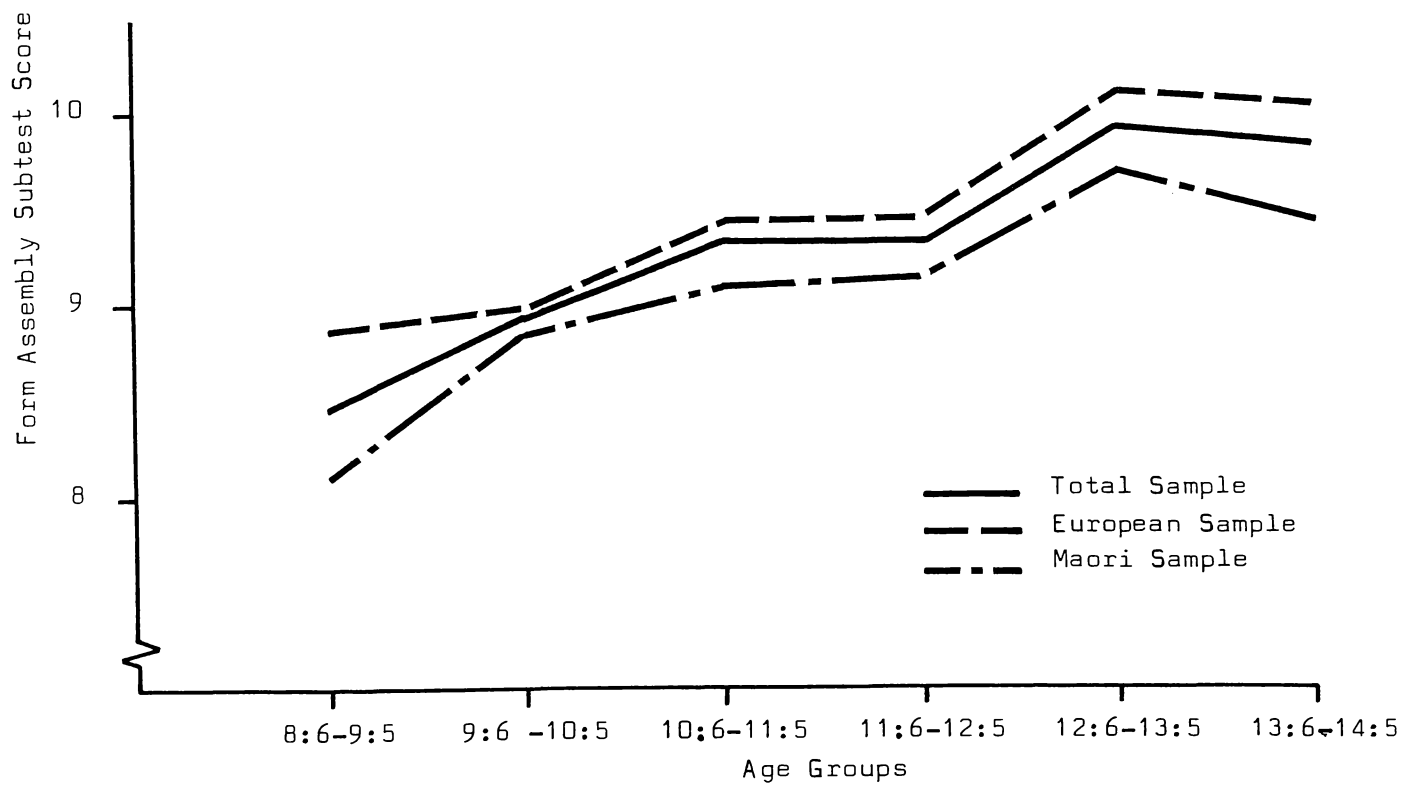
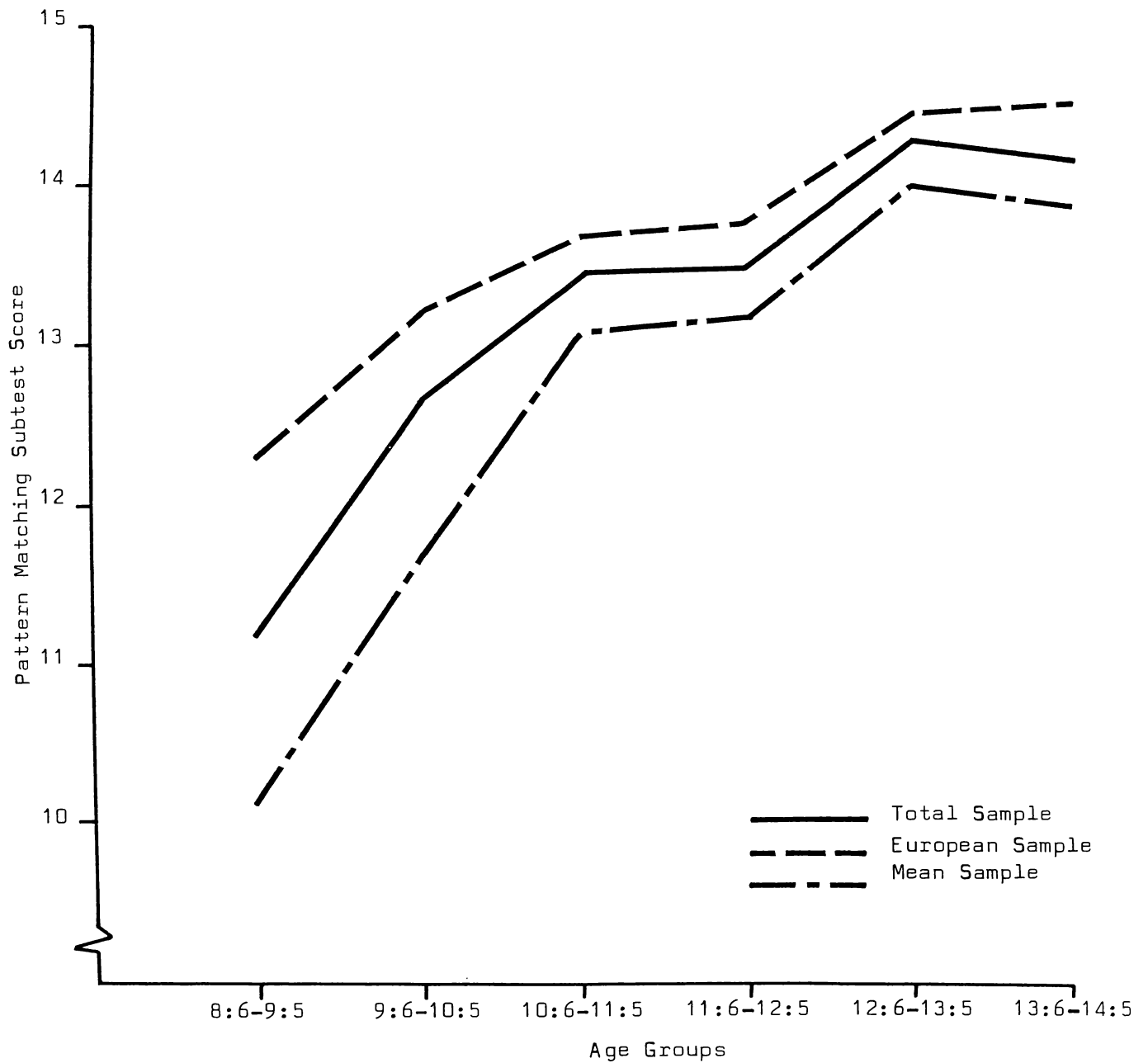


FIGURE 6.7 Mean Scores for Each Total Sample, European and Maori Age Group on the Pattern Matching Subtest



performance of European subjects on most subtests at each age level can also be observed. An exception is the Knox subtest where initially European subjects obtained higher mean scores, but for the age group 10:6 - 11:5 and over higher mean scores were obtained by the Maori samples. This particular interaction was, however, non-significant in the ANOVA.

While it has been noted that QT performance increases with age it is apparent from the figures that the increments are not constant across the age range. There is very little difference in the QT Total Score mean performances between the ages 10:6 - 11:5 and 11:6 - 12:5. A result which is also by and large evident in the five subtests. In addition, the level of performance of the oldest age group declines somewhat in relation to the previous age group. Both phenomena will be given consideration when drawing conclusions from the research.

(d) Score Distributions

Data on the distribution of QT Total Scores and QT subtest scores provide another perspective to the comparison of European and Maori performances. This data for the combined ethnic group samples are also presented.

QT Total Scores obtained by the separate and combined ethnic groups by both numbers and proportions are presented in Tables 6.1.23 to 6.1.25. The corresponding frequency distributions based on numbers of subjects are presented in Figures 6.8 to 6.10.

The QT Total Score distributions for the European and Maori samples would appear by inspection to be of the same general form. Both distributions cover an almost identical QT Total Score range and both evidence a degree of negative skew.

The distribution for the combined ethnic group sample presented in Figure 6.10 replicates features of the European and Maori distributions. The negative skew is evident although the distribution tends

FIGURE 6.8 Frequency Distribution of QT Scores for European Sample (N = 397)

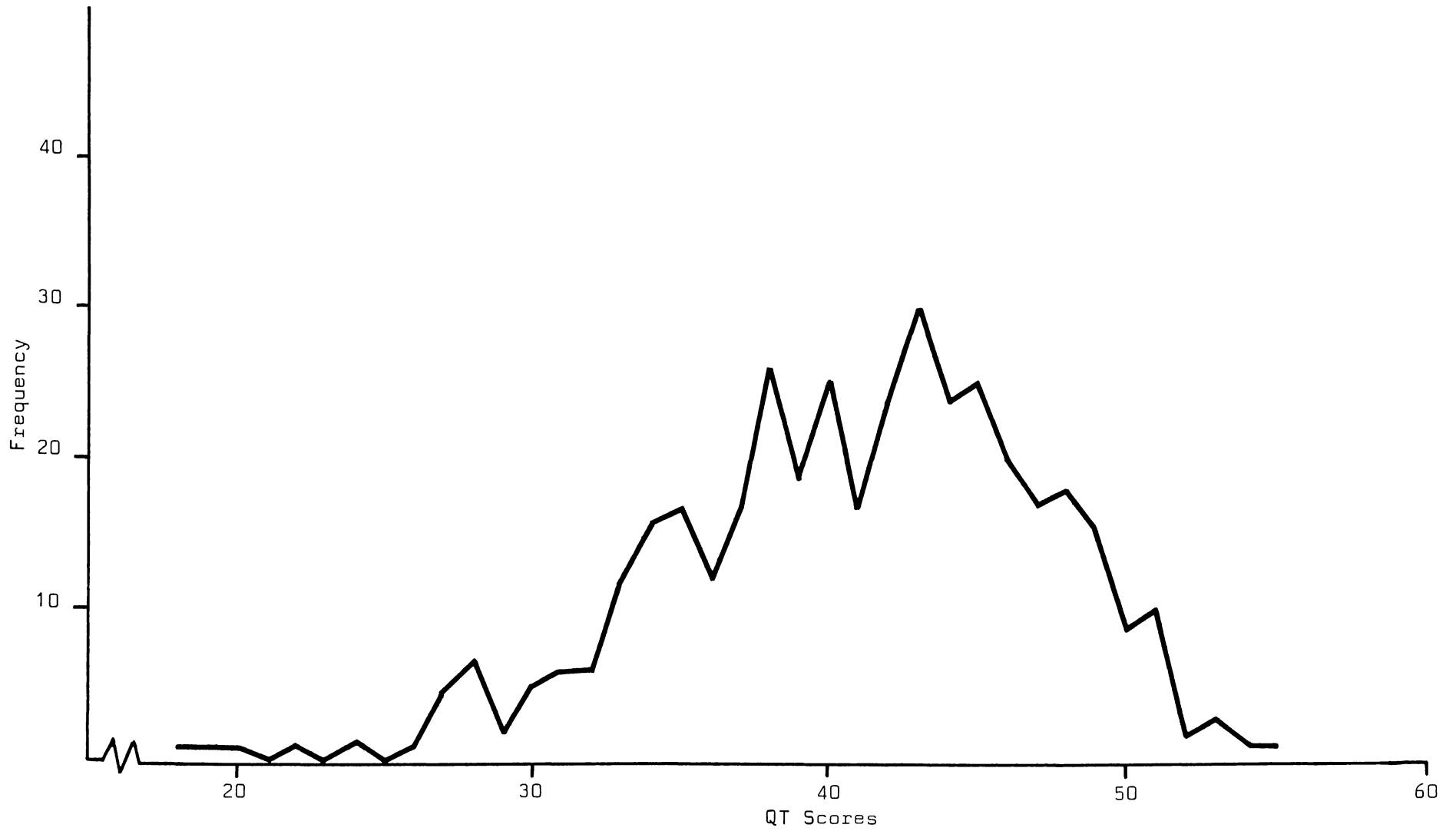


FIGURE 6.9

Frequency Distribution of QT Scores for Maori Sample (N = 303)

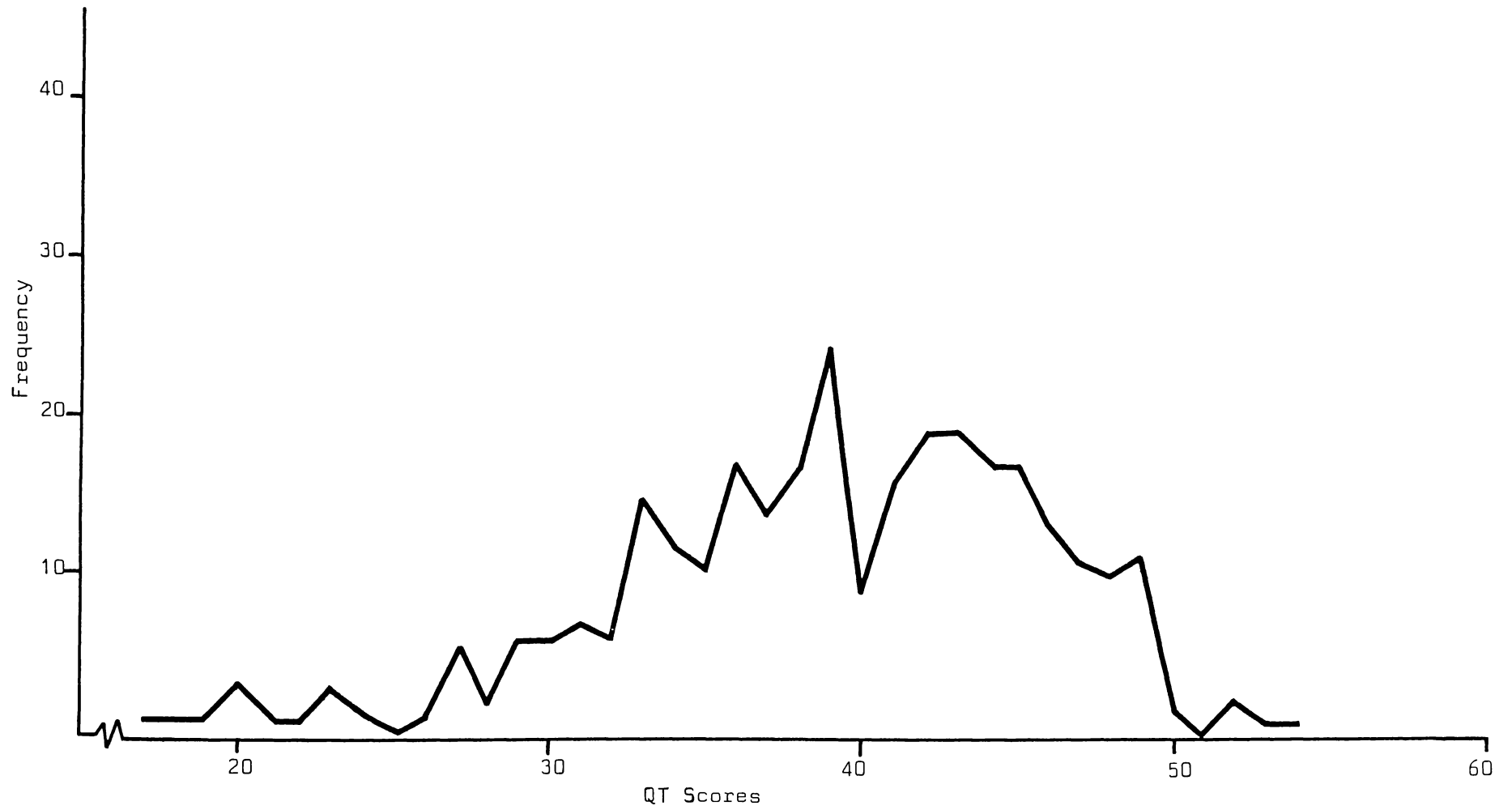
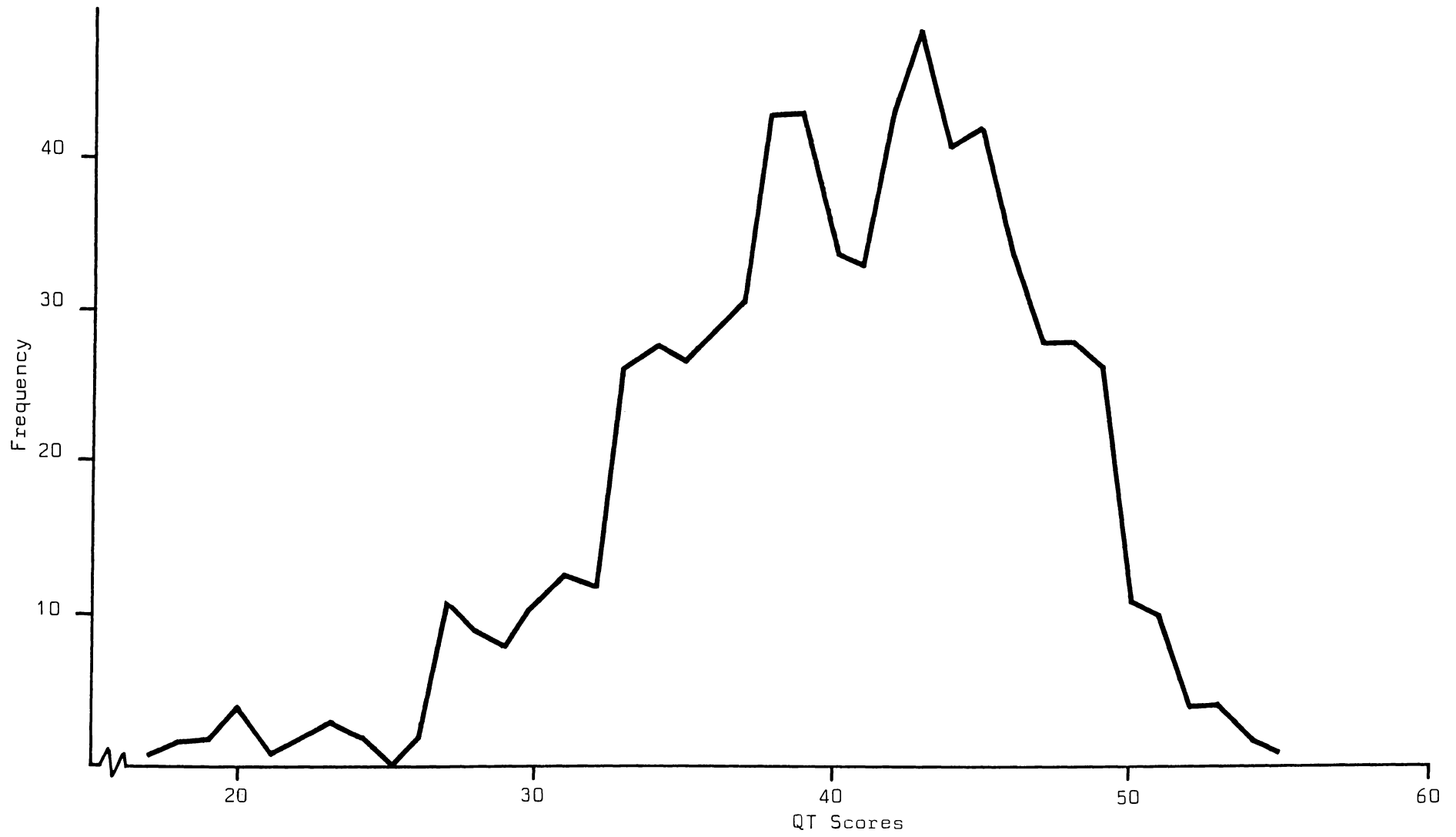


FIGURE 6.10

Frequency Distribution of QT Scores for Total Sample (N = 700)



towards that of the normal distribution curve. Distributions of this nature are, in fact, to be expected on ability measures according to Burt (1963). Burt pointed out that -

"The general acceptance of the theory that individual differences in ability are distributed in strict accordance with the Gaussian curve seems largely due to the advocacy of Thorndike, the acknowledged leader during the earlier decades of the century in the field of educational measurement" (p.176).

Burt is referring to E.L. Thorndike and his influential volume The Measurement of Intelligence published in 1927.

Arguing from work in the field of genetics, Burt considers that there are "..... strong theoretical grounds for believing that innate mental abilities are not distributed in exact conformity with the normal curve" (p.176). Burt's position is that both multifactorial and unifactorial inheritance is involved in the transmission of intellectual abilities. The genetic constitutional balance of the human species is however, so delicate that the effects of "..... exceptional genes, or of the mutations that produce them, are more likely to be unfavourable than favourable. Hence, the final outcome will be a distribution that is more or less skewed, the longer tail being in the downward direction" (p.176).

In the case of the present QT data from a restricted sample of school aged children it cannot be argued that the tail represented by low QT scores is necessarily the result of subjects with low innate intellectual ability. Low innate intellectual ability may well be implicated since Kearney (1966b) has argued that the QT measures a general intelligence factor that is an "innate property of the higher nervous system" (p.269) but this does not dismiss the possible effects of variables not controlled in research such as motivational factors or 'non-starters' where the nature of the task was not adequately comprehended despite practice.

TABLE 6.1.23 Frequency Distribution of QT Scores for European Sample by Numbers and Proportions(N = 397)

	QT Scores																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Numbers																		1	1	1
Proportion (100)																		.3	.3	.3

	QT Scores																			
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Numbers	0	1	0	1	0	1	5	7	2	5	6	6	12	16	17	12	17	26	19	25
Proportion (100)	0	.3	0	.3	0	.3	1.3	1.8	.5	1.3	1.5	1.5	3.0	4.0	4.3	3.0	4.3	6.5	4.8	6.3

	QT Scores																			
	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Numbers	17	24	30	24	25	20	17	18	15	9	10	2	3	1	1					
Proportion (100)	4.3	6.0	7.6	6.0	6.3	5.0	4.3	4.5	3.8	2.3	2.5	.5	.8	.3	.3					

TABLE 6.1.24 Frequency Distribution of QT Scores for Maori Sample by Numbers and Proportion (N = 303)

	QT Scores																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Numbers																	1	1	1	3
Proportion (100)																	.3	.3	.3	1.0

	QT Scores																			
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Numbers	1	1	3	1	0	1	6	2	6	6	7	6	15	12	10	17	14	17	24	9
Proportion (100)	.3	.3	1.0	.3	0	.3	2.0	.7	2.0	2.0	2.3	2.0	5.0	4.0	3.3	5.6	4.6	5.6	7.9	3.0

	QT Scores																			
	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Numbers	16	19	19	17	17	13	11	10	11	2	0	2	1	1						
Proportion (100)	5.3	6.3	6.3	5.6	5.6	4.3	3.6	3.3	3.6	.7	0	.7	.3	.3						

TABLE 6.1.25 Frequency Distribution of QT Scores for Total Sample by Numbers and Proportions (N = 700)

	QT Scores																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	29	20
Numbers																	1	2	2	4
Proportion (100)																	.1	.3	.3	.6

	QT Scores																			
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Numbers	1	2	3	2	0	2	11	9	8	11	13	12	27	28	27	29	31	43	43	34
Proportion (100)	.1	.3	.4	.3	0	.3	1.6	1.3	1.1	1.6	1.9	1.7	3.9	4.0	3.9	4.1	4.4	6.1	6.1	4.9

	QT Scores																			
	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Numbers	33	43	49	41	42	33	28	28	26	11	10	4	4	2	1					
Proportion (100)	4.7	6.1	7.0	5.9	6.0	4.7	4.0	4.0	3.7	1.6	1.4	.6	.6	.3	.1					

FIGURE 6.11 Frequency Distribution of Knox Subtest Scores for Total Sample, European and Maori Samples

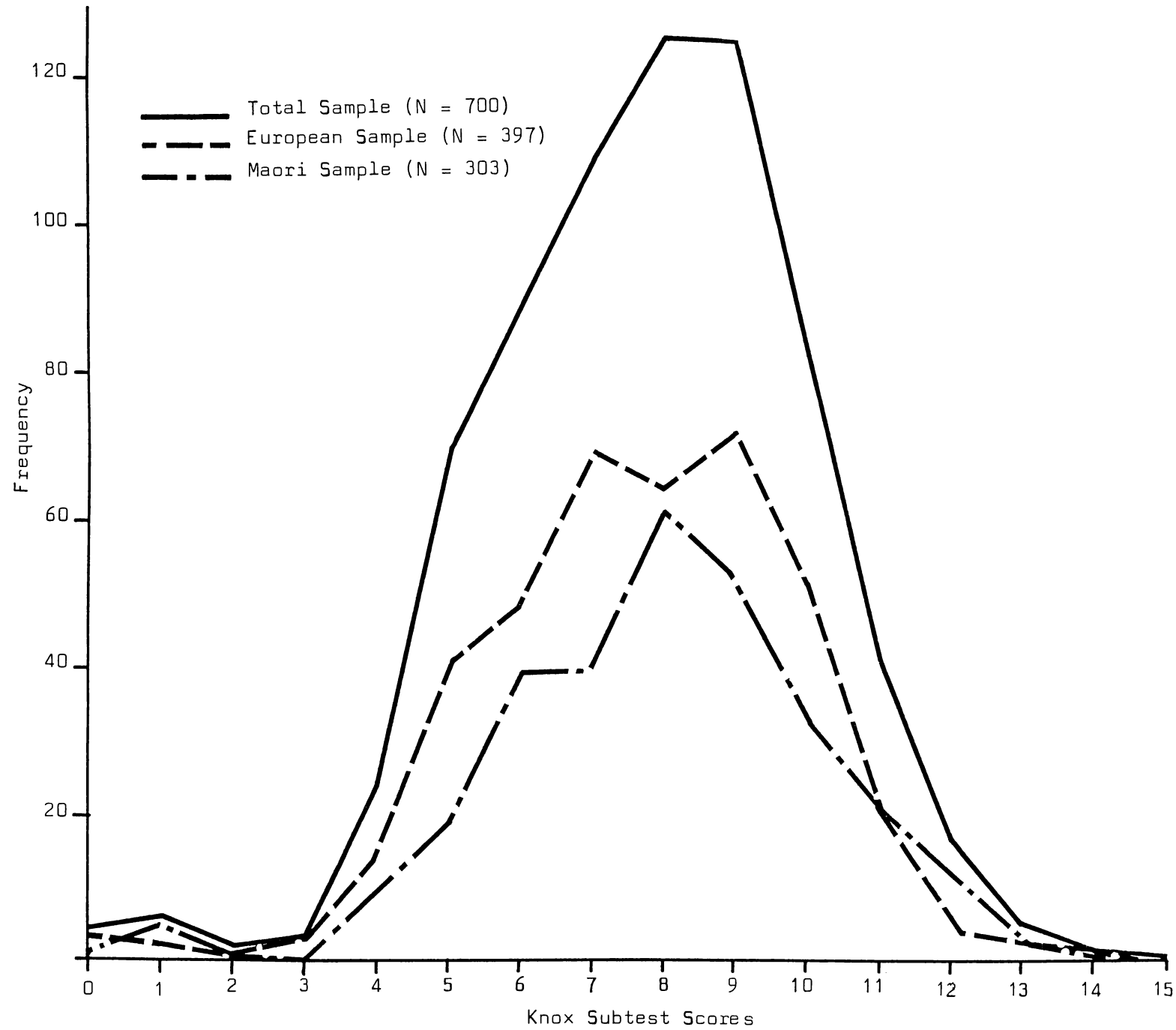


FIGURE 6.12

Frequency Distribution of Beads Subtest Scores for Total Sample, European and Maori Samples

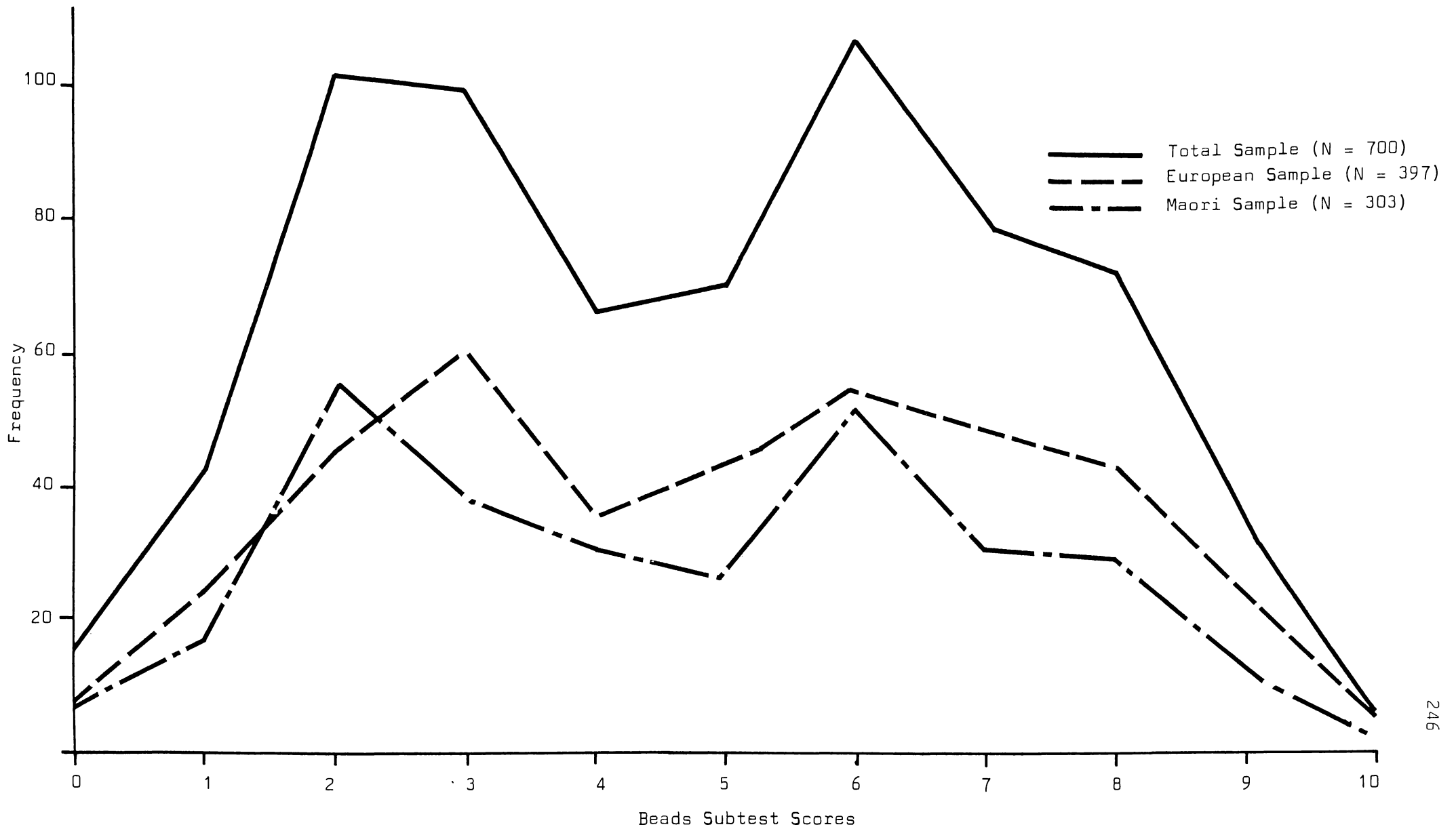


FIGURE 6.13 Frequency Distribution of Passalong Subtest Scores for Total Sample, European and Maori Samples

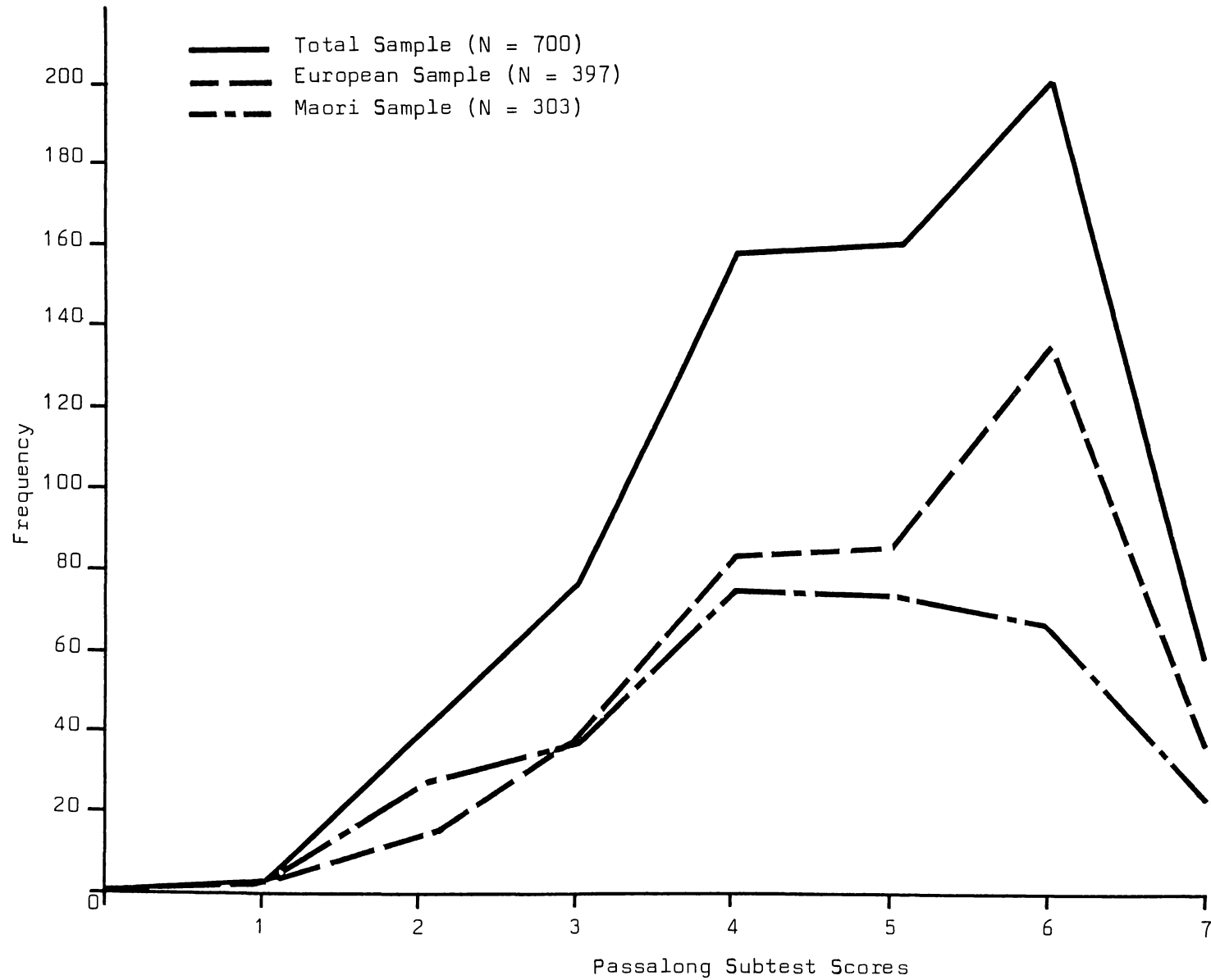


FIGURE 6.14

Frequency Distribution of Form Assembly Subtest Scores for Total Sample, European and Maori Samples

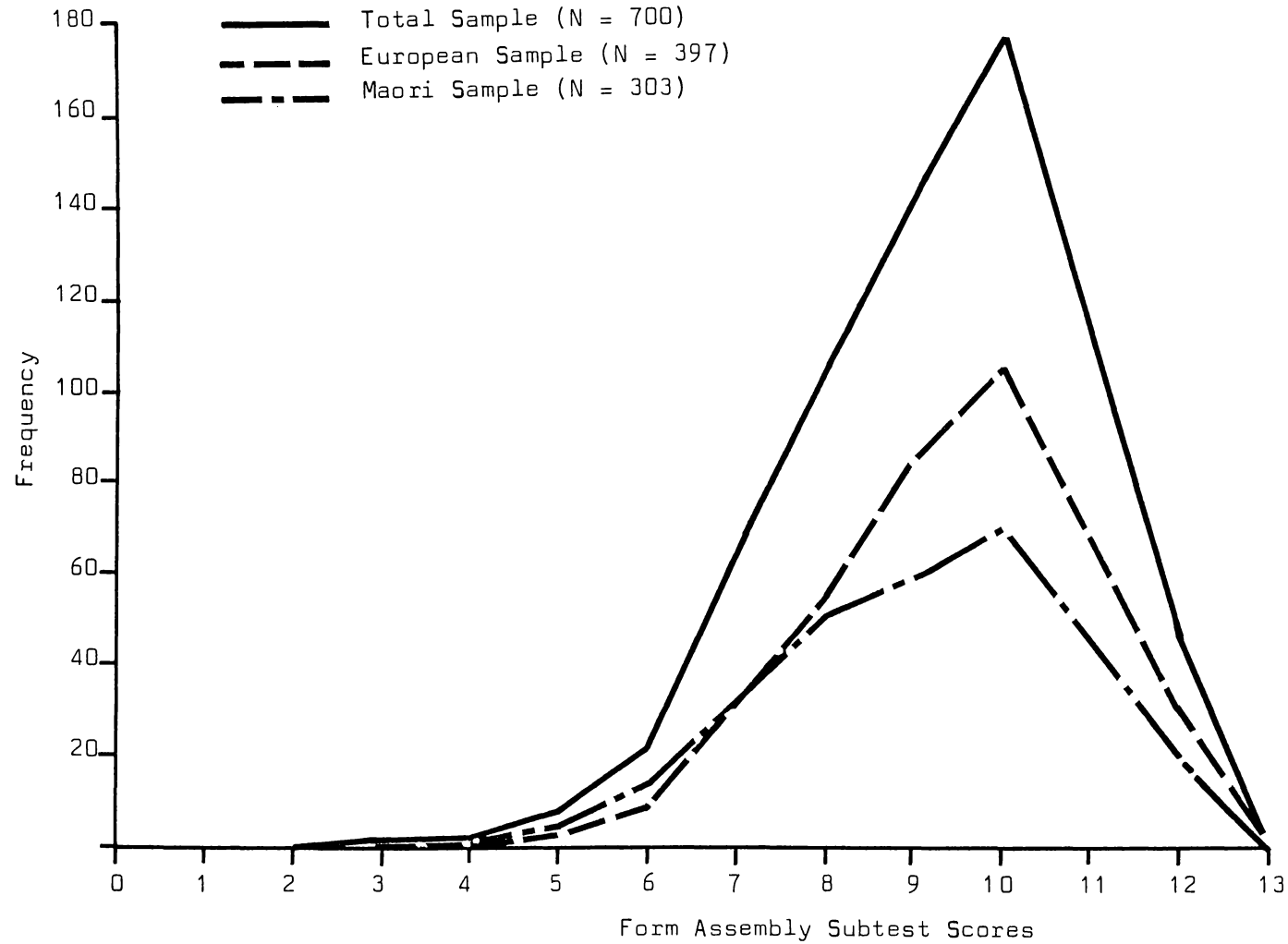
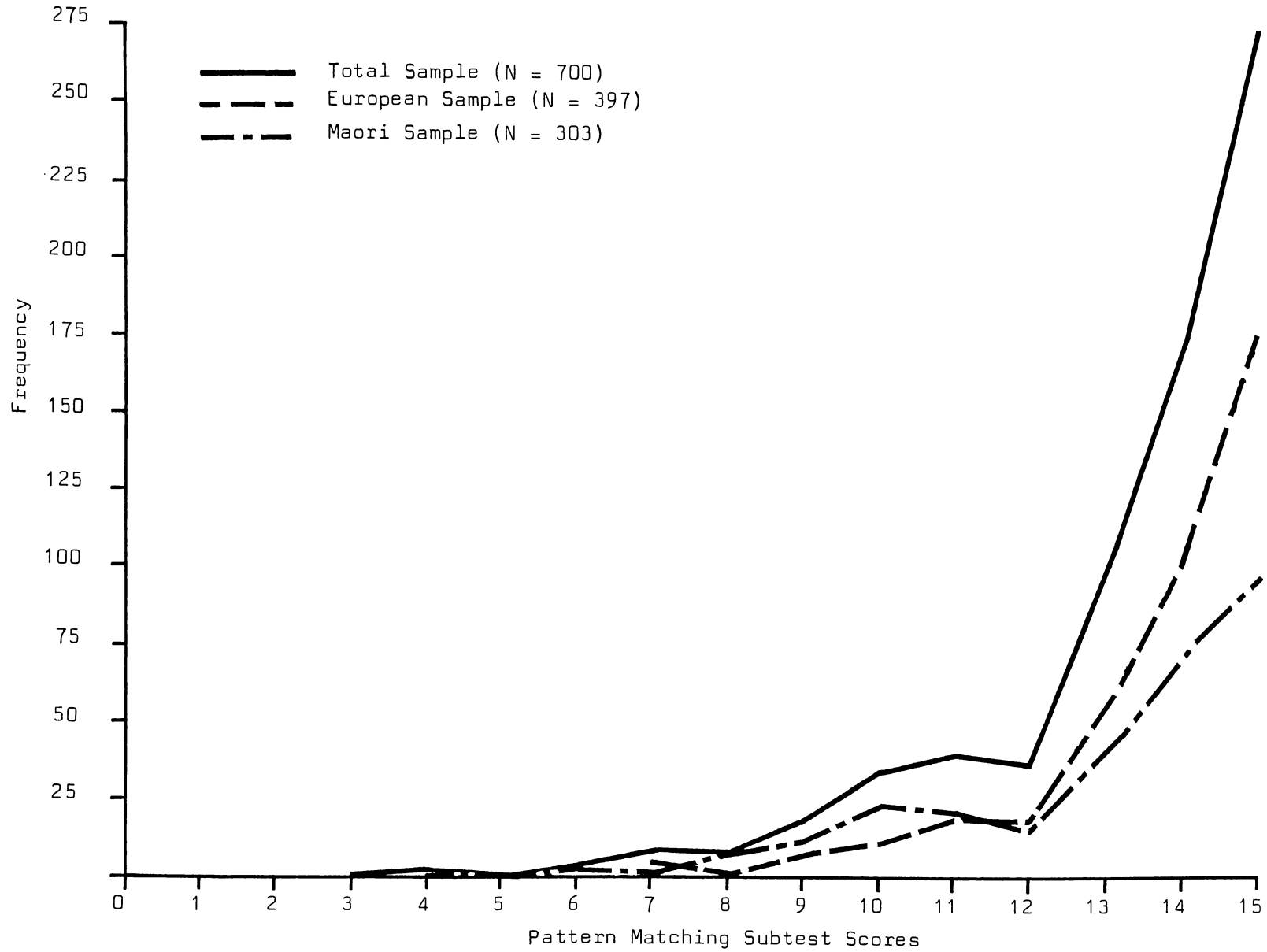


FIGURE 6.15 Frequency Distribution of Pattern Matching Subtest Scores for Total Sample, European and Maori Samples



While the data available cannot resolve the question of the causes of the asymmetrical distribution, the reference to Burt and his studies demonstrates that distributions of this nature may not be unusual. The fact that ability test data often assume a distribution similar to the theoretical mathematical distribution of the normal distribution curve simply allows certain dependable mathematical characteristics of the theoretical distribution to be applied to empirical data and this is not disputed by Burt. He says that "..... so long as we are concerned with small samples or with the general run of the population, a normal curve might still be trusted to yield a plausible fit" (p.176).

Tables 6.1.26 to 6.1.30 report the QT subtest score distributions. Numbers and proportions of subjects obtaining particular subtest scores are given for the separate and combined ethnic group samples. Figures 6.11 to 6.15 present frequency distributions of scores obtained by these samples on the QT subtests.

Inspection of these figures show that within subtests the score distributions for European and Maori samples are similar. The score distributions for the Knox and Beads subtests tend towards symmetry with the Knox distribution being unimodal, but the Beads distribution bimodal. The score distributions for the Passalong and Form Assembly subtests both evidence a degree of negative skew.

The score distribution for the Pattern Matching subtest clearly indicates that the most frequent scores were the higher subtest scores of 13 to 15. The distribution shows extreme negative skew and does not form a curve as such. With over a third of all subjects obtaining the maximum subtest score of 15, the psychometric properties of this subtest must be seriously questioned.

TABLE 6.1.26 Frequency Distributions of Knox Subtest Scores for European, Maori and Total Sample by Numbers and Proportions

	Knox Subtest Score															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N	3	2	1	3	15	40	48	69	64	72	51	21	5	2	1	0
European (N=397) Proportion (100)	.8	.5	.3	.8	3.8	10.0	12.1	17.4	16.1	18.1	12.8	5.3	1.3	.5	.3	.0
N	1	4	1	0	9	29	39	39	61	52	34	20	12	2	0	0
Maori (N=303) Proportion (100)	.3	1.3	.3	.0	3.0	9.6	12.9	12.9	20.1	17.2	11.2	6.6	4.0	.7	.0	.0
N	4	6	2	3	24	69	87	108	125	124	85	41	17	4	1	0
Total Sample (N=700) Proportion (100)	.6	.9	.3	.4	3.4	9.9	12.4	15.4	17.9	17.7	12.1	5.9	2.4	.6	.1	.0

TABLE 6.1.27 Frequency Distributions of Beads Subtest Scores for European, Maori and Total Sample by Numbers and Proportions

	Beads Subtest Score										
	0	1	2	3	4	5	6	7	8	9	10
N	8	25	46	61	36	44	55	49	44	24	5
European (N=397) Proportion (100)	2.0	6.3	11.6	15.4	9.1	11.1	13.9	12.3	11.1	6.0	1.3
N	7	18	56	39	31	27	52	31	29	12	1
Maori (N=303) Proportion (100)	2.3	5.9	18.5	12.9	10.2	8.9	17.2	10.2	9.6	4.0	.3
N	15	43	102	100	67	71	107	80	73	36	6
Total Sample (N=700) Proportion (100)	2.1	6.1	14.6	14.3	9.6	10.1	15.3	11.4	10.4	5.1	.9

While the data available cannot resolve the question of the causes of the asymmetrical distribution, the reference to Burt and his studies demonstrates that distributions of this nature may not be unusual. The fact that ability test data often assume a distribution similar to the theoretical mathematical distribution of the normal distribution curve simply allows certain dependable mathematical characteristics of the theoretical distribution to be applied to empirical data and this is not disputed by Burt. He says that "..... so long as we are concerned with small samples or with the general run of the population, a normal curve might still be trusted to yield a plausible fit" (p.176).

Tables 6.1.26 to 6.1.30 report the QT subtest score distributions. Numbers and proportions of subjects obtaining particular subtest scores are given for the separate and combined ethnic group samples. Figures 6.11 to 6.15 present frequency distributions of scores obtained by these samples on the QT subtests.

Inspection of these figures show that within subtests the score distributions for European and Maori samples are similar. The score distributions for the Knox and Beads subtests tend towards symmetry with the Knox distribution being unimodal, but the Beads distribution bimodal. The score distributions for the Passalong and Form Assembly subtests both evidence a degree of negative skew.

The score distribution for the Pattern Matching subtest clearly indicates that the most frequent scores were the higher subtest scores of 13 to 15. The distribution shows extreme negative skew and does not form a curve as such. With over a third of all subjects obtaining the maximum subtest score of 15, the psychometric properties of this subtest must be seriously questioned.

TABLE 6.1.26 Frequency Distributions of Knox Subtest Scores for European, Maori and Total Sample by Numbers and Proportions

	Knox Subtest Score																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
N	3	2	1	3	15	40	48	69	64	72	51	21	5	2	1	0	
European (N=397) Proportion (100)	.8	.5	.3	.8	3.8	10.0	12.1	17.4	16.1	18.1	12.8	5.3	1.3	.5	.3	.0	
N	1	4	1	0	9	29	39	39	61	52	34	20	12	2	0	0	
Maori (N=303) Proportion (100)	.3	1.3	.3	.0	3.0	9.6	12.9	12.9	20.1	17.2	11.2	6.6	4.0	.7	.0	.0	
N	4	6	2	3	24	69	87	108	125	124	85	41	17	4	1	0	
Total Sample (N=700) Proportion (100)	.6	.9	.3	.4	3.4	9.9	12.4	15.4	17.9	17.7	12.1	5.9	2.4	.6	.1	.0	

TABLE 6.1.27 Frequency Distributions of Beads Subtest Scores for European, Maori and Total Sample by Numbers and Proportions

	Beads Subtest Score										
	0	1	2	3	4	5	6	7	8	9	10
N	8	25	46	61	36	44	55	49	44	24	5
European (N=397) Proportion (100)	2.0	6.3	11.6	15.4	9.1	11.1	13.9	12.3	11.1	6.0	1.3
N	7	18	56	39	31	27	52	31	29	12	1
Maori (N=303) Proportion (100)	2.3	5.9	18.5	12.9	10.2	8.9	17.2	10.2	9.6	4.0	.3
N	15	43	102	100	67	71	107	80	73	36	6
Total Sample (N=700) Proportion (100)	2.1	6.1	14.6	14.3	9.6	10.1	15.3	11.4	10.4	5.1	.9

TABLE 6.1.28 Frequency Distributions of Passalong Subtest Scores for European, Maori and Total Sample by Numbers and Proportions

		Passalong Subtest Score							
		0	1	2	3	4	5	6	7
N		0	1	14	40	83	86	137	36
European (N=397)	Proportion (100)	0	.3	3.5	10.7	20.9	21.7	34.5	9.1
N		0	1	26	38	76	74	65	23
Maori (N=303)	Proportion (100)	0	.3	8.9	12.5	25.1	24.4	21.2	7.6
N		0	2	40	78	159	160	202	59
Total Sample (N=700)	Proportion (100)	0	.3	5.7	11.1	22.7	22.9	28.9	8.4

TABLE 6.1.29 Frequency Distributions of Form Assembly Subtest Scores for European, Maori and Total Sample by Numbers and Proportions

		Form Assembly Subtest Score													
		0	1	2	3	4	5	6	7	8	9	10	11	12	13
N		0	0	0	0	1	2	8	32	55	86	107	74	30	2
European (N=397)	Proportion (100)	0	0	0	0	.3	.5	2.0	8.1	13.9	21.7	27.0	18.6	7.6	.5
N		0	0	0	1	1	6	13	33	51	59	73	45	21	0
Maori (N=303)	Proportion (100)	0	0	0	.3	.3	2.0	4.3	10.9	16.8	19.5	24.1	14.9	6.9	0
N		0	0	0	1	2	8	21	65	106	145	180	119	51	2
Total Sample (N=700)	Proportion (100)	0	0	0	.1	.3	1.1	3.0	9.3	15.1	20.7	25.7	17.0	7.3	.3

TABLE 6.1.30 Frequency Distributions of Pattern Matching Subtest Scores of European, Maori and Total Sample by Numbers and Proportions

		Pattern Matching Subtest Score															
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N		0	0	0	1	0	0	0	5	2	7	11	19	19	55	103	175
European (N=397)																	
Proportion (100)		0	0	0	.3	0	0	0	1.3	.5	1.8	2.8	4.9	4.9	13.9	25.9	44.1
N		0	0	0	0	3	1	5	4	6	12	24	21	17	42	71	97
Maori (N=303)																	
Proportion (100)		0	0	0	0	1.0	.3	1.7	1.3	2.0	4.0	7.9	6.9	5.6	13.9	23.4	32.0
N		0	0	0	1	3	1	5	9	8	19	35	40	36	97	174	272
Total Sample (N=700)																	
Proportion (100)		0	0	0	.1	.4	.1	.7	1.3	1.1	2.7	5.0	5.7	5.1	13.9	24.9	38.9

(e) QT Subtest and QT Total Score Intercorrelations

The relationships between QT subtest and Total Score performances for the separate and combined ethnic groups within age levels was investigated by the calculation of Pearson Product-Moment correlation coefficients between all subtests scores and QT Total Score. The resulting matrices of coefficients are presented in Table 6.1.31 to 6.1.48. The required correlation coefficient value for statistical significance at $p < .01$ for a two-tailed test are indicated.

Focusing on the subtest correlations with QT Total Score, it can be seen that with the exception of one coefficient all correlations were significant for all groups at all age levels. The coefficient that failed to reach significance was that between Knox subtest scores and QT Total Score for Maoris aged 9:6 to 10:5.

With a single construct measure such as the QT, it is desirable for the subtests to have high positive correlations with the total score with correlations between subtests of a low positive order. Such a pattern is indicative of the unique contribution of each subtest to the total score and of the relative independence of each subtest one from another. By and large the QT subtests evidence moderately high, positive and statistically significant correlations with the QT Total Score. Correlations amongst the QT subtests are lower and not statistically significant in a good many cases.

It is apparent that there is an absence of relationship between a number of subtests, especially with some Maori age group samples. Performance on the Knox subtest for the Maori sample aged 9:6 to 10:5 bears no relationship to performance on the other QT subtests. This would also appear to be the case for the Maori sample aged 11:6 to 12:5.

In relation to the Knox subtest McElwain and Kearney (1970) have commented that it "is usually the least reliable and the least valid" (p.13). The Knox subtest intercorrelation data suggests that

this could also be the case with respect to the New Zealand data. It may also be that ethnic group membership influences the nature of the intellectual ability tapped by this subtest. The application of factor-analysis to subtest intercorrelation matrices may throw some light on this question.

TABLE 6.1.31 Product-Moment Correlation Coefficients for European Age Group 8:6 - 9:5 (N=36) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.05					
3. Passalong	.25	.40				
4. Form Assembly	.28	.19	.37			
5. Pattern Matching	.42	.24	.46	.43		
6. QT Total	.63	.55	.67	.63	.82	

All values of .42 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.32 Product-Moment Correlation Coefficients for Maori Age Group 8:6 - 9:5 (N=37) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.17					
3. Passalong	.51	.10				
4. Form Assembly	.37	.25	.42			
5. Pattern Matching	.31	.38	.50	.42		
6. QT Total	.65	.55	.69	.69	.83	

All values of .42 and above significant; $p < .01$, two-tailed test.

TABLE 6.1.33 Product-Moment Correlation Coefficients for Age Group 8:6 - 9:5 (N=73) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.10					
3. Passalong	.37	.26				
4. Form Assembly	.32	.22	.38			
5. Pattern Matching	.35	.30	.45	.46		
6. QT Total	.63	.54	.66	.67	.83	

All values of .30 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.34 Product-Moment Correlation Coefficients for European Age Group 9:6 - 10:5 (N=57) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.41					
3. Passalong	.40	.45				
4. Form Assembly	.24	.38	.29			
5. Pattern Matching	.18	.42	.28	.32		
6. QT Total	.66	.81	.66	.64	.64	

All values of .34 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.35 Product-Moment Correlation Coefficients for Maori Age Group 9:6 - 10:5 (N=33) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	-.09					
3. Passalong	-.03	.03				
4. Form Assembly	.03	.24	.09			
5. Pattern Matching	-.05	.29	.31	.45		
6. QT Total	.31	.56	.44	.62	.79	

All values of .44 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.36 Product-Moment Correlation Coefficients for Age Group 9:6 - 10:5 (N=90) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.26					
3. Passalong	.27	.35				
4. Form Assembly	.18	.34	.23			
5. Pattern Matching	.10	.40	.34	.36		
6. QT Total	.55	.75	.62	.62	.71	

All values of .27 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.37 Product-Moment Correlation Coefficients for European Age Group 10:6 - 11:5 (N=125) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.24					
3. Passalong	.31	.14				
4. Form Assembly	.14	.09	.15			
5. Pattern Matching	.33	.32	.26	.13		
6. QT Total	.71	.67	.54	.41	.67	

All values of .22 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.38 Product-Moment Correlation Coefficients for Maori Age Group 10:6 - 11:5 (N=69) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.18					
3. Passalong	-.05	.22				
4. Form Assembly	.15	-.02	-.08			
5. Pattern Matching	.29	.26	.16	.18		
6. QT Total	.60	.69	.37	.38	.68	

All values of .31 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.39 Product-Moment Correlation Coefficients for Age Group 10:6 - 11:5 (N=194) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.20					
3. Passalong	.16	.18				
4. Form Assembly	.12	.05	.07			
5. Pattern Matching	.29	.31	.24	.17		
6. QT Total	.64	.68	.48	.41	.68	

All values of .18 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.40 Product-Moment Correlation Coefficients for European Age Group 11:6 - 12:5 (N=63) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.32					
3. Passalong	.10	.31				
4. Form Assembly	.06	.19	.10			
5. Pattern Matching	.29	.25	.40	.28		
6. QT Total	.63	.76	.52	.47	.66	

All values of .31 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.41 Product-Moment Correlation Coefficients for Maori Age Group 11:6 - 12:5 (N=59) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	-.06					
3. Passalong	-.04	.20				
4. Form Assembly	-.01	.44	.35			
5. Pattern Matching	.13	.13	.47	.53		
6. QT Total	.37	.58	.57	.72	.71	

All values of .36 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.42 Product-Moment Correlation Coefficients for Age Group 11:6 - 12:5 (N=122) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.15					
3. Passalong	.01	.26				
4. Form Assembly	.02	.31	.26			
5. Pattern Matching	.18	.19	.46	.44		
6. QT Total	.48	.67	.57	.62	.70	

All values of .24 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.43 Product-Moment Correlation Coefficients for European Age Group 12:6 - 13:5 (N=78) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.34					
3. Passalong	.15	.21				
4. Form Assembly	.25	.16	.27			
5. Pattern Matching	.42	.54	.23	.43		
6. QT Total	.73	.75	.45	.57	.75	

All values of .29 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.44 Product-Moment Correlation Coefficients for Maori Age Group 12:6 - 13:5 (N=60) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.19					
3. Passalong	.25	.02				
4. Form Assembly	.08	.25	.13			
5. Pattern Matching	.16	.27	.24	.30		
6. QT Total	.65	.64	.48	.55	.60	

All values of .33 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.45 Product-Moment Correlation Coefficients for Age Group 12:6 - 13:5 (N=138) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.28					
3. Passalong	.18	.12				
4. Form Assembly	.16	.19	.22			
5. Pattern Matching	.27	.40	.27	.38		
6. QT Total	.69	.71	.46	.57	.67	

All values of .21 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.46 Product-Moment Correlation Coefficients for European Age Group 13:6 - 14:5 (N=38) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.27					
3. Passalong	.54	.13				
4. Form Assembly	.23	.25	.38			
5. Pattern Matching	.32	.27	.32	.30		
6. QT Total	.74	.69	.65	.62	.55	

All values of .42 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.47 Product-Moment Correlation Coefficients for Maori Age Group 13:6 - 14:5 (N=45) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.28					
3. Passalong	.17	.27				
4. Form Assembly	.15	.22	.10			
5. Pattern Matching	.29	.31	.46	.33		
6. QT Total	.67	.69	.56	.54	.71	

All values of .38 and above significant: $p < .01$, two-tailed test.

TABLE 6.1.48 Product-Moment Correlation Coefficients for Age Group 13:6 - 14:5 (N=83) Among QT Subtests and QT Total Score

	1	2	3	4	5	6
1. Knox						
2. Beads	.28					
3. Passalong	.31	.28				
4. Form Assembly	.17	.25	.25			
5. Pattern Matching	.28	.29	.43	.34		
6. QT Total	.68	.69	.61	.59	.65	

All values of .28 and above significant: $p < .01$, two-tailed test.

PART 2. DATA ON QT RELIABILITY

The reliability of the QT and its subtests has been investigated by both internal consistency and test-retest methods. Data on the standard error of measurement are also reported.

(a) Internal Consistency Estimates of Reliability

Internal consistency estimates have been obtained by the Kuder-Richardson Formula 20 (henceforth K-R 20) procedure (Kuder and Richardson, 1937). This procedure utilizes information on the proportion of subjects passing and failing each item rendering a measure of error in terms of the extent to which items are correlated.

Reliability estimates by the K-R 20 method were obtained for the European and Maori samples and the Total Sample at each age group level. The estimates have been derived for the QT as a whole and for each of the QT subtests at each age group level. These are reported in Tables 6.2.1 to 6.2.6.

The K-R 20 coefficients for the QT indicate a relatively high level of internal consistency at the six age levels for the European, Maori and combined ethnic group samples. The data supports the proposition that the items are sufficiently well selected from the universe of such items as to be representative of them, an interpretation of such estimates put forward by Cronbach (1951).

The coefficients range from .70 to .86 which is very similar to the range obtained by McElwain and Kearney (1970) with Australian Aboriginal and European subjects aged between seven and fourteen years. They obtained QT K-R 20 coefficients of .73 to .92.

Internal consistency estimates of reliability have been reported for the 'Performance Scale' in the Wechsler tests. Employing the Spearman-Brown split half method the internal consistency of the WAIS Performance Scale is reported as between .93 and .94 for adult samples (Wechsler, 1955), the WISC Performance Scale between .86 to .90 with

TABLE 6.2.1 K-R 20 Reliability Coefficients for QT by Age Groups and Separate and Combined Ethnic Group Samples

Age Group	European		Maori		Total Sample	
	N	K-R 20	N	K-R 20	N	K-R 20
8:6 - 9:5	36	.86	37	.85	73	.86
9:6 - 10:5	57	.82	33	.72	90	.80
10:6 - 11:5	125	.76	69	.76	194	.76
11:6 - 12:5	63	.77	59	.77	122	.78
12:6 - 13:5	78	.78	60	.70	138	.75
13:6 - 14:5	38	.82	45	.82	83	.82

TABLE 6.2.2 K-R 20 Reliability Coefficients for Knox Subtest by Age Groups and Separate and Combined Ethnic Group Samples

Age Group	European		Maori		Total Sample	
	N	K-R 20	N	K-R 20	N	K-R 20
8:6 - 9:5	36	.68	37	.60	73	.64
9:6 - 10:5	57	.55	33	.53	90	.55
10:6 - 11:5	125	.60	69	.62	194	.62
11:6 - 12:5	63	.59	59	.50	122	.54
12:6 - 13:5	78	.63	60	.63	138	.63
13:6 - 14:5	38	.67	45	.72	83	.70

TABLE 6.2.3 K-R 20 Reliability Coefficients for Beads Subtest by Age Groups and Separate and Combined Ethnic Group Samples

Age Group	European		Maori		Total Sample	
	N	K-R 20	N	K-R 20	N	K-R 20
8:6 - 9:5	36	.70	37	.61	73	.66
9:6 - 10:5	57	.68	33	.69	90	.69
10:6 - 11:5	125	.69	69	.80	194	.74
11:6 - 12:5	63	.77	59	.61	122	.71
12:6 - 13:5	78	.70	60	.52	138	.63
13:6 - 14:5	38	.76	45	.66	83	.71

TABLE 6.2.4 K-R 20 Reliability Coefficients for Passalong Subtest
by Age Groups and Separate and Combined Ethnic Group Samples

Age Group	European		Maori		Total Sample	
	N	K-R 20	N	K-R 20	N	K-R 20
8:6 - 9:5	36	.61	37	.53	73	.56
9:6 - 10:5	57	.50	33	.44	90	.50
10:6 - 11:5	125	.49	69	.54	194	.51
11:6 - 12:5	63	.42	59	.52	122	.50
12:6 - 13:5	78	.23	60	.41	138	.35
13:6 - 14:5	38	.61	45	.62	83	.62

TABLE 6.2.5 K-R 20 Reliability Coefficients for Form Assembly Subtest
by Age Groups and Separate and Combined Ethnic Group Samples

Age Group	European		Maori		Total Sample	
	N	K-R 20	N	K-R 20	N	K-R 20
8:6 - 9:5	36	.47	37	.47	73	.48
9:6 - 10:5	57	.44	33	.19	90	.36
10:6 - 11:5	125	.12	69	.45	194	.27
11:6 - 12:5	63	.23	59	.57	122	.43
12:6 - 13:5	78	.30	60	.24	138	.27
13:6 - 14:5	38	.56	45	.54	83	.55

TABLE 6.2.6 K-R 20 Reliability Coefficients for Pattern Matching Subtest
by Age Groups and Separate and Combined Ethnic Group Samples

Age Group	European		Maori		Total Sample	
	N	K-R 20	N	K-R 20	N	K-R 20
8:6 - 9:5	36	.86	37	.85	73	.87
9:6 - 10:5	57	.68	33	.78	90	.76
10:6 - 11:5	125	.69	69	.71	194	.70
11:6 - 12:5	63	.66	59	.78	122	.74
12:6 - 13:5	78	.58	60	.57	138	.57
13:6 - 14:5	38	.38	45	.74	83	.69

three age group samples of children (Wechsler, 1949) and for the new WISC-R Performance Scale between .89 to .91 over eleven age groups of children (Wechsler, 1974). Relative to the internal consistency estimates from the performance scales in the Wechsler tests those obtained with the QT compare quite favourably.

The internal consistency estimates of reliability for the QT subtests are not as high as for the whole test which is to be expected. Test length is a critical component in the estimation of reliability (see Mehrens and Lehmann, 1973, p.115) and a test of 60 items should be more reliable than one of, for instance, 7 items as is the case with the Passalong Subtest.

If the Tables are inspected for each subtest it can be seen that for the Knox Subtest the K-R 20 internal consistency estimates are of a similar order across the age range for both ethnic groups and are of a moderate order. The same could also be said in relation to the Beads Subtest.

The Passalong Subtest with its small number of items shows somewhat lower levels of internal consistency in comparison with the preceding subtests. The estimates are lowest for the 12:6 - 13:5 age group on this subtest.

The Form Assembly K-R 20 coefficients are also of a comparatively low order, although it contains more items than both the Beads and Passalong Subtests. The estimates suggest especially poor internal consistency in relation to the European sample age 10:6 to 11:5 and Maori sample aged 9:6 to 10:5. On the other hand, the Pattern Matching Subtest K-R 20 coefficients are generally of a moderately high order. A good many are on a par with those obtained for the QT as a whole for the equivalent sample group.

(b) Test-Retest Estimates of Reliability

Test-retest reliability (r_{tt}) data on the QT and its subtests

have been obtained from four class group samples with a test-retest interval of six months. Details on the class groups are to be found in Chapter 4. The test-retest correlation coefficients (sometimes referred to as stability coefficients) for the QT subtests and QT Total Score by class groups are reported in Tables 6.2.7 to 6.2.10. Test-retest correlation coefficients for the combined class groups are given in Table 6.2.11. Data on the means and standard deviations for the QT and its subtests and the correlation coefficients amongst these variables at both test and retest condition for the separate and combined class group samples are reported in Appendix B.

Inspection of Tables 6.2.7 to 6.2.10 show that statistically significant test-retest coefficients were obtained for the QT Total Score with each class group. While this is evidence of stability in performances on the QT over a six-month interval, it must be noted that three of test-retest coefficients are below those reported for eight age groups of Australian Aboriginal children over a one-year interval. In the latter case, the test-retest coefficients were

TABLE 6.2.7 QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 1 (N=27)
(Product-Moment Coefficients, Interval 6 months)

	r_{tt}
Knox	.50
Beads	.45
Passalong	.50
Form Assembly	.40
Pattern Matching	.73
QT Total	.76

Values of .49 and above significant, $p < .01$, two-tailed test.

TABLE 6.2.8 QT Subtest and QT Total Score Test-Retest Correlation
Coefficients for Class Group 2 (N=35)
(Product-Moment Coefficients, interval 6 months)

	r_{tt}
Knox	.44
Beads	.53
Passalong	.32
Form Assembly	.06
Pattern Matching	.68
QT Total	.54

Values of .43 and above significant, $p < .01$, two-tailed test.

TABLE 6.2.9 QT Subtest and QT Total Score Test-Retest Correlation
Coefficients for Class Group 3 (N=34)
(Product-Moment Coefficients, interval 6 months)

	r_{tt}
Knox	.19
Beads	.17
Passalong	.73
Form Assembly	.37
Pattern Matching	.46
QT Total	.50

Values of .43 and above significant, $p < .01$, two-tailed test.

TABLE 6.2.10 QT Subtest and QT Total Score Test-Retest Correlation
Coefficients for Class Group 4 (N=31)
(Product-Moment Coefficients, interval 6 months)

	r_{tt}
Knox	.63
Beads	.40
Passalong	.27
Form Assembly	.47
Pattern Matching	.22
QT Total	.66

Values of .46 and above significant, $p < .01$, two-tailed test.

TABLE 6.2.11 QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Groups 1 to 4 Combined (N=127)
(Product-Moment Coefficients, interval 6 months)

	r_{tt}
Knox	.42
Beads	.40
Passalong	.46
Form Assembly	.33
Pattern Matching	.65
QT Total	.66

Values of .23 and above significant, $p < .01$, two-tailed test.

between .741 and .897 (reported in McElwain and Kearney, 1970, p.71). The combined Class Group test-retest coefficient of .66 reported in Table 6.2.11 while statistically significant, must also be compared with that of .875 obtained by McElwain and Kearney for the combined Australian samples mentioned above.

Mehrens and Lehmann (1973) discussing reliability and test use in decision making, rightly point out that absolute standards on the level of reliability for tests cannot be made. Reliability coefficients of all types are relative to factors in the test instrument itself and the samples. They see the more important issue as being the 'goodness' of the decision that can be made with test information of even low reliability over no information or that of unknown reliability. They do, however, point out that -

"it is generally accepted that standardized tests used to assist in making decisions about individuals should have reliability coefficients of at least .85" (Mehrens and Lehmann, 1973, p.122).

Against such a criterion it could be argued that while the K-R 20 internal consistency estimates of the reliability of the QT obtained with New Zealand samples measure up reasonably well, the test-retest reliability coefficients do not.

The test-retest reliability coefficients for the QT subtests are of a lower order for each of the class groups than those for the QT overall. Inspection of the tables shows that there is an absence of a statistically significant relationship between test and retest performances on some subtests over the relatively short six-month interval.

Test-retest reliability coefficients for the QT Total Score and the five subtests for the European and Maori samples in each class group are reported in Tables 6.2.12 to 6.2.15. The coefficients from the pooled samples of European and Maori subjects are reported in Table 6.2.16. The data concerning means, standard deviations and intercorrelations at test and retest condition for these samples are included in Appendix B.

TABLE 6.2.12 QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 1 European and Maori Subjects
(Product-Moment Coefficients, interval 6 months)

	r_{tt}	
	European ¹ N=17	Maori ² N=10
Knox	.56	.29
Beads	.20	.73
Passalong	.38	.64
Form Assembly	.57	.12
Pattern Matching	.63	.83
QT Total	.73	.79

- 1 Values of .61 and above significant, $p < .01$, two-tailed test.
- 2 Values of .77 and above significant, $p < .01$, two-tailed test.

TABLE 6.2.13 QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 2 European and Maori Subjects
(Product-Moment Coefficients, interval 6 months)

	r_{tt}	
	European ¹ N=27	Maori ² N=8
Knox	.59	-.08
Beads	.32	.94
Passalong	.36	.00
Form Assembly	.00	.21
Pattern Matching	.60	.93
QT Total	.54	.54

1 Values of .49 and above significant, $p < .01$, two-tailed test.

2 Values of .83 and above significant, $p < .01$, two-tailed test.

TABLE 6.2.14 QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 3 European and Maori Subjects
(Product-Moment Coefficients, interval 6 months)

	r_{tt}	
	European ¹ N=27	Maori ² N=8
Knox	.23	-.02
Beads	.53	.18
Passalong	.54	.84
Form Assembly	.27	.49
Pattern Matching	.59	.37
QT Total	.44	.54

1 Values of .59 and above significant, $p < .01$, two-tailed test.

2 Values of .62 and above significant, $p < .01$, two-tailed test.

TABLE 6.2.15 QT Subtest and QT Total Score Test-Retest Correlation Coefficients for Class Group 4 European and Maori Subjects
(Product-Moment Coefficients, interval 6 months)

	r_{tt}	
	European ¹ N=18	Maori ² N=13
Knox	.71	.56
Beads	.34	.48
Passalong	.21	.29
Form Assembly	.34	.52
Pattern Matching	.21	.16
QT Total	.65	.65

1 Values of .59 and above significant, $p < .01$, two-tailed test.

2 Values of .68 and above significant, $p < .01$, two-tailed test.

TABLE 6.2.16 QT Subtest and QT Total Score Test-Retest Correlation Coefficients for European and Maori Subjects
From Class Groups 1 to 4 Combined
(Product-Moment Coefficients, interval 6 months)

	r_{tt}	
	European ¹ N=80	Maori ² N=47
Knox	.50	.26
Beads	.39	.42
Passalong	.37	.56
Form Assembly	.31	.38
Pattern Matching	.65	.67
QT Total	.64	.66

1 Values of .29 and above significant, $p < .01$, two-tailed test.

2 Values of .35 and above significant, $p < .01$, two-tailed test.

Within each class group the test-retest coefficients for the QT Total Score are of a similar order for European and Maori subjects, although not all of the correlations achieved statistical significance.

There was no statistically significant relationship between test and retest QT Total Score results for Maori subjects of Class Group 2,

both European and Maori subjects of Class Group 3 and Maori subjects of Class Group 4. The relatively low and statistically non-significant coefficients are most probably the result of restriction of range effects given the relatively small and homogeneous samples.

The test-retest correlations for the QT Total Score from the combined class group European and Maori samples were both statistically significant. Again however, the coefficients are of a lower order than those reported for the QT with Australian subjects over a longer time interval.

The data shows that QT performance is of the same order of stability for European and Maori subjects over a six-month interval. However, the test-retest reliability coefficients are for both ethnic groups below the level considered desirable for the use of the QT in individual decision making.

The test-retest reliability coefficients for the QT subtests for the European and Maori samples were in general of a lower order than those for the QT overall. It is apparent from a comparison of the coefficients for the subtests that the test-retest reliabilities varied considerably between European and Maori samples. Within class groups there are some quite marked differences in the levels of correlation. For instance, in Class Group 2, while there is a statistically significant positive relationship between test and retest performances on the Knox Subtest for European subjects, there is a small non-significant negative relationship for Maori subjects. Here again the effects of small and relatively homogeneous samples probably play a part in the results along with measurement error. A comparison of the coefficients from the pooled ethnic group samples from each class does suggest ethnic group differences in subtest stability. On a comparative basis it would appear that the Knox Subtest does not yield particularly reliable results with Maori subjects in comparison with Europeans. Meanwhile,

the reverse would appear to be the case with the Passalong Subtest.

(c) Estimates of the Standard Error of Measurement

Another approach to estimating test reliability is by determining the standard error of measurement (SEm). Ideally, measures of psychological functions aspire to low standard errors of measurement so that any single score is very close to the theoretical true score in the intra-individual case. Thus, differences in scores in the inter-individual situation represent real differences as opposed to departures from the true scores brought about by errors of measurement. Guilford and Fruchter (1973) give a clear account of the meaning and method of estimating the standard error of measurement.

The standard error of measurement for the QT Total Score and the subtests are reported in Tables 6.2.17 to 6.2.22. These estimates are reported for European, Maori and Total Samples for each of the six age group levels.

The standard error of measurement estimates for the QT Total Score are of a comparable order for the European and Maori samples at each age group level. The values obtained within the ethnic groups evidence a relatively wide range. There is a tendency, however, for the estimates to decrease with age.

In the case of European subjects, the lowest standard error of measurement estimate was 4.63 QT points, the highest 6.52 QT points. For Maori subjects, 3.95 QT points and 6.73 QT points respectively. A range of similar magnitude was also evident in the combined ethnic group samples. The magnitude of the range is highlighted by the fact that the same estimates from Australian samples were in the order of 2.25 to 2.91 QT points (McElwain and Kearney, 1970, p.74). In the light of the Australian standard error of measurement estimates cited

TABLE 6.2.17 Standard Error of Measurement Estimates for QT
Total Score by Age Group and for Separate and Combined Ethnic Groups

Age Group	European		Maori		Total Sample	
	N	SEm	N	SEm	N	SEm
8:6 - 9:5	36	6.52	37	6.73	73	6.79
9:6 - 10:5	57	5.58	33	4.25	90	5.40
10:6 - 11:5	125	4.66	69	4.76	194	4.72
11:6 - 12:5	63	4.73	59	4.98	122	4.93
12:6 - 13:5	78	4.63	60	3.95	138	4.37
13:6 - 14:5	38	5.19	45	5.51	83	5.46

TABLE 6.2.18 Standard Error of Measurement Estimates for Knox
Subtest by Age Group and for Separate and Combined Ethnic Groups

Age Group	European		Maori		Total Sample	
	N	SEm	N	SEm	N	SEm
8:6 - 9:5	36	1.96	37	1.60	73	1.78
9:6 - 10:5	57	1.41	33	1.34	90	1.39
10:6 - 11:5	125	1.63	69	1.68	194	1.67
11:6 - 12:5	63	1.52	59	1.35	122	1.43
12:6 - 13:5	78	1.72	60	1.68	138	1.71
13:6 - 14:5	38	1.83	45	2.09	83	1.97

TABLE 6.2.19 Standard Error of Measurement Estimates for Beads
Subtest by Age Group and for Separate and Combined Ethnic Groups

Age Group	European		Maori		Total Sample	
	N	SEm	N	SEm	N	SEm
8:6 - 9:5	36	1.78	37	1.40	73	1.59
9:6 - 10:5	57	1.84	33	1.60	90	1.80
10:6 - 11:5	125	1.87	69	2.35	194	2.07
11:6 - 12:5	63	2.15	59	1.61	122	1.92
12:6 - 13:5	78	1.96	60	1.37	138	1.68
13:6 - 14:5	38	1.97	45	1.79	83	2.01

TABLE 6.2.20 Standard Error of Measurement Estimates for Passalong Subtest by Age Group and for Separate and Combined Ethnic Groups

Age Group	European		Maori		Total Sample	
	N	SEm	N	SEm	N	SEm
8:6 - 9:5	36	1.03	37	1.01	73	1.01
9:6 - 10:5	57	0.94	33	0.81	90	0.94
10:6 - 11:5	125	0.90	69	1.01	194	0.95
11:6 - 12:5	63	0.72	59	0.99	122	0.92
12:6 - 13:5	78	0.43	60	0.76	138	0.64
13:6 - 14:5	38	1.03	45	1.12	83	1.10

TABLE 6.2.21 Standard Error of Measurement Estimates for Form Assembly Subtest by Age Group and for Separate and Combined Ethnic Groups

Age Group	European		Maori		Total Sample	
	N	SEm	N	SEm	N	SEm
8:6 - 9:5	36	1.17	37	1.29	73	1.27
9:6 - 10:5	57	1.12	33	0.62	90	0.96
10:6 - 11:5	125	0.45	69	1.08	194	0.75
11:6 - 12:5	63	0.67	59	1.38	122	1.07
12:6 - 13:5	78	0.76	60	0.71	138	0.74
13:6 - 14:5	38	1.24	45	1.29	83	1.29

TABLE 6.2.22 Standard Error of Measurement Estimates for Pattern Matching Subtest by Age Group and for Separate and Combined Ethnic Groups

Age Group	European		Maori		Total Sample	
	N	SEm	N	SEm	N	SEm
8:6 - 9:5	36	2.67	37	3.02	73	3.04
9:6 - 10:5	57	1.48	33	2.15	90	1.89
10:6 - 11:5	125	1.37	69	1.55	194	1.46
11:6 - 12:5	63	1.25	59	1.90	122	1.62
12:6 - 13:5	78	0.76	60	0.98	138	0.89
13:6 - 14:5	38	0.50	45	1.45	83	1.16

it is apparent that the QT is a less reliable instrument with New Zealand subjects. The higher standard error of measurement estimates applying to the New Zealand samples means that considerably wider confidence limits must be employed in relation to any score.

The standard error of measurement of the five QT subtests for European, Maori and Total Samples at each age group level while reported are not discussed. It is appropriate to note, however, that values obtained were in many instances higher than those reported for Australian subjects (see McElwain and Kearney, 1970, pp.75-79).

PART 3 DATA ON QT VALIDITY

The validity of the QT for use under New Zealand conditions with European and Maori subjects has been investigated. Consideration has been given to the content validity, the criterion-related validity and the construct validity of the QT.

(a) Content Validity of the QT

The description of the QT subtests given in Chapter 5 illustrates its 'non-verbal' performance type nature. The description of the method of administration and the subtest content suggests that there is an emphasis on spatial-perceptual skills in item tasks. It was also evident that considerable effort went into the selection and development of item types of this nature known to be useful in the assessment of intellectual ability with European subjects. Furthermore, evidence from the use of these item types prior to the QT with non-European subjects has suggested that they do measure aspects of intellectual ability considered important in some 'Western' constructs of intelligence.

The QT as a measure of general intellectual ability in the tradition of Spearman's 'g' is said to sample aspects of intellectual ability found to be of practical significance in selecting "persons for training in complex European skills from groups where the psychological tests usually used in European groups were inapplicable" (McElwain and Kearney, 1970, p.2). The authors of the QT Handbook acknowledge that there are other approaches to the study of intelligence and suggest that 'cognitive style' considerations might arise from some QT data.

In Chapter 1 issues relating to the cross-cultural use of psychological measures of intellectual ability were discussed. It was noted that the questions of use and content must be seen in terms of context and need. The emergence of cultural relativism as a principle

underlying research in this field has also encouraged the more cautious use of such measures and the more circumspect interpretation of test data.

While the content of the QT is not obviously related to a particular cultural context in the way that, say, Guthrie's (n.d.) 'Information Test of Intelligence' for use in the rural Phillipines is (with all items based on the knowledge and day-to-day experiences of rural Filipino subjects), or in the way instruments based upon Irvine's (1970) or Serpell's (1974) analysis of the components of intelligence behaviour amongst African tribal groups might be, it is still not said to be 'culture-free'. Already it has been noted that culturally determined skills may facilitate QT performance. McElwain⁶ reflecting on the evidence from the use of the QT in Australia and elsewhere has said that he believes the frame of reference of the test and its content would favour modes of thought most common to the Indo-European family of languages. McElwain, while naturally reticent about these speculations was, however, pointing out that the form and content of the QT arose in a cultural-linguistic context which will have both explicitly and implicitly influenced its development.

Subsequently it was noted in the QT Handbook (McElwain and Kearney, 1970) that "unpublished studies using the Illinois Test of Psycholinguistic Ability (Kirk and McCarthy, 1961) together with the QT suggest that QT performance is enhanced by the ability to speak (and presumably to think) in English". (p.136).

As a result, the assumption that its content is universally valid would be difficult to maintain even though efforts were made to eliminate or control obvious cultural influences.

⁶Comment by Prof. D.W. McElwain at the 'Research Workshop on Cross-Cultural Ability Testing', University of Waikato, 4-6 February, 1970.

While such efforts can be applauded it is perhaps instructive to consider the content-validity and 'cultural appropriateness' debate in somewhat wider terms. An intriguing proposition is the possibility that the emerging theory and procedures for the adaptation of psychological tests to new cultural contexts may be as culture bound as the theory and procedures underlying the original test. It has been argued that the study and measurement of intelligence has been a pursuit dominated by a science of psychology based on preconceptions, modes of thought and methods of enquiry arising out of the intellectual tradition of Europe (see Serpell, 1972). The application of psychological techniques - such as the intelligence test - to peoples whose cultures and languages differed from those of the communities of Europe and North America brought about the adaptation of measures of intelligence so that, in theory at least, they were more 'culturally appropriate'. The result has been the 'culture-free', 'culture-fair' and 'culture-reduced' tests which have placed strong emphasis on the solution of items presented visually as puzzles of one form or another as opposed to answering questions presented orally or in a written form. The point to consider is that the definition of 'cultural appropriateness' and the content and form of such tests have also been largely determined by psychologists well versed in the science as practised in the European and North American communities. Thus, it remains a distinct possibility that a belief in the 'cultural appropriateness' of the new or modified scales could be misleading. The dilemma that becomes apparent is that those involved in the enterprise of measuring intelligence, while on the one hand stepping back from claims concerning the universal applicability of intelligence tests, may have advanced what might well be culture-bound views on suitable test adaptations and content. If there is an element of truth in this proposition then it would add weight to the view that the content validity of a test such as the QT must be seen initially in conditional and relative terms until satisfactory evidence on psychological equivalence is available.

In discussing previous comparative research on the intellectual abilities of European and Maori subjects, it was noted that the suitability of the content of a number of measures had been questioned in relation to Maori subjects. Riske (1950), it will be recalled, saw the task of devising tests of intelligence with content equally familiar to Maori and European subjects as difficult, if not impossible. Instead, he advocated the development of a test or tests with 'equally unfamiliar' content.

It would seem reasonable to argue that the QT content is of necessity culturally determined. Nonetheless, in some respects its content could also be viewed as more 'equally unfamiliar' to differing cultural groups in comparison with the content of many other measures of ability.

Cronbach (1970) makes the points that content validity is essentially concerned with the way a measure fairly represents a universe of observations of interest. He argues that while the judgement of the suitability of content ultimately rests with the test user there is an obligation on the part of test authors to help in defining the universe of observations the test intends to sample. Content validity is, however, only mentioned briefly in the QT Handbook (McElwain and Kearney, 1970) in relation to subtest item homogeneity. Unfortunately, no clear definition of the universe of observations to be sampled is given in the way Cronbach suggests. An assessment of the content validity of the QT remains, therefore, a judgement on the part of the test user. This judgement must take account of the origin and development of the item types and the evidence on their utility as measures of intelligence.

Consideration of the item types, method of administration, general format and the successful application of the QT to subjects from divergent cultural circumstances has resulted in the test being judged suitable in

terms of content validity for the study of the general intellectual ability of European and Maori subjects. This judgement may or may not be supported by the criterion-related and construct validity studies.

(b) Criterion-Related Validity of the QT

In this research the term criterion-related validity covers empirical studies of the relationships between the QT and a number of independent external measures or criteria.

Criterion-related validities allow inferences to be made from an individual's test score as to his or her probable standing on some other variable known as the criterion. Texts on psychological measurement have often made a distinction between two types of criterion-related validity; concurrent validity and predictive validity (see Anastasi, 1961). This distinction is primarily procedural in relation to the time at which the criterion data are gathered. When the criterion data are collected at approximately the same time as the test data, the investigation is one of concurrent validation. When the criterion data are gathered at a later date, the study is concerned with predictive validity. Statements on concurrent validity concern the extent to which a test score may be used in estimating a subject's current standing on the criterion. Statements on predictive validity concern the extent to which a test score may be used in estimating a subject's future standing on the criterion. While concurrent validity reflects the status quo, predictive validity reflects the estimated outcomes under certain circumstances. The predictive validity model is critical to the use of a test for selection or assignment purposes, but the implementation of the model is dependent upon the replication of, or at least approximation to, the intervening conditions to which the predictive validity estimate applied.

McElwain and Kearney (1970) when discussing the validity of the QT detail a series of studies into both concurrent and predictive validity. Following the approach taken in the recent Standards for Educational and Psychological Tests (APA, 1974), these two types of

validity are considered jointly under the heading 'criterion-related validity' in the present study.

The QT criterion-related validity data were obtained from four class groups. Each class group contained both European and Maori subjects enabling separate and combined ethnic group analyses. The data consisted of performances on the Pacific Reasoning Series Test, the Otis Intermediate Test, the Progressive Achievement Tests of Reading Vocabulary and Reading Achievement, class performance data for Language Skills, Mathematics and Nature and Social Studies and a teacher ranking of 'General Ability'. Details relating to the criterion measures have been given in Chapter 4 in the section on 'Supplementary Data'.

Prior to the presentation and discussion of the results, it is appropriate to consider some of the conditional and limiting factors that must be borne in mind when considering any and all of the criterion-related validity coefficients.

First, it is unlikely that the conditions of a validation study would be repeated exactly, yet the logic of predictive validation does make this assumption. Thus, changed and changing conditions always limit the usefulness of such studies.

Second, the criterion measures must themselves possess certain characteristics if they are to be considered adequate. Mehrens and Lehmann(1973) state that criterion measures must be relevant, reliable and free from bias or contamination. It is important to consider the criterion measures employed in this research in relation to the above points.

The relevance of any criterion measure is, of course, a value judgement. Nonetheless, there is probably some scope for agreement. The Pacific Reasoning Series Test and the Otis Intermediate Test were adopted as acceptable group measures of general ability, the former

being a test with low English language content and the latter a test with high English language content.

The criterion measures of an educational achievement nature - the Progressive Achievement Tests and class performance data - are considered relevant to any consideration of the utility of the QT in educational settings with New Zealand subjects of a similar age or class level to those in the study.

The teacher assessment criterion was employed because teachers are, by virtue of their position, critical agents in the maintenance of current behaviours and in the development of new behaviours on the part of their students. A complex of evaluations, judgements, attitudes, perceptions and expectations on the part of the teacher can be subtly but effectively translated into patterns of behaviour between student and teacher that have causal consequences for the educational attainment of the student (see Brophy and Good, 1974). A teacher assessment of the 'General Ability' of students was, therefore, a relevant criterion.

The reliability of criterion measures is as important to any criterion-related validity relationship as is the reliability of the test or 'predictor' measure itself. The respective reliabilities will affect the maximum relationship that can theoretically be expected between any test and criterion measure.

Evidence on the reliability of the Pacific Reasoning Series Test, the Otis Intermediate Test and the Progressive Achievement Tests is to be found in the respective manuals. In all cases the evidence is of a statistical nature and is considered satisfactory. The reliability of the class performance data and the teacher assessment of 'General Ability' could not be estimated in the context of this investigation.

The biasing or contamination of criterion measures occurs when a criterion score is influenced by information concerning the 'predictor'

score. To avoid such contamination, QT data were secured until the collection of all criterion data was complete.

A third issue in criterion-related validity studies is the assumption that the samples upon which the studies are based are, in fact, representative of the population(s) for which later inferences are to be drawn. In practice it is difficult to claim that samples are totally representative. In the case of this research, class groups were not selected from schools known to be atypical, but nonetheless selective factors such as the need to obtain a Maori sample within each class group, considerations relating to the availability of class performance data and continuing school and teacher co-operation must place limits on the representativeness of the class group samples.

A fourth point which must be acknowledged is that there are restrictions in terms of sample size. At times it would seem criterion-related validity studies involve 'doing one's best with the data to hand'. While attempting to investigate criterion-related validity, if only in a limited way, may be preferable to accepting untested assertions, there is a contrary point of view. It is noted in the Standards for Educational and Psychological Tests (A.P.A., 1974) for instance, that -

".....'doing something' is not necessarily better than doing nothing; the results of an inadequate study may be quite misleading. Results of validation studies with severely restricted ranges or small N's are especially open to question" (pp.27-28).

These appeals for caution in the interpretation of criterion-related validity data certainly apply to these studies.

Tables 6.3.1 to 6.3.12 present for each class group as a whole and for the ethnic group samples in each class, the matrix of Product-Moment correlation coefficients between QT subtests, QT Total Score (at the test condition of the test-retest reliability studies) and all

criteria excepting the teacher ranking. Tables 6.3.13 to 6.3.20 present the rank correlation coefficients between QT Total Score, the criteria and the teacher ranking of 'General Ability' for the class group samples. The related descriptive statistics are presented in Appendix C.

Since the validity of the QT as a unified scale is the major concern of this aspect of the research, consideration of the results focuses primarily on the relationships between QT Total Score performance and the various criteria. The intercorrelations between QT subtests and the criteria (excepting teacher rankings) are reported and may be of interest. It should be noted that in the case of both Class Groups 1 and 2, class performance data on Language Skills and Nature and Social Studies are absent. The data were not consistently maintained in a form suitable for inclusion in the analyses.

To study more easily the QT Total Score, criterion-related validity data summary tables have been presented. Table 6.3.21 reports the coefficients from the class groups as a whole. Statistically significant and relatively consistent positive correlations are evident between the QT and the Pacific Reasoning Series Test, the QT and the class performance measure of Mathematics and the QT and the teacher ranking of 'General Ability'. Correlations with the Otis Intermediate Test were of a moderate positive order and statistically significant in two cases. QT correlations with the Progressive Achievement Tests and the class performance data on Language Skills and Nature and Social Studies were somewhat lower and generally non-significant.

TABLE 6.3.1 Product-Moment Correlation Coefficients for Class Group 1 (N = 27)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11
1. Knox											
2. Beads	.67										
3. Passalong	.29	.32									
4. Form Assembly	-.03	.48	.10								
5. Pattern Matching	.23	.61	.35	.41							
6. QT Total	.46	.76	.57	.63	.83						
7. P.R.S.T.	.13	.39	.43	.35	.63	.60					
8. Otis Int.	.15	.48	.22	.33	.55	.54	.59				
9. PAT RV	.08	.35	.29	.32	.32	.42	.56	.78			
10. PAT RC	-.09	.56	.28	.38	.42	.48	.63	.81	.77		
11. Mathematics	.23	.54	.34	.48	.71	.72	.75	.88	.73	.74	

Values of .49 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.2 Product-Moment Correlation Coefficients for European Subjects of Class Group 1 (N = 17)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11
1. Knox											
2. Beads	-.04										
3. Passalong	.38	.15									
4. Form Assembly	.07	.60	.14								
5. Pattern Matching	.38	.40	.14	.40							
6. QT Total	.57	.65	.48	.72	.76						
7. P.R.S.T.	.27	.06	.48	.41	.38	.49					
8. Otis Int.	.29	.47	.34	.53	.48	.66	.36				
9. PAT RV	.19	.38	.50	.41	.25	.51	.55	.69			
10. PAT RC	-.09	.68	.46	.54	.13	.49	.45	.65	.70		
11. Mathematics	.49	.52	.44	.62	.65	.85	.61	.87	.70	.58	

Values of .61 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.3 Product-Moment Correlation Coefficients for Maori Subjects of Class Group 1 (N = 10)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11
1. Knox											
2. Beads	.34										
3. Passalong	.14	.50									
4. Form Assembly	-.39	.28	.04								
5. Pattern Matching	-.02	.85	.57	.52							
6. QT Total	.25	.91	.68	.48	.93						
7. P.R.S.T.	-.10	.78	.39	.36	.90	.77					
8. Otis Int.	-.06	.55	.04	.13	.60	.44	.74				
9. PAT RV	-.20	.27	-.12	.34	.27	.20	.29	.73			
10. PAT RC	-.07	.59	.09	.50	.78	.63	.73	.83	.58		
11. Mathematics	-.23	.57	.24	.44	.76	.60	.82	.83	.65	.84	

Values of .77 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.4 Product-Moment Correlation Coefficients for Class Group 2 (N = 35)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11
1. Knox											
2. Beads	.17										
3. Passalong	.33	.22									
4. Form Assembly	.04	.09	.13								
5. Pattern Matching	.14	.30	.35	.16							
6. QT Total	.66	.65	.62	.43	.57						
7. P.R.S.T.	.47	.26	.25	.21	.23	.50					
8. Otis Int.	.42	.25	.04	.06	.17	.37	.51				
9. PAT RV	.34	.24	.10	-.08	.43	.36	.48	.67			
10. PAT RC	.20	.25	-.16	-.11	.17	.17	.46	.57	.63		
11. Mathematics	.59	.10	.07	.07	.18	.40	.41	.67	.63	.39	

Values of .43 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.5 Product-Moment Correlation Coefficients for European Subjects of Class Group 2 (N = 27)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11
1. Knox											
2. Beads	.16										
3. Passalong	.50	.29									
4. Form Assembly	.06	.15	.17								
5. Pattern Matching	.34	.31	.34	.16							
6. QT Total	.72	.63	.71	.44	.63						
7. P.R.S.T.	.45	.22	.31	.24	.29	.49					
8. Otis Int.	.41	.16	.11	.22	.22	.38	.53				
9. PAT RV	.32	.21	.16	-.03	.49	.36	.45	.65			
10. PAT RC	.05	.25	-.07	-.01	.36	.18	.45	.59	.66		
11. Mathematics	.60	-.13	.21	.05	.32	.36	.35	.69	.61	.40	

Values of .49 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.6 Product-Moment Correlation Coefficients for Maori Subjects of Class Group 2 (N = 8)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11
1. Knox											
2. Beads	.12										
3. Passalong	-.82	.00									
4. Form Assembly	.30	.23	-.37								
5. Pattern Matching	-.62	.31	.45	.14							
6. QT Total	.34	.89	-.21	.57	.32						
7. P.R.S.T.	.60	.48	-.48	.83	-.12	.74					
8. Otis Int.	.32	.45	-.29	.04	.33	.47	.25				
9. PAT RV	.32	.21	-.30	.58	-.56	.53	.62	.63			
10. PAT RC	.79	.11	-.93	.27	-.60	.21	.46	.09	.02		
11. Mathematics	.63	.43	-.38	.63	-.20	.71	.81	.63	.80	.21	

Values of .83 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.7 Product-Moment Correlation Coefficients for Class Group 3 (N = 34)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Knox													
2. Beads	.12												
3. Passalong	.02	-.17											
4. Form Assembly	-.19	.22	.43										
5. Pattern Matching	.30	.06	.32	.19									
6. QT Total	.45	.47	.46	.52	.72								
7. P.R.S.T.	.16	.30	.36	.34	.67	.69							
8. Otis Int.	.07	.18	.35	.43	.45	.53	.78						
9. PAT RV	-.02	.04	.35	.39	.35	.37	.63	.91					
10. PAT RC	.08	.21	.22	.39	.32	.45	.67	.90	.90				
11. Language Skills	.04	.30	.31	.46	.42	.56	.78	.89	.83	.79			
12. Mathematics	-.10	.28	.36	.43	.40	.48	.78	.89	.79	.81	.88		
13. Nature-Social St.	-.16	.02	.42	.45	.30	.32	.56	.79	.86	.79	.71	.79	

Values of .43 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.8 Product-Moment Correlation Coefficients for European Subjects of Class Group 3 (N = 18)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Knox													
2. Beads	-.27												
3. Passalong	-.16	-.09											
4. Form Assembly	-.31	.25	.33										
5. Pattern Matching	.24	-.04	.39	.23									
6. QT Total	.28	.50	.42	.50	.70								
7. P.R.S.T.	-.03	.42	.28	.39	.69	.72							
8. Otis Int.	-.08	.41	.11	.30	.31	.46	.81						
9. PAT RV	-.09	.26	.20	.33	.26	.38	.68	.92					
10. PAT RC	.08	.46	-.11	.26	.20	.44	.68	.88	.86				
11. Language Skills	-.19	.47	.13	.30	.38	.48	.77	.93	.89	.79			
12. Mathematics	-.17	.56	.12	.27	.42	.54	.86	.93	.80	.80	.93		
13. Nature-Social St.	-.04	.19	.24	.36	.47	.48	.76	.88	.93	.79	.87	.79	

Values of .59 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.9 Product-Moment Correlation Coefficients for Maori Subjects of Class Group 3 (N = 16)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Knox													
2. Beads	-.05												
3. Passalong	.21	-.24											
4. Form Assembly	-.06	.24	.48										
5. Pattern Matching	.39	.17	.26	.16									
6. QT Total	.56	.45	.51	.56	.75								
7. P.R.S.T.	.27	.08	.50	.37	.74	.68							
8. Otis Int.	.19	-.08	.54	.54	.56	.58	.78						
9. PAT RV	.14	-.18	.46	.42	.42	.41	.70	.93					
10. PAT RC	.14	-.08	.51	.50	.43	.49	.73	.93	.94				
11. Language Skills	.28	.04	.51	.67	.49	.66	.79	.87	.80	.81			
12. Mathematics	-.02	-.10	.60	.59	.40	.46	.71	.85	.79	.82	.80		
13. Nature-Social St.	-.14	-.06	.52	.49	.22	.30	.59	.79	.84	.84	.70	.88	

Values of .62 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.10 Product-Moment Correlation Coefficients for Class Group 4 (N = 31)
Between QT Subtests, QT Total Score and Criteria Data .

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Knox													
2. Beads	.25												
3. Passalong	.19	.28											
4. Form Assembly	.28	.43	.22										
5. Pattern Matching	.13	.23	.31	.44									
6. QT Total	.65	.76	.54	.70	.52								
7. P.R.S.T.	.38	.38	.49	.49	.48	.65							
8. Otis Int.	.12	.22	.38	.41	.39	.41	.78						
9. PAT RV	.27	.28	.50	.08	.10	.39	.59	.41					
10. PAT RC	.09	.10	.20	.49	.25	.31	.67	.85	.42				
11. Language Skills	.01	.15	.16	.21	.19	.20	.55	.78	.26	.68			
12. Mathematics	.30	.23	.50	.50	.43	.55	.87	.84	.42	.69	.72		
13. Nature-Social St.	.05	.17	.50	.19	.41	.33	.64	.73	.26	.44	.69	.81	

Values of .46 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.11 Product-Moment Correlation Coefficients for European Subjects of Class Group 4 (N = 18)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Knox													
2. Beads	.40												
3. Passalong	-.09	.45											
4. Form Assembly	.19	.27	.08										
5. Pattern Matching	.06	-.01	.19	-.19									
6. QT Total	.68	.86	.47	.50	.14								
7. P.R.S.T.	.38	.63	.52	.38	.30	.75							
8. Otis Int.	-.12	.24	.33	.37	.39	.30	.65						
9. PAT RV	.24	.61	.48	-.05	-.15	.49	.60	.12					
10. PAT RC	-.05	-.08	.06	.47	.04	.17	.53	.80	.23				
11. Language Skills	-.08	.26	.28	.24	.38	.28	.52	.84	.06	.59			
12. Mathematics	.10	.38	.56	.51	.42	.56	.78	.80	.18	.50	.73		
13. Nature-Social St.	-.21	.21	.61	.07	.63	.25	.40	.61	.03	.19	.69	.72	

Values of .59 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.12 Product-Moment Correlation Coefficients for Maori Subjects of Class Group 4 (N = 13)
Between QT Subtests, QT Total Score and Criteria Data

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Knox													
2. Beads	-.01												
3. Passalong	.45	-.01											
4. Form Assembly	.33	.65	.24										
5. Pattern Matching	.11	.44	.26	.78									
6. QT Total	.62	.64	.52	.88	.70								
7. P.R.S.T.	.38	.13	.30	.49	.41	.50							
8. Otis Int.	.30	.13	.24	.34	.23	.37	.80						
9. PAT RV	.27	-.48	.40	.07	.07	.06	.45	.62					
10. PAT RC	.22	.08	.22	.45	.30	.36	.78	.90	.64				
11. Language Skills	.09	-.11	-.08	.12	-.01	.01	.62	.77	.55	.80			
12. Mathematics	.48	-.48	.29	.41	.29	.42	.90	.81	.58	.87	.73		
13. Nature-Social St.	.28	.00	.18	.16	.08	.22	.72	.76	.39	.68	.74	.83	

Values of .68 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.13 Spearman-Rank Correlation Coefficients
for Class Group 1 (N = 27) Between QT Total Score,
Criteria Data and Teacher Ranking of 'General Ability'

	Teacher Ranking General Ability
QT Total	.64
P.R.S.T.	.66
Otis Int.	.91
PAT RV	.76
PAT RC	.81
Mathematics	.95

Values of .52 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.14 Spearman-Rank Correlation Coefficients for
European (N=17) and Maori (N=10) Subjects of Class Group 1 Between
QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'

	European ¹	Maori ²
	Teacher Ranking General Ability	Teacher Ranking General Ability
QT Total	.69	.40
P.R.S.T.	.50	.60
Otis Int.	.88	.65
PAT RV	.77	.74
PAT RC	.73	.60
Mathematics	.90	.83

1 Values of .67 and above significant, $p < .01$, two-tailed test.
2 Values of .79 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.15 Spearman-Rank Correlation Coefficients for Class Group 2 (N=35) Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'

	Teacher Ranking General Ability
QT Total	.37 ns
P.R.S.T.	.54 s
Otis Int.	.77 s
PAT RV	.68 s
PAT RC	.35 ns
Mathematics	.64 s

Correlations tested for significance by t for r_s , $p < .01$, two-tailed test.

s = significant, ns = non-significant.

TABLE 6.3.16 Spearman-Rank Correlation Coefficients for European (N=27) and Maori (N=8) Subjects of Class Group 2 Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'

	European ¹	Maori ²
	Teacher Ranking General Ability	Teacher Ranking General Ability
QT Total	.38	.44 ns
P.R.S.T.	.60	.36 ns
Otis Int.	.82	.47 ns
PAT RV	.63	.70 ns
PAT RC	.46	-.36 ns
Mathematics	.57	.90 s

1 Values of .52 and above significant, $p < .01$, two-tailed test.

2 Values of .88 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.17 Spearman-Rank Correlation Coefficients for Class Group 3 (N = 34) Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'

	Teacher Ranking General Ability	
QT Total	.53	s
P.R.S.T.	.76	s
Otis Int.	.89	s
PAT RV	.86	s
PAT RC	.83	s
Language Skills	.96	s
Mathematics	.81	s
Nature-Social St.	.77	s

Correlations tested for significance by t for r_s , $p < .01$, two-tailed test, s = significant.

TABLE 6.3.18 Spearman-Rank Correlation Coefficients for European (N = 18) and Maori (N = 16) Subjects of Class Group 3 Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'

	European ¹	Maori ²
	Teacher Ranking General Ability	Teacher Ranking General Ability
QT Total	.49	.50
P.R.S.T.	.10	.76
Otis Int.	.91	.84
PAT RV	.88	.79
PAT RC	.80	.83
Language Skills	.97	.95
Mathematics	.92	.79
Nature-Social St.	.81	.72

1 Values of .63 and above significant, $p < .01$, two-tailed test.

2 Values of .67 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.19 Spearman-Rank Correlation Coefficients for Class Group 4 (N = 31) Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'

	Teacher Ranking General Ability
QT Total	.53
P.R.S.T.	.78
Otis Int.	.92
PAT RV	.57
PAT RC	.73
Language Skills	.80
Mathematics	.86
Nature-Social St.	.82

Values of .49 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.20 Spearman-Rank Correlation Coefficients for European (N=18) and Maori (N=13) Subjects of Class Group 4 Between QT Total Score, Criteria Data and Teacher Ranking of 'General Ability'

	European ¹	Maori ²
	Teacher Ranking General Ability	Teacher Ranking General Ability
QT Total	.37	.57
P.R.S.T.	.58	.78
Otis Int.	.89	.88
PAT RV	.45	.58
PAT RC	.55	.90
Language Skills	.82	.76
Mathematics	.77	.91
Nature-Social St.	.78	.71

1 Values of .63 and above significant, $p < .01$, two-tailed test.

2 Values of .78 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.21 Summary of QT Criterion-Related Validity Coefficients - Class Groups 1 to 4

Class Group	Criterion:							
	P.R.S.T.	Otis Int	PAT RV	PAT RC	Lang Sk.	Maths.	Nat-Soc. Studies	T.R. Gen Ability
Class Group 1 N = 27	.60*	.54*	.42	.48	-	.72*	-	.64*
Class Group 2 N = 35	.50*	.37	.36	.17	-	.40	-	.37
Class Group 3 N = 34	.69*	.53*	.37	.45*	.56*	.48*	.32	.53*
Class Group 4 N = 31	.65*	.41	.39	.31	.20	.55*	.33	.53*

* Coefficient significant, $p < .01$, two-tailed test.

TABLE 6.3.22 Summary of QT Criterion-Related Validity Coefficients for European Subjects - Class Groups 1 to 4

Class Group	Criterion:							
	P.R.S.T.	Otis Int	PAT RV	PAT RC	Lang Sk.	Maths	Nat-Soc. Studies	T.R. Gen Ability
Class Group 1 N = 17	.49	.66	.51	.49	-	.85*	-	.69*
Class Group 2 N = 27	.49*	.38	.36	.18	-	.36	-	.38
Class Group 3 N = 18	.72*	.46	.38	.44	.48	.54	.48	.49
Class Group 4 N = 18	.75*	.30	.49	.17	.28	.56	.25	.37

* Coefficient significant, $p < .01$, two-tailed test.

TABLE 6.3.23 Summary of QT Criterion-Related Validity Coefficients for Maori Subjects -
Class Groups 1 to 4

Class Group	Criterion:							
	P.R.S.T.	Otis Int	PAT RV	PAT RC	Lang Sk.	Maths.	Nat-Soc. Studies	T.R. Gen Ability
Class Group 1 N = 10	.77*	.44	.20	.63	-	.60	-	.40
Class Group 2 N = 8	.74	.47	.53	.21	-	.71	-	.44
Class Group 3 N = 16	.68*	.58	.41	.49	.66*	.46	.30	.50
Class Group 4 N = 13	.50	.37	.06	.36	.01	.42	.22	.57

* Coefficient significant, $p < .01$, two-tailed test.

Tables 6.3.22 and 6.3.23 present the separate European and Maori criterion-related validity coefficients for the QT from the class groups. Inspection of these tables shows no distinctive or systematic ethnic group differences in the pattern of coefficients. The QT's positive relationship with the Pacific Reasoning Series Test performance is also apparent in the within ethnic group analyses.

The evidence from the combined ethnic group analyses suggests that the QT has a measure of validity in relation to the school based criteria of mathematics and teacher rankings of 'General Ability'. The QT correlates less well with criteria more dependent upon English language skills.

Australian data on the criterion-related validity of the QT does indicate a moderate degree of relationship with teacher assessments and scholastic attainment.

A QT validation study of some importance in attempting to assess the utility of the test is that of Smith (1966). Smith correlated QT results from samples of European and Aboriginal subjects with class performance data. He found few significant correlations with the European sample, but considerably more in the case of Aboriginal subjects. Interestingly the mathematics results of Aboriginal children correlated well with QT performance. However, it was found that the Otis Intermediate Test was a better predictor of class results than the QT for both European and Aboriginal subjects although the mean differences between the groups were far greater.

If the tables of correlations for the New Zealand class group samples are inspected it can be seen that the Otis Intermediate Test correlates extremely well with the class performance data and that this applies to European and Maori samples alike.

McElwain and Kearney (1970) in discussing Smith's findings argue that "These results are to be expected on theoretical grounds since the Otis Test and the criterion behaviour both contain educational material"

(p.129). The same argument would apply with regards the high criterion related validity coefficients obtained for the Otis Intermediate Test with European and Maori subjects.

It should also be noted that the Pacific Reasoning Series Test correlated with performances on the criterion measures at levels superior to the QT. Smith employed the Raven Progressive Matrices Test in his study and an inspection of his data shows that this group 'non-verbal' measure was also a better 'predictor' of the school based criterion measures than the QT.

The evidence, then, for the criterion-related validity of the QT in relation to scholastic achievement measures is not particularly strong. This is particularly so in situations where alternative scales with a degree of educational content could be used.

It is apparent from the evidence available from both Australian and New Zealand contexts that useful criterion measures for the QT other than conventional educational achievement data will need to be sought. Since the QT was designed as a relatively unique measure of general ability for use in culturally diverse or extreme situations, the above results are probably not surprising. Possibly there is more to be gained by validation research of the type undertaken by Kearney (1966a) amongst the Orokaiva people using a 'real-life' index of economic productivity - in this case the ownership of cash crop coconut trees.

The principle reason for the development of the QT was the need for a measure of general ability that would distinguish 'learners' from 'non-learners', enabling a judicious selection to be made of subjects for training in complex skills, in situations of considerable cultural diversity. The distinction drawn between 'learners' and 'non-learners' in situations where conventional criterion (educational attainment) may be unavailable, irrelevant or unreliable suggests that

the criterion-related validity of the QT should be studied in relation to the task of learning (measured by rate of acquisition, mastery, etc.) specified new skills.

A further dimension to the criterion-related validity of the QT concerns its use in situations where other instruments could not be administered. Australian research with the QT has employed an acculturation continuum in relation to Aboriginal samples. Groups are described as being of 'medium' or 'low' contact (with European society, institutions and ways). Thus Smith (1966) has argued in relation to his findings that -

"..... the Otis Test would be of limited value in Aboriginal groups of lower acculturation levels, because a minimal standard of (European) education must be reached before it may be effectively administered" (from McElwain and Kearney, 1970, p.129).

The implication of Smith's argument is that with subjects whose experiences with things European is strictly limited the QT could still be used, while other scales simply could not be administered.

McElwain and Kearney (1970) have extended this argument by first suggesting that the QT be used when there is a communication barrier between the tester and subject, when selection decisions must be made on the grounds of educability in situations where attained levels of education are largely fortuitous and when an indication of 'basic capacity' is required. Second, they suggest three strategies for the use of the QT along with additional instruments in non-western settings:-

- (a) Where contact is low, communication difficult and formal education absent, the QT is likely to be a useful predictor of the acquisition of new skills.
- (b) Where contact is high and formal school available, the usual group verbal intelligence and attainment tests would be more appropriate although the QT may provide useful supplementary data.

- (c) Where contact and formal education is uneven then the QT and achievement test results in combination might prove to be a suitable psychometric strategy.

Evidence on the contribution of QT data in these situations has yet to be produced. This will be necessary before more confident statements can be made on its validity in relation to relevant criteria.

(c) Construct Validity of the QT

Construct validity is concerned with the extent to which the results of a measure can be accounted for by certain explanatory constructs in psychological theory. If scores on a test are to be interpreted as measuring some theoretical variable then both the theoretical derivation and empirical support for the hypothesized variable must be considered.

Already it has been noted that the QT was developed as a measure of general intellectual ability akin to 'g' in Spearman's two-factor theory. Evidence for or against an interpretation of this nature of QT results for European and Maori subjects can be sought in a number of ways.

First, it has been argued that the types of items employed by the QT - its content - are known to be good measures of the hypothetical construct referred to as 'general ability'. Any decision as to the construct validity of the QT on the basis of its content remains, however, a judgement the test user must make in relation to particular conceptualizations of the construct of intelligence.

A second source of evidence bearing upon construct validity is the relationship between QT performance and age. It was noted when discussing McElwain and Kearney's (1970) view on the meaning of QT results that they expected and obtained evidence of a monotonic increase

in performance with age over the period of general biological maturation. The development of general ability was seen as being subject to this maturational process.

Evidence on the relationships between age and QT performance for New Zealand European and Maori subjects comes from the correlation coefficients and regression equations between the two variables. These data have been presented previously in Table 6.1.4 for European and Maori samples and the linear regression lines are plotted for each sample in Figure 6.1. Inspection of this figure reveals support for the hypothesis of a positive relationship between QT Score and age. For both European and Maori samples the slope of the regression line is ascending and the positive correlations between QT Score and age are statistically significant. The regression lines for the two samples are very similar.

When tests for the homogeneity of regression coefficients were undertaken it was found the slopes could be considered parallel, but were separated by a small absolute difference. These data support the view that QT performance increases with age during maturation and that the relationship is essentially the same for both the European and Maori samples.

Factor analytic studies are a third and very important source of evidence on the construct validity of the QT. Three sets of factor analytic data are presented:

- i) the factor analysis of QT subtest data by age groups;
- ii) the factor analysis of age and QT subtest data for the European and Maori samples; and
- iii) the factor analysis of age data, QT subtest data and the additional psychological test data from the class group studies.

For all analyses both Principal Factor and Varimax Rotated Factor solutions are presented. The Principal Factor solutions are 'direct'

solutions arising from the correlation matrices through the application of particular mathematical models. However, the interpretation of 'direct' factor solutions presents a number of difficulties. Bipolar factors occur frequently in all but the first factor. These are a result of the type of analysis and give rise to interpretation problems. In addition, Guilford (1968), an eminent factor theorist, has argued that incidental mathematical features inherent with these analysis procedures have been used as a basis for psychological theories. Criticising 'g' theory he claims that the presence of a major share of common variance in the first factor of a 'direct' solution is a function of factorial design and not necessarily a structural feature of human ability. Thus, further analysis by 'derived' factorial methods are recommended.

'Derived' solutions involve the adjustment of the frame of reference (the axes) of the original 'direct' factor analysis solution. Kaiser's (1959) Varimax solution with an orthogonal rotation has, therefore, been applied to the data.

i) Age Group Factor Analyses: These factor analyses are based on the age group intercorrelations amongst QT subtests presented in Part 1 of this chapter. The analyses presented are for the combined ethnic group samples at each of the six age levels. This is because the separate ethnic group sample sizes at each age group level were in most cases too small for the sample size guidelines for factor analysis proposed by Brislin, Lonner and Thorndike (1973).

From previous studies and comment on constructs underlying the QT it was anticipated that evidence would be found of a general factor, a spatial-perceptual factor and possibly a memory factor. Thus, three factors were sought in the analyses. However, the low eigenvalues and small percentage of variance accounted for by the third factor in each analysis indicated that there was little point in considering this factor further. Consequently, the data reported is for only the first two factors. The combined effect of not reporting the third factor and any factor

loadings less than .30 means that the squared and summed factor loadings for each variable will not equal the reported communalities.

Tables 6.3.24 to 6.3.35 present the Principal Factor Matrix and Varimax Rotated Factor Matrix, first two factors only, for the five QT subtests for the six age groups.

TABLE 6.3.24 Principal Factor Matrix for QT Subtests: Age Group 8:6-9:5(N=73)
(Loadings less than .30 not shown)

Variables	Factors		Commynality (h^2)
	1	2	
Knox	-.527		.369
Beads	-.382	.340	.269
Passalong	-.663		.477
Form Assembly	-.620		.413
Pattern Matching	-.719		.539
Total % of Variance	85.4	10.5	95.9
Eigenvalue	1.76	0.22	

TABLE 6.3.25 Varimax Rotated Factor Matrix for QT Subtests:
Age Group 8:6-9:5 (N=73)
(Loadings less than .30 not shown)

Variables	Factors		Commynality (h^2)
	1	2	
Knox	.555		.369
Beads		.493	.269
Passalong	.550	.368	.477
Form Assembly	.355		.413
Pattern Matching	.377	.402	.539
Total % of Variance	42.8	29.2	72.0
Eigenvalue	0.88	0.60	

TABLE 6.3.26 Principal Factor Matrix for QT Subtests: Age Group 9:6-10:5(N=90)
(Loadings less than .30 not shown)

Variables	Factors		Commynality (h^2)
	1	2	
Knox	-.385	-.424	.334
Beads	-.630		.402
Passalong	-.575		.415
Form Assembly	-.526		.340
Pattern Matching	-.653	.331	.541
Total % of Variance	77.7	15.9	93.6
Eigenvalue	1.57	0.32	

TABLE 6.3.27 Varimax Rotated Factor Matrix for QT Subtests:
Age Group 9:6-10:5 (N=90)
(Loadings less than .30 not shown)

Variables	Factors		Commynality (h^2)
	1	2	
Knox			.334
Beads	.476		.402
Passalong		.510	.415
Form Assembly	.543		.340
Pattern Matching	.595	.432	.541
Total % of Variance	46.1	27.1	73.2
Eigenvalue	0.94	0.55	

TABLE 6.3.28 Principal Factor Matrix for QT Subtests:
 Age Group 10:6-11:5 (N=194)
 (Loadings less than .30 not shown)

Variables	Factors		Commynality (h ²)
	1	2	
Knox	-.433		.193
Beads	-.487		.288
Passalong	-.381		.191
Form Assembly		(.286)	.136
Pattern Matching	-.652		.429
Total % of Variance	84.7	11.3	96.0
Eigenvalue	1.04	0.14	

TABLE 6.3.29 Varimax Rotated Factor Matrix for QT Subtests:
 Age Group 10:6-11:5 (N=194)
 (Loadings less than .30 not shown)

Variables	Factors		Commynality (h ²)
	1	2	
Knox			.193
Beads	.496		.288
Passalong			.191
Form Assembly		.361	.136
Pattern Matching	.485	.350	.429
Total % of Variance	49.8	27.3	77.1
Eigenvalue	0.62	0.34	

TABLE 6.3.30 Principal Factor Matrix for QT Subtests:
 Age Group 11:6-12:5 (N=122)
 (Loadings less than .30 not shown)

Variables	Factors		Commynality (h ²)
	1	2	
Knox			.410
Beads	.492	.614	.624
Passalong	.502		.273
Form Assembly	.516		.297
Pattern Matching	.911	-.397	.993
Total % of Variance	63.0	21.4	84.4
Eigenvalue	1.63	0.56	

TABLE 6.3.31 Varimax Rotated Factor Matrix for QT Subtests:
 Age Group 11.6-12:5 (N=122)
 (Loadings less than .30 not shown)

Variables	Factors		Commynality (h ²)
	1	2	
Knox			.410
Beads		.759	.624
Passalong	.469		.273
Form Assembly	.452	.303	.297
Pattern Matching	.972		.993
Total % of Variance	54.0	27.9	81.9
Eigenvalue	1.40	0.73	

TABLE 6.3.32 Principal Factor Matrix for QT Subtests:
 Age Group 12:6-13:5 (N=138)
 (Loadings less than .30 not shown)

Variables	Factors		Communality (h ²)
	1	2	
Knox	-.436		.277
Beads	-.559	-.343	.434
Passalong	-.379		.231
Form Assembly	-.482		.289
Pattern Matching	-.739		.578
Total % of Variance	78.7	13.9	92.6
Eigenvalue	1.42	0.25	

TABLE 6.3.33 Varimax Rotated Factor Matrix for QT Subtests:
 Age Group 12:6-13:5 (N=138)
 (Loadings less than .30 not shown)

Variables	Factors		Communality (h ²)
	1	2	
Knox		.320	.277
Beads		.621	.434
Passalong	.369		.231
Form Assembly	.502		.289
Pattern Matching	.599	.454	.578
Total % of Variance	43.7	39.7	83.4
Eigenvalue	0.79	0.72	

TABLE 6.3.34 Principal Factor Matrix for QT Subtests:
 Age Group 13:6-14:5 (N=83)
 (Loadings less than .30 not shown)

Variables	Factors		Commynality (h ²)
	1	2	
Knox	-.496		.340
Beads	-.476	.302	.322
Passalong	-.616		.453
Form Assembly	-.466		.272
Pattern Matching	-.672		.489
Total % of Variance	81.1	10.0	91.1
Eigenvalue	1.52	0.19	

TABLE 6.3.35 Varimax Rotated Factor Matrix for QT Subtests:
 Age Group 13.6-14:5 (N=83)
 (Loadings less than .30 not shown)

Variables	Factors		Commynality (h ²)
	1	2	
Knox			.340
Beads		.383	.322
Passalong	.598		.453
Form Assembly		.463	.272
Pattern Matching	.489	.470	.489
Total % of Variance	38.7	33.9	72.6
Eigenvalue	0.73	0.64	

The Principal Factor analyses all show clear evidence of an important and significant first factor (based on an eigenvalue criterion of > 1). All subtests load in this first factor for four of the six age levels with factor loadings of .30 or more. In the case of the Age Group 10:6-11:5 the Form Assembly Subtest fails to load appreciably in the first factor and for Age Group 11:6-12:5 the Knox Subtest makes only a minor contribution to the first factor. The most parsimonious interpretation of the first factor from the Principal Factor analyses is in terms of a 'general ability' or 'g' factor.

The second factor accounts for far less variance than the first factor for all the age groups. In four cases the Beads Subtest evidences a factor loading greater than .30 which may suggest a relatively unique contribution on its part. The nature of the factor is not clear, however, since other subtests implicated in the second factor do not appear in any consistent manner. Interpretation is also confused by the bipolar loadings in some instances.

The Varimax Rotated Factor Matrices present evidence of a more varied factorial structure across the six age groups. While there is still clear evidence of an important first factor (accounting for between 38 and 53 percent of the variance) which is interpreted as a general factor, the pattern of subtest loadings is not altogether uniform.

Table 6.3.36 presents the frequencies with subtests loaded into the first or second factors for the six age groups. The first factor would appear to be associated primarily with the Pattern Matching, Passalong and Form Assembly Subtests. The Pattern Matching Subtest is based on Koh's (1923) Block Design Test which has been used in the Wechsler scales and is, according to Cohen (1959), a useful measure of

TABLE 6.3.36 Frequencies of Factor Loadings $> .30$ on QT Subtests for the Six Age Groups (Varimax Rotated Analyses)

	1	2	Absence of loading $> .30$
Knox	1	1	4
Beads	2	4	
Passalong	4	2	1
Form Assembly	4	3	
Pattern Matching	6	5	

both 'g' and a 'perceptual organisation' factor.

The Passalong Subtest was for Alexander (1932) a measure of general intelligence. Sliding block puzzles of a similar form have appeared for many years as popular children's games and more serious attention has been given to the complex qualities of such puzzles in terms of their mathematical properties (Gardner, 1964).

The Form Assembly Subtest was also found to load with some consistency into the first factor. Like other form board type tests it was anticipated that this subtest would tap spatial-ability skills. Cronbach (1970) has pointed to the use of form boards in assessing space-relations and spatial ability. Nevertheless, this may not mean that more generalized reasoning skills are not also strongly implicated in such tasks. Recently a paper and pencil group form of this test was administered as part of a battery of tests to Cook Island Maori subjects (St. George and St. George, 1974). Here, factor analysis showed this test to be more closely allied to the 'non-verbal general ability' measures than to those whose content more obviously reflected 'spatial organisation', 'perceptual speed' and 'perceptual discrimination'.

The evidence from what is known or can be surmised on the common content of these three subtests lends weight to a general ability factor interpretation of the first factor.

Table 6.3.36 shows the second factor to be primarily associated with the Pattern Matching, Beads and Form Assembly Subtests. The variance accounted for by the second factor is between 27 and 40 percent.

It was noted that Cohen (1959) saw block design tasks as tapping a 'perceptual organisation' factor as well as 'g'. Furthermore, form board tasks are believed to tap spatial-perceptual skills. The content of the Beads Subtest tasks would also suggest the use of perceptual discrimination and spatial relation skills in addition to memory. The element most common amongst these second factor subtests not accounted for in the first factor appears to be a spatial factor involving both perceptual discrimination and organisation components. This second factor might tentatively be identified as being of a 'k' or spatial nature.

While it is possible to suggest the above interpretations by compressing the data across age groups, it must be recognised that it does not fully reflect the variation in the data overall. Table by table inspection of the factor loadings does highlight the presence of some subtests at particular age levels and their absence at others. In addition, subtests do not always contribute consistently to the first or second factors. Perhaps all that can be said at this stage is that empirically all five subtests of the QT are making a contribution to the underlying construct(s), but are tending to do so in a differing manner at the various age levels.

The absence of an invariant pattern of subtest factor loadings across age may indeed highlight the changing and complex nature of intellectual development across the age range studied. The Piagetian model of cognitive development (Piaget, 1950) suggests that this is indeed the case. Therefore, the assumption that measures of intelligence such as the QT measure and reflect the same intellectual processes for all subjects, irrespective of age or stage of intellectual develop-

ment may not be wholly justified.

At present the data does not allow any clear decision to be made as to the reasons for the changing factor structures in this QT data. Viewed one way, some consistencies do emerge in the factor analysis data. However, viewed another way, the variation is apparent and this in itself is intriguing. A better understanding of the construct validity of the QT will depend on research employing both enlarged samples, which would allow for greater differentiation in terms of sample variables, and appropriate reference tests keyed to an explicit theory on the nature of intellectual growth and functioning.

ii) Age and QT Subtest Factor Analyses for European and Maori Samples

Principal Factor and Varimax Rotated Factor analyses were undertaken upon age and QT subtest data for the separate samples of European and Maori subjects. Here again three factors were sought. However, the third factor in each analysis was not of statistical or practical significance. As with the previous age group analyses the third factors are not reported. Similarly, the communalities will not equal the squared and summed factor loadings for the reasons outlined earlier. Descriptive data and the matrices of correlation coefficients are presented in Tables 6.3.37 to 6.3.40. Tables 6.3.41 to 6.3.44 report the Principal Factor and Varimax Rotated Factor matrices for the two groups. The Principal Factor solution shows clearly the presence of a significant first factor for both ethnic groups that loads on all five QT subtests and age. This factor is interpreted for both groups as being of a general intelligence or 'g' nature. The loading of age in this factor supports this interpretation inasmuch as 'g' is presumed to be determined largely by maturation.

The second factor loadings differ for the ethnic groups. For European subjects this factor loads primarily on the Knox Subtest and for Maori subjects on the Beads Subtest. In both cases the factors are not significant (Eigenvalues < 1), but the fact that different subtests are implicated might suggest some element of difference in cognitive organisation.

TABLE 6.3.37 Means and Standard Deviations for Age, QT Subtests and QT Total Score: European Sample (N=397)

	Mean	S.D.
Age (in months)	137.74	17.22
Knox	7.67	2.16
Beads	4.92	2.49
Passalong	5.00	1.30
Form Assembly	9.50	1.52
Pattern Matching	13.72	1.77
QT Total Score	40.80	6.36

TABLE 6.3.38 Product-Moment Correlation Coefficients Between Age, QT Subtests and QT Total Score: European Sample (N=397)

	1	2	3	4	5	6
1. Age						
2. Knox	.17					
3. Beads	.35	.33				
4. Passalong	.34	.33	.33			
5. Form Assembly	.26	.23	.27	.30		
6. Pattern Matching	.35	.36	.40	.39	.34	
7. QT Total Score	.42	.69	.75	.63	.58	.72

Values of .13 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.39 Means and Standard Deviations for Age, QT Subtests and QT Total Score: Maori Sample (N=303)

	Mean	S.D.
Age (in months)	139.92	18.90
Knox	7.85	2.23
Beads	4.57	2.41
Passalong	4.59	1.39
Form Assembly	9.12	1.72
Pattern Matching	12.87	2.45
QT Total Score	38.98	6.85

TABLE 6.3.40 Product-Moment Correlation Coefficients Between Age, QT Subtests and QT Total Score: Maori Sample (N=303)

	1	2	3	4	5	6
1. Age						
2. Knox	.22					
3. Beads	.42	.22				
4. Passalong	.28	.18	.26			
5. Form Assembly	.22	.19	.28	.21		
6. Pattern Matching	.46	.31	.39	.42	.43	
7. QT Total Score	.50	.61	.69	.56	.61	.79

Values of .15 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.41 Principal Factor Matrix for Age and QT Subtests:
European Sample (N=397)
(Loadings less than .30 not shown)

Variables	Factors		Commynality (h ²)
	1	2	
Age	-.519		.338
Knox	-.574	-.469	.549
Beads	-.640		.521
Passalong	-.596		.377
Form Assembly	-.478		.252
Pattern Matching	-.655		.436
Total % of Variance	81.8	12.3	94.1
Eigenvalue	2.02	0.30	

TABLE 6.3.42 Varimax Rotated Factor Matrix for Age and QT Subtests:
European Sample (N=397)
(Loadings less than .30 not shown)

Variables	Factors		Commynality (h ²)
	1	2	
Age	.502		.338
Knox		.694	.549
Beads	.320		.521
Passalong	.516		.377
Form Assembly	.447		.252
Pattern Matching	.520	.306	.436
Total % of Variance	45.7	30.1	75.8
Eigenvalue	1.13	0.75	

TABLE 6.3.43 Principal Factor Matrix for Age and QT Subtests:
Maori Sample (N=303)
(Loadings less than .30 not shown)

Variables	Factors		Commynality (h^2)
	1	2	
Age	-.597		.422
Knox	-.373		.139
Beads	-.601	.365	.508
Passalong	-.468		.228
Form Assembly	-.508		.395
Pattern Matching	-.844		.812
Total % of Variance	81.6	11.7	93.3
Eigenvalue	2.05	0.29	

TABLE 6.3.44 Varimax Rotated Factor Matrix for Age and QT Subtests:
Maori Sample (N=303)
(Loadings less than .30 not shown)

Variables	Factors		Commynality (h^2)
	1	2	
Age	.393	.511	.422
Knox			.139
Beads		.648	.508
Passalong	.388		.228
Form Assembly			.395
Pattern Matching	.783		.812
Total % of Variance	42.8	34.5	77.3
Eigenvalue	1.07	0.86	

The rotated factor analyses confirm the importance of the first factor in that it accounts for over 40 percent of the variance for both ethnic groups. Age loads into this factor for both groups, but the contribution of QT subtests varies. With European subjects the Beads, Passalong, Form Assembly and Pattern Matching subtests all contribute suggesting a general intelligence factor. With Maori subjects it is only the Passalong and Pattern Matching subtests which clearly contribute to the first factor. It was, however, noted earlier that items of the type contained in these two subtests are said to be good measures of general intelligence. This would lend some support to a general intelligence factor interpretation for the Maori sample also.

The second factors extracted in the rotated analyses for both ethnic groups were insignificant in terms of the eigenvalue criterion. Nevertheless, it is interesting to note the ethnic group differences in this factor. In the case of the European sample the predominant loading was on the Knox Subtest with a much smaller loading on the Pattern Matching Subtest whereas with the Maori sample the major loadings occurred on the Beads Subtest and Age.

In the case of the European sample the loading on the Knox Subtest suggests a short-term memory factor in that the subject must recall the block tapping sequence. The earlier suggestion that the Pattern Matching type items tap 'g' and perceptual organization do not, however, fit particularly well with this conceptualization.

The Maori sample second factor loadings on Beads and Age is unusual. It suggests an age related ability specific to the performance of Maori subjects on bead-chaining type items. Sattler (1965) in a logical analysis of the 1960 Stanford-Binet contents suggests that this type of item is one of 'visual memory'. A conceptualization of this kind might be appropriate if there were additional evidence in support of

age and ethnic group variables influencing an ability of this nature.

As well as considering the ways in which subtests loaded into the factors identified by the analyses, it should be noted that both the Knox and Form Assembly Subtests actually failed to load in any appreciable way in the Varimax solution for Maori subjects.

These differences in the first and second factor loadings between the ethnic groups certainly indicates that a single set of hypothesised constructs cannot be advanced as an explanation of the meaning of QT results even though the ethnic group differences in actual levels of performance are small. While a general intelligence factor can be proposed as the most parsimonious construct underlying QT performance on the basis of variance accounted for in the first factor, there is also a suggestion of some ethnic group differences in the organisation and utilisation of intellectual abilities in relation to the items presented in the test. Consideration of further research on these findings will be taken up later.

iii) Factor Analysis of QT Subtests and Four Additional Measures

Following the factor analytic studies of the QT with various Australian age and ethnic group samples, McElwain and Kearney (1970) suggest that the factorial structure of the QT needs to be further explored through studies employing a wider range of psychological tests.

A limited study of this nature was possible by combining the New Zealand class group samples. Data were available for 127 subjects on the following variables; age, the five QT subtests, the Pacific Reasoning Series Test, the Otis Intermediate Test and the Progressive Achievement Tests - Reading Comprehension and Reading Vocabulary. With ten variables the sample of 127 subjects is, however, only a 'bare minimum sample' (see Brislin, Lonner and Thorndike, 1973, p.279).

Table 6.3.45 presents the summary statistics for the ten variables and Table 6.3.46 the matrix of correlation coefficients amongst the variables. The Principal Factor Matrix is given in Table 6.3.47 and the Varimax Rotated Factor Matrix in Table 6.3.48.

The first factor in the rotated solution involves the four group tests. The Otis and the two Progressive Achievement Tests are highly verbal in nature, requiring skill in English. All three have been found to be good measures of educational attainment in New Zealand (see Elley and Reid, 1969). The Pacific Reasoning Series Test, while not stressing English language skills to the same degree, requires a basic understanding of English since it is administered via English language instructions and a proportion of its items involve the solving of Roman alphabet and English word series problems. In other research it has been found to correlate positively with measures of English attainment (see St. George and St. George, 1974). This first factor is interpreted as being a verbal-educational factor (v:ed) and one which most probably has a high English language skill component. The lower factor loading of the Pacific Reasoning Series Test into this factor would support this view.

The loadings in the second factor are upon age, the Knox, Passalong and Pattern Matching subtests of the QT, the Pacific Reasoning Series Test and, to a lesser extent, the Otis and Reading Vocabulary Tests. The involvement of three of the QT subtests and age point towards the factor being of a reasoning ability nature that is related to maturation. This argument could accommodate the loading on the Pacific Reasoning Series Test and the Otis Test if the loadings were viewed as reflecting the reasoning components of performance as distinct from the English language skills. The same argument if applied to the Reading Vocabulary Test loading would need to be premised on the view that an element of performance on this test depends upon a good

TABLE 6.3.45 Means and Standard Deviations for Age, QT Subtests and Four Additional Tests: Combined Factor Analysis Sample (N=127)

	Mean	S.D.
Age (in months)	133.13	12.93
Knox	7.53	2.11
Beads	4.24	2.17
Passalong	4.80	1.33
Form Assembly	9.20	1.59
Pattern Matching	13.46	1.83
P.R.S.T.	25.13	10.86
Otis Int.	33.94	15.44
PAT RV	30.40	13.10
PAT RC	17.18	8.56

TABLE 6.3.46 Product-Moment Correlation Coefficients Among Age, QT Subtests and Four Additional Tests: Combined Factor Analysis Sample (N=127)

	1	2	3	4	5	6	7	8	9	10
1. Age										
2. Knox	.19									
3. Beads	.27	.14								
4. Passalong	.20	.25	.21							
5. Form Assembly	.10	.05	.30	.24						
6. Pattern Matching	.29	.24	.34	.40	.31					
7. P.R.S.T.	.15	.35	.37	.42	.35	.54				
8. Otis Int.	.24	.23	.33	.34	.33	.47	.69			
9. PAT RV	.14	.19	.27	.38	.22	.40	.59	.72		
10. PAT RC	-.13	.06	.23	.15	.30	.29	.56	.75	.67	

Values of .23 and above significant, $p < .01$, two-tailed test.

TABLE 6.3.47 Principal Factor Matrix for Age, QT Subtests and Four Additional Tests: Combined Factor Analysis Sample (N=127)
(Loadings less than .30 not shown)

Variables	Factors			Commynality (h^2)
	1	2	3	
Age		.439		.251
Knox	-.306			.217
Beads	-.446			.322
Passalong	-.472			.315
Form Assembly	-.429		.384	.337
Pattern Mátching	-.612	.309		.473
P.R.S.T.	-.814			.672
Otis Int.	-.866			.771
PAT RV	-.765			.647
PAT RC	-.762	-.640		.997
Total % of Variance	74.2	19.2	6.7	100.0
Eigenvalue	3.7	0.9	0.3	

TABLE 6.3.48 Varimax Rotated Factor Matrix for Age, QT Subtests and Four Additional Tests: Combined Factor Analysis Sample (N=127)
(Loadings less than .30 not shown)

Variables	Factors			Commynality (h^2)
	1	2	3	
Age		.446		.251
Knox		.449		.217
Beads			.482	.322
Passalong		.480		.315
Form Assembly			.539	.337
Pattern Matching		.490	.396	.473
P.R.S.T.	.592	.461	.330	.672
Otis Int.	.759	.337		.771
PAT RV	.726	.323		.647
PAT RC	.954			.997
Total % of Variance	51.4	28.0	20.6	100.0
Eigenvalue	2.6	1.4	1.0	

general knowledge of words and their meaning.

The third factor extracted in this analysis features the Beads, Form Assembly and Pattern Matching Subtests of the QT and the Pacific Reasoning Series Test. This factor loads highest on the Form Assembly Subtest. In the Australian factor analytic studies this subtest has been interpreted as a measure of spatial ability (McElwain and Kearney, 1970) and in the New Zealand age group factor analyses it was also interpreted in a similar manner. Both the Beads and Pattern Matching subtests require spatial/perceptual skills and the Pacific Reasoning Series Test contains many items of a figural reasoning nature. This third factor is, therefore, interpreted as a spatial factor.

This analysis with a small set of additional measures tends to support other factorial evidence on the hypothetical constructs underlying the QT. As is obvious from its content, the QT does not tap verbal-educational skills in any direct sense. There is some evidence, however, of an important reasoning factor underlying QT performance and, as might be expected from the test content, a subsidiary spatial component.

The results of this analysis are not unlike those of Richardson (1970). He factor analysed QT data, group test data, education level and age data from a small sample of Papua New Guinean armed service recruits. The first factor loaded predominantly on measures of verbal skill and educational attainment. While Richardson chose to call this an acculturation factor, it also appears to be very much of 'verbal' nature. The second factor, with loadings on the Knox, Beads and Pattern Matching subtests and two group ability measures, was interpreted as an intelligence factor which was independent of cultural factors. If read as a factor somewhat distinct from verbal skills and educational attainment, its interpretation would be close to the reasoning ability

factor proposed above in relation to the present analysis.

Richardson's third factor loaded on the Knox, Beads and Passalong subtests and he suggested it might be a speed factor. It would appear to be a different type of factor to the spatial factor suggested in this analysis. Nonetheless, the concordance in the nature of the first two factors in these analyses in contrasting cultural contexts is interesting. The data does suggest that QT performance is based upon a set of rather general reasoning abilities that are relatively independent of those skills and attainments of a more verbal and scholastic nature.

PART 4 DATA ON QT ITEMS

Data are presented on item difficulty, item discrimination indices, item correlations with subtest score and QT Total Score. A factor analysis of item intercorrelations is also presented.

(a) Item Difficulty

Tables 6.4.1 to 6.4.5 report the proportion of the Total Sample, European and Maori samples passing each item of the subtests. Figures 6.16 to 6.18 summarize these data graphically.

From Figure 6.16 for the Total Sample it can be seen that the level of ease falls relatively smoothly for the Knox, Beads, Passalong and Form Assembly subtests. The items of the Pattern Matching subtest are an exception to this pattern and the items stand out as being particularly easy.

Performance on Knox subtest, the initial subtest, shows a tendency to improve over the first six items. This may detract somewhat from the psychometric properties of the subtest. After all, over 80% of subjects passed all six items. Nonetheless, the introductory nature of the subtest to the QT and its role in inducing the 'set to imitate' may favour the retention of these items. The introduction of longer block tapping sequences with Item 7 initiates a more conventional difficulty gradient.

The Beads subtest item difficulty order evidences a 'saw-tooth' effect which was also apparent in the equivalent Australian data. These first eight items are in pairs requiring the same number of beads and the item difficulty data show that for the first three pairs of items the second item is easier than the first. For McElwain and Kearney (1970) -

"A possible explanation is that the number of beads used creates a set which interferes with the first performance where there is an increase in the number of beads required" (p.88).

The Passalong subtest with its relatively small number of items shows

TABLE 6.4.1 Proportion of Total Sample, European and Maori
Samples Passing Each Item of Knox Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Total	700	84	83	91	91	85	97	48	68	26	40	38	14	03	05	01
European	397	83	83	93	90	82	97	47	68	26	39	38	13	04	04	01
Maori	303	84	83	89	91	87	97	50	68	27	42	38	16	03	06	02

Decimal points omitted

TABLE 6.4.2 Proportion of Total Sample, European and Maori
Samples Passing Each Item of Beads Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10
Total	700	81	89	43	64	34	41	35	32	43	16
European	397	81	89	46	65	37	43	35	35	45	17
Maori	303	81	89	39	63	29	39	34	28	39	14

Decimal points omitted

TABLE 6.4.3 Proportion of Total Sample, European and Maori
Samples Passing Each Item of Passalong Subtest

Sample	N	1	2	3	4	5	6	7
Total	700	99	99	88	65	61	51	18
European	397	99	99	90	71	68	53	18
Maori	303	99	99	85	58	52	48	17

Decimal points omitted

TABLE 6.4.4 Proportion of Total Sample, European and Maori
Samples Passing Each Item of Form Assembly Subtest

Samples	N	1	2	3	4	5	6	7	8	9	10	11	12	13
Total	700	96	96	98	98	89	97	68	81	54	51	06	36	61
European	397	97	97	99	98	93	98	72	83	52	56	07	38	60
Maori	303	94	96	97	97	84	96	64	79	56	45	04	33	63

Decimal points omitted

TABLE 6.4.5 Proportion of Total Sample, European and Maori Samples Passing Each Item of Pattern Matching Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Total	700	99	99	98	99	99	99	99	91	91	92	94	79	72	53	71
European	397	99	99	99	100	99	99	100	94	91	95	95	85	80	57	76
Maori	303	99	97	97	98	98	98	99	87	90	89	92	71	61	46	64

Decimal points omitted

a rapid but ordered rise in difficulty. In terms of item format Items 1 and 2, 3 and 4, and 5 and 6 form convenient pairs. It can be seen that almost all subjects passed Items 1 and 2. While 88% of subjects passed Item 3, only 65% passed Item 4. The latter item is the first in which the subject must contend with an oblong blue tile which makes it necessary to move one or either pair of blue tiles from a position of one above the other to a position side by side each other in order to continue to move the large red tile. Inspection of the illustration of this item in Figure 5.8 shows the necessity of this move. This is a move which is critical to the solution of the remaining subtest items.

Items 5 and 6 are passed by 61% and 51% of subjects respectively and is followed by a considerable rise in difficulty for Item 7. Only 18% of subjects successfully solved this item within the prescribed time limit.

In the case of the Form Assembly subtest 89% or more of subjects passed all of the first six items. Of these items the first three require 'one-piece solutions' and the second three 'two one-piece solutions'. The difficulty of the items in this subtest increases with the introduction of the 'two-piece solution' requirement for Items 7 to 10. The first 'three-piece solution' item - Item 11 - is passed by only 6% of subjects. However, there is a steady increase in the proportion of

FIGURE 6.16

Proportion of Total Sample Passing Each Item (N = 700)

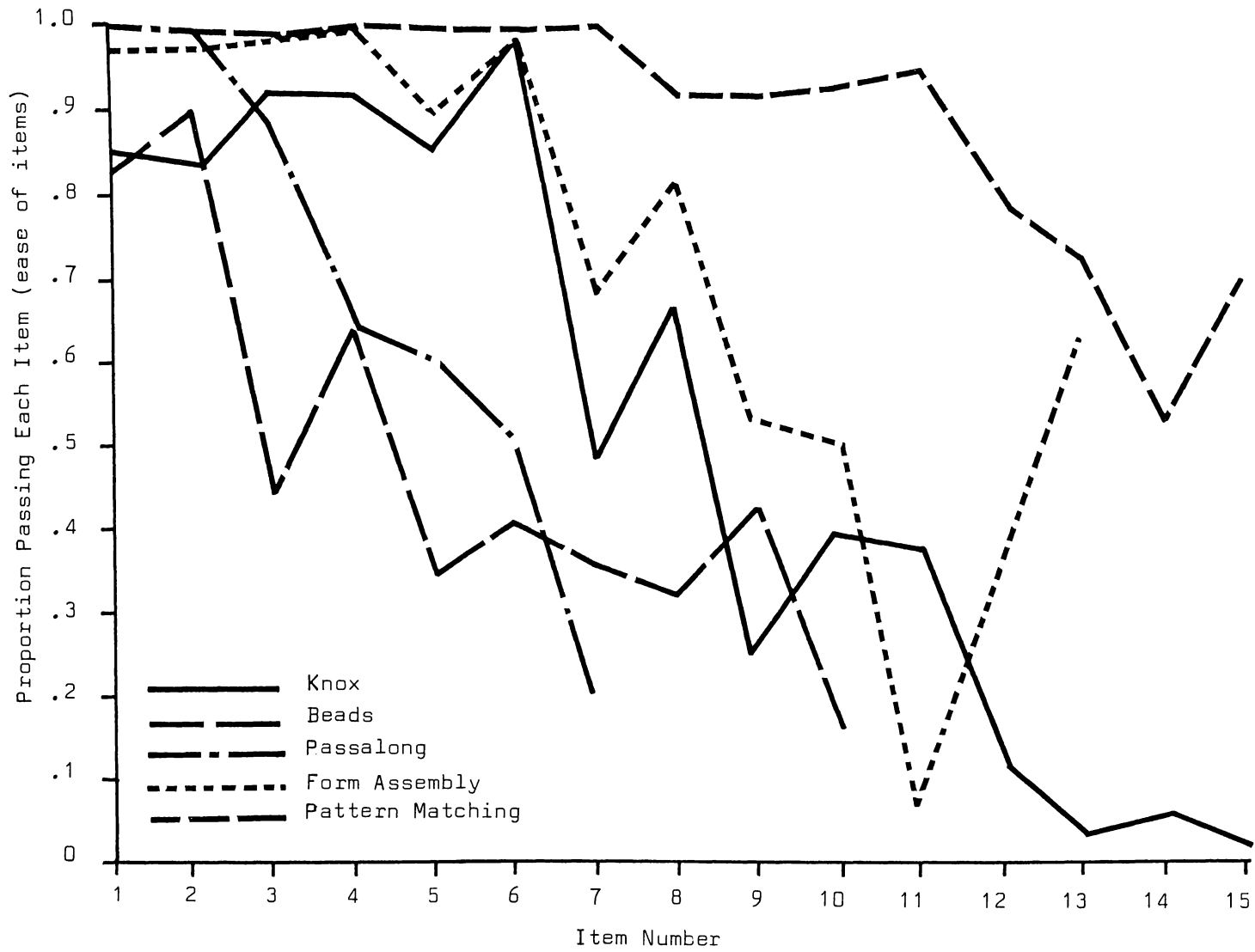


FIGURE 6.17 Proportion of European Sample Passing Each Item (N = 397)

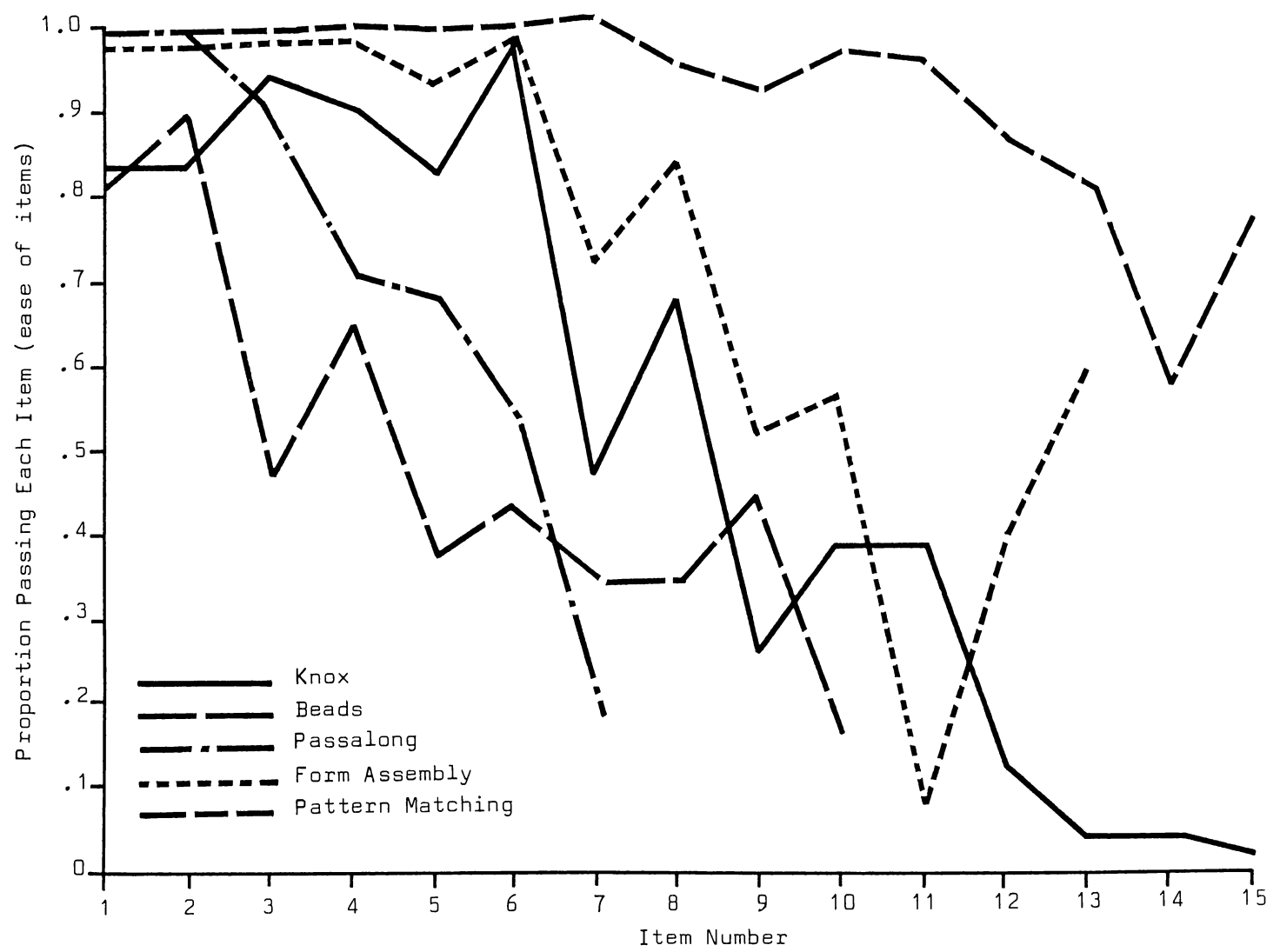
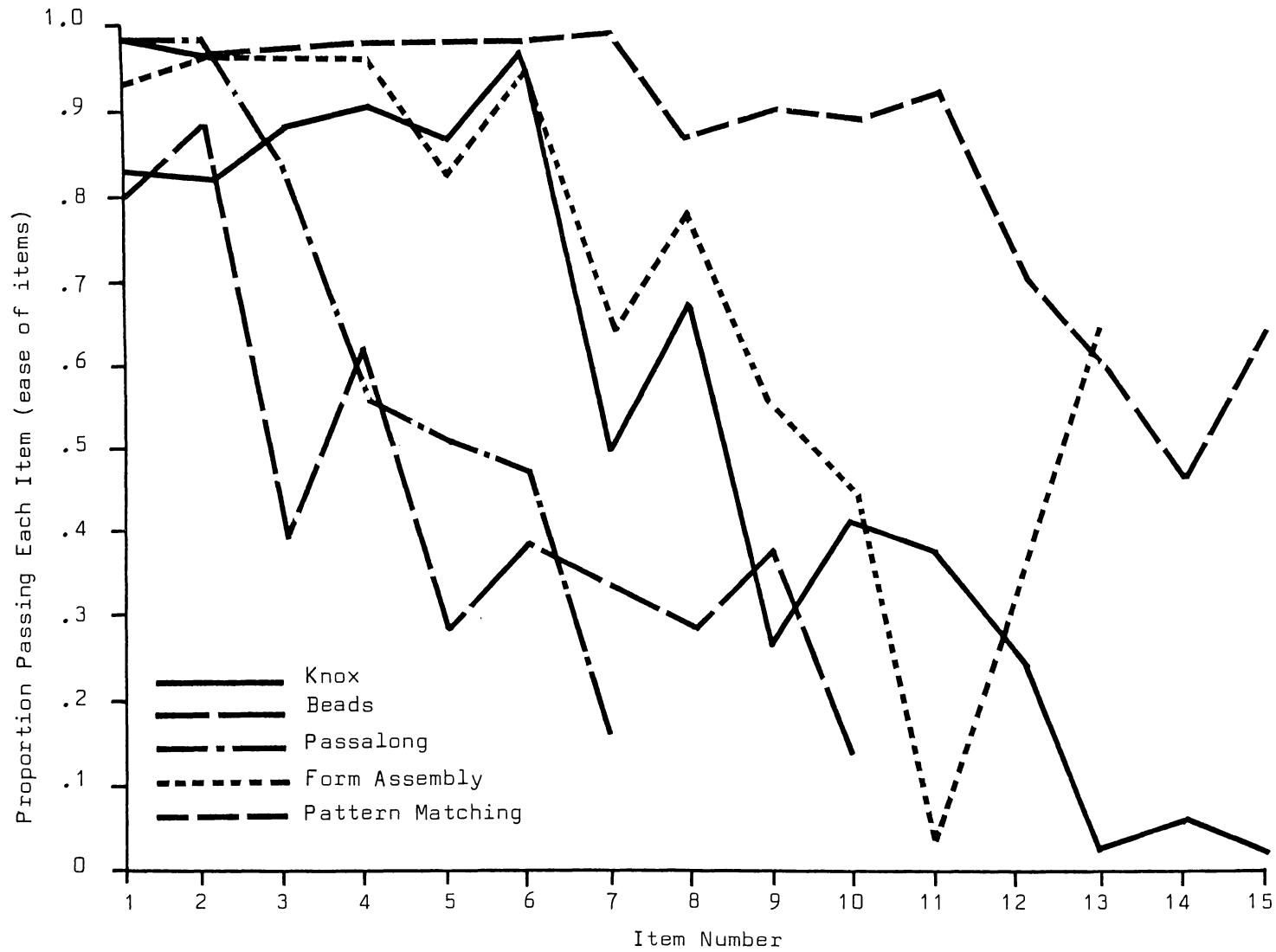


FIGURE 6.18 Proportion of Maori Sample Passing Each Item (N = 303)



subjects passing Items 12 and 13. It must be remembered that Item 13 is always administered. This in itself will have increased the proportion of subjects passing this item.

A very high proportion of subjects, some 90%, passed the first eleven items of the Pattern Matching subtest. Furthermore, over half the Total Sample passed the remaining four items. Clearly this subtest could be improved psychometrically by the exclusion of a number of the very easy items and the inclusion of a number of more difficult items.

Figures 6.17 and 6.18 present graphs of the proportion of European and Maori subjects passing each item. Inspection of the two Figures shows that the proportion of subjects passing each item in the five QT subtests are very similar. The pattern of proportions of European and Maori subjects passing each item follow closely those noted in relation to the Total Sample data, and the comments already made apply equally to the separate ethnic group data.

The proportions of Total Sample, European sample and Maori sample subjects passing each item of the five QT subtests by age groups are reported in Appendix D.

(b) Item Discrimination Indices

The item discrimination index-D (Findley, 1956) was calculated for each item of the QT on the basis of the upper and lower 27 percent of QT Scores in the Total Sample. The procedures set out in Stanley and Hopkins (1972) were employed. Englehart (1965) in comparing several item discrimination index methods has found D to be very effective in identifying poor items.

The QT item discrimination indices for the Total Sample are presented in Table 6.4.6. Stanley and Hopkins state that:-

"Items that yield a discrimination index of .4 or more are high in discrimination. Those with D-values below .2 are low....." (p.269).

TABLE 6.4.3. QT Item Discrimination Indices-D, Based on Upper and Lower 27% of Total Sample

Knox Subtest															
Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
D-index	.12	.20	.20	.20	.22	.08	.36	.35	.23	.39	.41	.29	.07	.11	.02

Beads Subtest										
Item	1	2	3	4	5	6	7	8	9	10
D-index	.29	.23	.45	.48	.46	.63	.57	.60	.78	.39

Passalong Subtest							
Item	1	2	3	4	5	6	7
D-index	.02	.03	.22	.39	.59	.52	.24

Form Assembly Subtest													
Item	1	2	3	4	5	6	7	8	9	10	11	12	13
D-index	.12	.08	.05	.08	.16	.05	.34	.20	.24	.52	.11	.26	.17

Pattern Matching Subtest															
Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
D-index	.02	.04	.06	.04	.04	.04	.01	.23	.29	.24	.17	.49	.59	.65	.63

In this analysis twenty-three of the sixty QT items were found to have D-values below .2. It cannot be concluded, however, that all these items should be automatically excluded from the test. Some very easy items are probably necessary for subjects to learn the solution strategies required in subsequent items.

Notwithstanding the need for some easy items, the low indices of discrimination from particular sets of items does call their inclusion into question. Inspection of the D-values does suggest that Items 1 to 6 of the Form Assembly Subtest and Items 1 to 7 of the Pattern Matching Subtest contribute very little to the discrimination power of the QT. It is also apparent that the first six items of the Knox Subtest contribute little in a statistical sense, although as noted, there may be good reasons for retaining these easy items in the first subtest.

In this analysis only some fourteen items exceed a D-value of .4 which is considered high by Stanley and Hopkins (1972). Of these fourteen items seven are in the Beads Subtest. The remaining items with high discrimination indices were Item 11 of the Knox Subtest, the last four items of the Pattern Matching Subtest and Items 5 and 6 of the Passalong Subtest.

(c) Item Correlations with Subtest Score and QT Total Score

The items were further analysed by determining the point biserial correlation of each item with the corresponding subtest total score and with the QT Total Score. This analysis was undertaken for the Total Sample and the two ethnic group samples. Tables 6.4.7 to 6.4.16 report the item point biserial correlations for the five subtests. The first table in each pair reports the item correlations with the subtest score and the second the item correlations with the QT Total Score.

It may be noted that the point biserial correlation values reported for the item-QT Total Score correlations for the Total Sample approximate

TABLE 6.4.7 Point Biserial Correlations of Knox Subtest Items with Knox Subtest Score

Sample	N	1	2	3	4	5	6	Items 7	8	9	10	11	12	13	14	15
Total	700	25	35	37	45	43	39	47	53	42	54	59	45	28	32	20
European	397	31	39	32	45	43	37	49	53	42	49	58	44	30	27	14
Maori	303	18	31	43	46	43	41	44	53	41	60	60	46	26	37	24

Decimal points omitted

TABLE 6.4.8 Point Biserial Correlations of Knox Subtest Items with QT Total Score

Sample	N	1	2	3	4	5	6	Items 7	8	9	10	11	12	13	14	15
Total	700	12	25	35	32	27	28	30	33	24	33	36	30	17	21	10
European	397	21	27	35	36	30	29	35	32	26	34	36	28	18	22	10
Maori	303	01	22	33	29	26	28	25	35	21	33	37	34	15	22	13

Decimal points omitted

TABLE 6.4.9 Point Biserial Correlations of Beads Subtest Items
with Beads Subtest Score

Sample	N	Items									
		1	2	3	4	5	6	7	8	9	10
Total	700	35	35	45	49	54	64	63	68	76	51
European	397	34	38	46	46	55	64	64	68	76	51
Maori	303	37	30	42	52	53	65	63	68	75	52

Decimal points omitted

TABLE 6.4.10 Point Biserial Correlations of Beads Subtest Items
with QT Total Score

Sample	N	Items									
		1	2	3	4	5	6	7	8	9	10
Total	700	27	27	34	39	38	45	42	48	56	37
European	397	30	28	32	40	40	48	44	51	57	38
Maori	303	24	25	36	38	34	41	40	42	55	36

Decimal points omitted

TABLE 6.4.11 Point Biserial Correlations of Passalong Subtest
Items with Passalong Subtest Score

Sample	N	Items						
		1	2	3	4	5	6	7
Total	700	05	13	46	63	70	70	51
European	397	04	16	43	61	69	72	52
Maori	303	07	11	48	64	70	69	52

Decimal points omitted

TABLE 6.4.12 Point Biserial Correlations of Passalong Subtest
Items with QT Total Score

Sample	N	Items						
		1	2	3	4	5	6	7
Total	700	04	11	31	35	48	42	25
European	397	05	05	31	34	48	45	30
Maori	303	03	15	31	32	45	38	20

Decimal points omitted

TABLE 6.4.13 Point Biserial Correlations of Form Assembly Subtest Items with Form Assembly Subtest Score

Sample	N	1	2	3	4	5	Items 6	7	8	9	10	11	12	13
Total	700	29	29	20	22	33	20	42	48	44	51	25	47	36
European	397	22	22	18	20	30	14	40	46	46	52	27	47	33
Maori	303	33	37	20	23	35	24	44	50	44	49	20	46	41

Decimal points omitted

TABLE 6.4.14 Point Biserial Correlations of Form Assembly Subtest Items with QT Total Score

Sample	N	1	2	3	4	5	Items 6	7	8	9	10	11	12	13
Total	700	27	17	16	22	26	15	27	23	19	40	18	22	16
European	397	21	12	15	22	25	14	25	20	21	41	18	24	10
Maori	303	31	23	17	22	24	14	28	26	19	38	15	18	25

Decimal points omitted

TABLE 6.4.15 Point Biserial Correlations of Pattern Matching Subtest Items with Pattern Matching Subtest Score

Sample	N	1	2	3	4	5	Items 6	7	8	9	10	11	12	13	14	15
Total	700	15	37	34	34	31	36	16	58	62	57	59	72	72	63	75
European	397	27	23	29	-	35	31	-	51	59	56	53	69	70	62	74
Maori	303	05	43	35	43	29	39	20	61	67	57	64	72	72	64	77

Decimal points omitted

TABLE 6.4.16 Point Biserial Correlations of Pattern Matching Subtest Items with QT Total Score

Sample	N	1	2	3	4	5	Items 6	7	8	9	10	11	12	13	14	15
Total	700	14	25	22	22	26	24	12	42	49	43	44	52	56	51	59
European	397	18	14	14	-	24	18	-	36	45	40	33	50	51	48	53
Maori	303	09	32	26	30	28	29	16	45	53	44	51	52	60	54	63

Decimal points omitted

the D-values reported above. Bridgman (1964) has shown that the values of D are almost perfectly linearly correlated with biserial coefficients.

Inspection of the point biserial correlations for the two ethnic groups between items and subtest score and items and QT Total Score does not show any marked differences in level or pattern. The absence of coefficients for Items 4 and 7 of the Pattern Matching Subtest for European subjects is the result of all European subjects passing these two items.

(d) Factor Analysis of Item Intercorrelations

An exploration of the factorial structure of the QT was also undertaken through the factor analysis of item intercorrelations. Initially attempts were made to calculate the tetrachoric correlation coefficients for all possible pairs of items. Three alternative root-solver routines were introduced into a computer tetrachoric correlation coefficient programme, but collectively failed to solve the coefficient for a considerable number of pairs of items. The root-solver methods attempted were the Newton-Raphson iterative technique (Guilford, 1956b), the quotient-difference algorithm (Rutishauser, 1957) and Bairstow's iterative method (Wilkinson, 1959).

Guilford and Fruchter (1973) have discussed limitations applying to the tetrachoric correlation coefficients noting particularly that unusual distributions of frequencies in the tetrachoric fourfold contingency table will produce a false estimate of correlation and may suggest a non-linear regression between the variables.

To explore the partitioning of the variables, the contingency tables were obtained for all pairs of items. The $(bc)/(ad)$ cell ratio was calculated where possible and converted to tetrachoric r by the method given in Glass and Stanley (1970, p.319). It was found that just on half the fourfold contingency tables had at least one cell with a frequency of 10 entries or less which, from the remarks of Guilford and

Fruchter (1973), would make the calculation of the tetrachoric \underline{r} suspect. Furthermore, tetrachoric \underline{r} could just not be obtained for a considerable number of pairs of items.

As an alternative, the use of the phi coefficient (ψ) was considered. Whereas the tetrachoric \underline{r} assumes that variables have been artificially reduced to two categories (in this case pass or fail an item) the phi coefficient, in theory, assumes that a real gap exists between the two categories. Here this would mean that no continua existed by which it might be said how near a subject came to solving an item correctly. In fact, if the measuring instrument was precise enough, the differentiation of finer degrees of correct performance would theoretically be possible. This might seem to make the application of the phi coefficient inappropriate. However, Guilford and Fruchter (1973) state that -

"The method can be applied to data that are measurable on continuous variables if we make certain allowances for the continuity, with appropriate corrections and interpretations" (p.306).

Phi coefficients were, therefore, calculated for all pairs of items and a QT item intercorrelation matrix obtained. Inspection of this matrix showed that the phi coefficients were on the whole lower than those that were able to be obtained by the tetrachoric method. In addition, negative phi coefficients, which could reasonably be expected, were absent from the matrix.

The phi coefficients had been computed by the procedures given by Nie, Hull, Jenkins, Steinbrenner and Bent (1975) in their Statistical Package for the Social Sciences (SPSS). In describing the procedures they note that by their method of derivation "Phi takes on the value of 0 when no relationship exists, and the value of +1 when the variables are perfectly related" (p.224). Mathematically, phi is related to chi square when the latter is applied to a 2 x 2 contingency table. The procedures used by Nie, Hull, Jenkins, Steinbrenner and Bent (1975) in

calculating phi depend on the prior calculation of chi square which, through the squaring of the numerator value, ensures that phi adopts a positive form. Thus, phi by the computational methods available rendered coefficients for all pairs of items giving an index of the magnitude of association, but no indication of direction. Understandably, both magnitude and direction of correlation are critical to any anticipated factor analysis.

Ferguson (1971) in discussing the phi coefficient notes that it is "a particular case of the product-moment correlation coefficient" (p.230). He goes on to say that -

"If we assign integers, say I and O, to represent the two categories of each variable and calculate the product moment correlation coefficient in the usual way, the result will be identical with ϕ " (p.230).

As a result, the QT item intercorrelations were re-calculated by the product-moment method. All correlations were indeed found to be identical with those derived via the phi coefficient procedures and in addition the direction of the relationship was indicated. This product-moment matrix was therefore employed in the subsequent factor analysis.

The matrix of product-moment correlation coefficients amongst all pairs of QT items is given in Table 6.4.17. This matrix was then subjected to a Principal Factor analysis and a Varimax Rotated Factor analysis.

Determining the number of factors to be extracted, especially from a matrix involving a large number of variables, has caused some disagreement amongst factor theorists. Child (1970) notes that a decision "technique in considerable use at present is Kaiser's criterion..." (p.43). In this case only factors with latent roots (often referred to as the Eigenvalue) greater than one are taken as common factors and considered for interpretation. Cattell (1966) had pointed out, however,

TABLE 4.4.11. Matrix of Pairwise Pearson Product-Moment Correlation Coefficients Amongst QT Items (N = 700)

1															
2	09														
3	00	10													
4	04	16	29												
5	-04	15	15	25											
6	07	18	32	34	25										
7	07	11	12	12	13	11									
8	04	06	12	19	12	21	10								
9	-06	04	10	11	08	06	09	12							
10	09	05	12	13	16	10	16	23	15						
11	01	09	09	19	19	11	20	39	18	22					
12	08	04	03	13	11	07	13	15	20	23	16				
13	06	06	03	00	08	03	05	06	11	09	14	22			
14	10	05	07	02	04	04	05	08	13	17	20	14	27		
15	04	05	03	03	04	02	08	07	01	06	10	24	06	05	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Items 1 - 15 Knox Subtest
 Items 16 - 25 Beads Subtest
 Items 26 - 32 Passalong Subtest
 Items 33 - 45 Form Assembly Subtest
 Items 46 - 60 Pattern Matching Subtest

Decimal points omitted

TABLE 1.1

31	05	03	10	15	09	13	07	09	04	15	13	06	07	08	04	12	06	02	17	12	14	14	15	16	08	05	00	15	31	33
32	02	04	07	05	00	03	03	05	08	06	06	02	00	08	-01	06	03	10	05	05	06	05	08	11	07	03	-01	06	15	20
33	03	06	06	06	03	05	07	04	06	01	29	02	04	-02	-05	08	11	05	04	09	07	08	13	14	05	-01	-02	14	11	13
34	-02	01	05	10	00	02	00	02	03	05	09	06	01	01	-06	06	-04	-02	05	07	00	04	12	09	08	-01	-01	05	02	02
35	09	09	12	00	03	05	08	-04	-05	06	-03	-01	02	-02	01	-04	06	03	11	05	02	05	09	05	03	-01	-01	05	04	08
36	01	10	02	05	04	-03	06	01	05	03	05	04	03	03	02	02	04	10	09	07	13	06	07	14	07	-01	11	03	08	06
37	-05	10	07	07	-01	03	02	02	00	02	04	04	01	01	03	02	09	08	05	09	11	06	12	15	10	-02	10	04	10	09
38	-03	17	14	04	10	03	02	02	02	10	02	07	03	-01	02	01	00	04	00	-05	00	-01	00	02	02	-01	10	04	06	07
39	04	04	01	12	04	-02	00	06	03	03	04	05	02	05	01	02	05	08	08	08	09	08	16	13	10	-04	07	12	11	15
40	-05	-01	12	05	01	05	02	01	04	03	-02	02	05	02	05	06	03	04	09	04	04	06	07	09	08	-03	01	-01	03	05
41	02	00	08	06	06	06	07	07	-05	02	-01	06	-05	-02	09	06	04	02	05	06	00	05	06	05	04	-05	-03	05	02	02
42	-03	05	12	09	06	08	05	09	03	11	05	10	07	05	10	07	11	13	14	10	12	14	12	19	11	05	00	12	13	17
43	-02	-03	01	08	06	04	10	10	-02	06	08	02	-01	-03	04	04	05	-01	02	04	00	07	02	08	01	01	02	07	10	07
44	-03	-02	04	06	02	08	03	08	01	01	06	06	-02	-01	07	02	05	00	14	00	08	00	04	08	02	-02	-06	03	05	03
45	04	-03	06	02	-01	01	05	07	07	01	04	02	06	03	02	04	-02	01	-02	-02	04	-03	05	04	07	-04	06	02	06	08
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Decimal points omitted

TABLE 6.4.17 (Contd.)

46	-03	-03	11	-02	-03	-01	04	03	05	02	02	03	01	02	01	06	-03	03	06	01	06	06	05	07	03	00	-01	09	06	10
47	05	07	05	09	05	05	02	07	04	05	07	05	02	03	01	00	-04	08	04	06	03	01	03	08	02	-01	15	21	06	08
48	-03	06	08	11	03	11	06	05	05	04	04	05	-04	03	01	02	-10	03	04	05	06	00	04	05	06	-01	-01	12	04	08
49	-04	04	03	-03	05	-02	03	10	06	04	04	04	02	02	01	-01	-03	05	09	03	08	00	00	08	04	00	20	15	03	08
50	-04	11	12	07	16	16	04	09	06	05	05	04	02	02	01	06	10	09	07	04	08	01	07	06	04	-01	18	01	11	13
51	-01	07	17	07	12	16	04	09	03	05	05	04	02	02	01	02	-03	06	10	04	08	04	07	06	04	-01	-01	09	05	10
52	05	05	-02	08	-02	-01	05	02	-03	04	-01	02	01	01	01	04	-02	05	02	04	04	04	04	05	02	00	00	06	07	07
53	02	10	19	08	02	17	08	13	09	04	10	07	03	07	-02	03	04	12	13	13	09	10	16	18	10	-02	04	13	16	17
54	05	17	17	21	15	22	13	11	07	06	12	12	00	05	03	09	05	15	21	10	17	17	14	20	12	-02	04	15	17	19
55	05	10	16	12	12	09	09	07	02	05	08	09	05	06	03	05	15	11	18	11	15	11	15	17	12	09	-02	07	13	15
56	04	14	11	11	19	11	07	10	06	02	11	09	05	06	03	08	11	13	12	12	17	10	11	16	11	-01	06	11	16	17
57	06	07	15	06	09	17	04	10	01	10	07	12	04	03	-02	05	07	16	12	16	12	15	18	22	15	04	05	19	27	23
58	00	11	21	11	10	13	10	14	08	10	12	11	06	12	03	10	09	14	20	16	17	15	17	27	14	09	-01	22	17	29
59	04	08	14	10	03	11	15	08	06	07	11	08	08	04	04	12	13	14	19	14	17	18	20	26	18	06	04	08	16	23
60	05	08	13	13	14	18	11	09	07	13	14	11	03	07	00	15	12	13	19	17	15	18	18	26	19	08	08	15	22	27
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Decimal points omitted

that this criterion is most reliable when the set of variables number 20 to 50. Child, following Cattell's argument, has concluded that -

"Where the number of variables is less than 20, there is a tendency, not too serious, for this method to extract a conservative number of factors. When more than 50 variables are involved, too many factors are taken out" (pp. 43-44).

It is Cattell's view that the latter event occurs because the unique variance present in all factors is proportionately of a sufficient order in the later factors extracted to swamp the common variance. As a result the optimum number of factors to be extracted must be determined before the intrusion of non-common variance becomes too serious.

One method for determining the optimum number of factors is Cattell's Scree Test which is based on a plot of latent roots against the order of factor extraction. The resulting curve is used in judging the cut-off point. The point at which the function of the curve changes from curvilinear to linear is taken as the determination point and the intersection of the perpendicular from this point to the factor number axis read as the number of factors to be extracted.

The initial attempts to apply factor analytic techniques to the QT item correlations encountered difficulties in that convergence amongst the variables could not be achieved within the iteration limits of the computer programmes available. Consequently, Cattell's Scree Test was applied in order to limit the extraction of factors. The Scree Test suggested the extraction of six factors.

The extraction of six factors could also have been argued for on theoretical grounds. The remarks by McElwain and Kearney (1970) in favour of the QT as a unified measure of general ability would be strengthened by evidence of a general first factor accounting for a large proportion of the variance. It might then be expected that some evidence might emerge on the homogeneity of items within subtests and

the unique aspects of intellectual ability tapped by each of the five subtests.

Table 6.4.18 presents the Principal Factor Matrix showing the six factors extracted. Only the first three factors account individually for more than 10% of the variance. The remaining factors each account for less than 10% of the variance and are not considered to be meaningful. The fact that the majority of items loaded in the first factor in this analysis would suggest that it is of general nature. The highest loadings in this factor were on the Pattern Matching subtest. Of the QT subtests, Pattern Matching has, on the whole, correlated best with other measures of general ability (see relevant tables in section on Criterion Related Validity). Of the five subtests the Form Assembly subtest contributes least to this first factor.

The second factor loaded in a bipolar manner on the items of the Beads and Pattern Matching subtests. In common with all other QT items perceptual recognition, spatial organisation and the physical manipulation of materials are involved. Nonetheless, it would be difficult to see these item types as occupying end points on a continuum on those grounds. However, a distinction between the item types, although purely speculative, that can be offered concerns the two versus three dimensional nature of these subtest items.

It will be recalled that considerable effort went into the modification of the Pattern Matching subtest to present the task in a two-dimensional form. It could be argued that while both tasks may involve similar underlying cognitive skills, the Beads subtest makes use of the critical third dimension through which subjects clearly distinguish the sequencing of the cubes, spheres and cylinders that go to make up the bead sequence.

TABLE 6.4.18 (Contd.)

Variables	Factors						Commynality (h^2)
	1	2	3	4	5	6	
Form Assembly Subtest							
Item 1	-.277						.107
2							.033
3							.062
4	-.237						.136
5	-.265						.133
6						-.251	.114
7							.064
8						-.387	.221
9							.050
10	-.369						.178
11							.032
12							.062
13							.018
Pattern Matching Subtest							
1							.052
2	-.339	-.407		.258			.366
3	-.269	-.224					.146
4	-.313	-.421		.385			.472
5	-.304	-.212					.155
6	-.334	-.384		.299			.356
7							.054
8	-.482						.290
9	-.554						.366
10	-.490	-.203					.301
11	-.517	-.306					.389
12	-.585			-.264			.463
13	-.601			-.240			.435
14	-.504						.307
15	-.627			-.271			.484
Total % of Variance	50.2	15.7	11.9	8.9	7.3	5.9	100.0
Eigenvalue	6.09	1.90	1.45	1.08	0.89	0.72	

The third factor is almost entirely specific to the items of the Knox Subtest. The items involve the imitative performance of a block tapping sequence. The factor is interpreted as primarily reflecting the short-term memory component inherent in making an immediate and imitative block tapping response on the part of the subject.

While it is reasonable to suggest the above interpretations of the three main factors arising from the Principal Factor analysis, it must still be remembered that the amount of variance accounted for by the first factor could be an artifact of the analysis technique. A clarification of the factor structure of the items was, therefore, sought by a Varimax rotation of the six Principal Factors.

The Varimax Rotated Factor Matrix is presented in Table 6.4.19. The first four rotated factors were found to account for just over three-quarters of the variance. Only these first four factors have been reported and interpreted in this instance.

The first rotated factor loaded on all but one item (the first) of the Beads Subtest. It also loaded on the fifth item of the Passalong Subtest and the fourteenth item of the Pattern Matching Subtest. This factor accounts for half the total variance. The interpretation of this factor is most clearly and simply as a 'Beads Subtest factor' indicating the high degree of homogeneity amongst items of this type. This is supported by the satisfactory internal consistency estimates of reliability obtained for this subtest which indicated a good measure of item homogeneity.

The second rotated factor loaded on all but three items of the Pattern Matching Subtest. It also loaded on two items of the Form Assembly Subtest. The loadings on the Pattern Matching Subtest were in the main higher on the easier items involving 4-tile constructions

TABLE 6.4.19 Varimax Rotated Factor Matrix for QT Items Based on Product Moment Correlation Coefficients Among Items: The First Four Factors
 Rotated Principal Factor Matrix After Applying Cattell's Scree Test (N=700)
 (Loadings less than .20 not shown)

Variables	Factors				Communality (h^2)
	1	2	3	4	
Knox Subtest					
Item 1					.027
2				.251	.095
3				.405	.241
4				.542	.330
5				.368	.188
6				.610	.391
7					.097
8				.251	.203
9					.114
10					.203
11				.209	.287
12					.222
13					.115
14					.135
15					.080
Beads Subtest					
Item 1					.073
2	.246				.077
3	.269				.110
4	.351				.180
5	.441				.221
6	.568				.352
7	.554				.326
8	.645				.439
9	.751				.606
10	.438				.212
Passalong Subtest					
Item 1					.009
2					.049
3			.266		.137
4			.400		.174
5	.201		.429		.262
6			.436		.245
7			.258		.088

TABLE 6.4.19 Varimax Rotated Factor Matrix for QT Items Based on Product Moment Correlation Coefficients Among Items: The First Four Factors From Rotated Principal Factor Matrix After Applying Cattell's Scree Test (N=700)
(Loadings less than .20 not shown)

Variables	Factors				Commynality (h ²)
	1	2	3	4	
Knox Subtest					
Item 1					.027
2				.251	.095
3				.405	.241
4				.542	.330
5				.368	.188
6				.610	.391
7					.097
8				.251	.203
9					.114
10					.203
11				.209	.287
12					.222
13					.115
14					.135
15					.080
Beads Subtest					
Item 1					.073
2	.246				.077
3	.269				.110
4	.351				.180
5	.441				.221
6	.568				.352
7	.554				.326
8	.645				.439
9	.751				.606
10	.438				.212
Passalong Subtest					
Item 1					.009
2					.049
3			.266		.137
4			.400		.174
5	.201		.429		.262
6			.436		.245
7			.258		.088

TABLE 6.4.19 (Contd.)

Variables	1	2	3	4	Commynality (h^2)
Form Assembly Subtest					
Item 1			.254		.107
2					.033
3					.062
4		.259			.136
5		.202			.113
6					.114
7					.064
8					.221
9					.050
10			.250		.178
11					.032
12					.062
13					.018
Pattern Matching Subtest					
Item 1					.052
2		.587			.366
3		.297			.146
4		.675			.472
5		.315			.155
6		.569			.356
7					.054
8		.301	.392		.290
9		.339	.375	.276	.366
10		.389	.254		.301
11		.522	.229		.389
12		.262	.561		.463
13		.216	.527		.435
14	.217		.393		.307
15		.211	.574		.484
Total % of Variance	22.6	20.4	20.8	12.7	76.5
Eigenvalue	2.73	2.47	2.52	1.54	
Loadings less than .20 not shown					

(Items 2, 3, 4, 5, 6, 8, 9 and 10) and on the first item involving a 9-tile construction (Item 11) than on the remaining four items.

While the internal consistency estimates for the Pattern Matching Subtest are satisfactory, the item difficulty analysis showed that over 90% of the Total Sample passed items 1 to 11. In addition, the item discrimination indices for these eleven items were, on the whole, low. This factor is seen as being one reflecting the high homogeneity amongst the set of eleven extremely easy items of the Pattern Matching Subtest.

The third rotated factor loaded primarily on the items of two subtests. First, Items 3 to 7 of the Passalong Subtest and second, Items 8 to 15 of the Pattern Matching Subtest. Two loadings on items of the Form Assembly Subtest also occurred in this factor. This factor provides some evidence of the homogeneity of the items of the Passalong Subtest, but is complicated by the loadings on the Pattern Matching Subtest.

A feature of the loadings on the latter subtest is that the higher loadings pertain to the last four items (Items, 12, 13, 14 and 15). These items, while not found to be particularly difficult, still had satisfactory discrimination indices. The items of the Passalong Subtest implicated in this factor were also those with high discrimination indices.

While outwardly the items in this factor do not appear to belong together, the analysis suggests that the common contribution of these different but discriminating items is being tapped by the factor.

The fourth rotated factor loaded on seven of the items of the Knox Subtest and one item of the Pattern Matching Subtest. This is read as evidence of a degree of homogeneity amongst items of the Knox Subtest.

To summarize, the rotated factor solution presents some evidence

of the clustering of items within the majority of QT subtests. Items of the Beads Subtest clearly belong together and contribute 23% of the total variance. Similarly, items of the Knox Subtest evidence homogeneity and as the fourth factor still accounts for 12% of the variance.

The second and third factors might best be understood in relation to one another. The second factor, while clearly evidencing the homogeneity of the Pattern Matching Subtest items, loads primarily on items found to be extremely easy. On the other hand, the third factor loaded on the difficult items of this subtest and on the majority of items of the short and difficult Passalong Subtest.

The final feature of the analysis that should not pass unnoticed is the absence of any high and/or consistent loadings by the items of the Form Assembly Subtest. Item discrimination analyses have shown that the majority of items in this subtest had poor discriminative power and the item difficulty order data were not altogether encouraging. Furthermore, the low internal consistency estimates also suggested that the items were not particularly homogeneous. Taken together, these data point to the Form Assembly Subtest having rather marginal psychometric qualities relative to the other QT subtests.

PART 5 PRELIMINARY DECILE RANK NORMS

In preparing New Zealand QT norms, a number of factors had to be taken into consideration. These were the importance of the ethnic group difference in performance, the sample sizes, the representativeness of the samples and the drop in the level of performance at the oldest age group level. These factors along with evidence of reliability and considerations related to the likely use of the QT have a bearing upon the norm method adopted.

The fact that a statistically significant ethnic group difference was found in QT performance would suggest the need for separate European and Maori norms. Investigation of the direct contribution of ethnicity to QT performance after the semi-partialling out of other variables showed, however, that ethnic group membership accounted for only two percent of the variance. Thus, knowledge of ethnic group membership contributes very little to an understanding of QT performance. Age of subject accounting for eighteen percent of the variance is of far greater importance. The small contribution of ethnicity as an 'explanatory' variable, the small estimated constant difference of 2.15 QT raw score points and the considerable overlap in raw score distributions suggest that a single set of norms would be appropriate.

A further consideration which favoured a single set of norms was the question of sample size. Separate ethnic group norms for each age level would have been based in some cases on samples of between 30 to 40 subjects. Samples of this size were considered to be somewhat inadequate as a basis for any proposed norms. Combined ethnic group samples at each age level would provide enlarged sets of data for the derivation of norms.

The question of the representativeness of the sample was considered in Chapter 4. It was concluded that the samples could not

be considered fully representative on a national basis. It was clear that on a national basis, Maori subjects are over-represented in the Total Sample and thus over-represented in each age group sample. Conversely, European subjects are under-represented in the Total Sample and age group samples. In addition, other ethnic and cultural groups that might legitimately claim representation (e.g. Pacific Island groups) have not been sampled.

Male and female subjects were equally well represented and the proportions of subjects from urban and rural settings were, however, in accord with national data. Within the ethnic group sample the distribution of socio-economic status levels conformed with what is known to be the case nationally. Nonetheless, the pooling of the ethnic group samples produced a socio-economic status level distribution with higher proportions in the lower status levels than is indicated by national data. This was the result of the higher proportion of Maori subjects in the lower status levels in conjunction with their over-representation in the Total Sample. On the question of geographical distribution, the sample was also considered to be unrepresentative.

The mean level of performance on the QT for the 13:6 - 14:5 age group was found to be a little below that of the previous and younger age group. This was the case for both European and Maori samples. McElwain and Kearney (1970) expected QT performance to increase during the period of general biological maturation and the Australian data show this to be the case. Their data for subjects aged 12 to 15 years show no evidence of a decline in performance (see McElwain and Kearney, 1970, pp. 62-66). Cronbach (1970) reviewing the evidence on ability changes with age confirms the expectation of relatively consistent increments during childhood and adolescence reflecting, he thinks, both biological maturation and educational effort and opportunity.

The decline in the mean level of performance on the QT by the

13:6 - 14:5 age group relative to the next youngest age group requires explanation. It is thought that in drawing the 13:6 - 14:5 age group sample from the primary school level the sample was biased in favour of subjects experiencing educational difficulties that may involve intellectual capacity. All subjects at this age group level were drawn from Form II classes. The median age for Form II classes as of July 1972 was approximately twelve years and nine months (refer Department of Education, 1973). With the majority of children in the ages between 13:6 - 14:5 at secondary school sampling at that level would be required to better assess the QT performance of this age group.

Given the suspect nature of the QT data for the 13:6 - 14:5 age group, a decision was made against deriving norms from the data for this particular age level.

In light of the above it was decided that a single set of norms should be produced for the combined ethnic group samples for the five twelve month age groups between 8 years 6 months and 13 years 5 months. It was also apparent that the present samples could not be considered fully representative. Consequently, the preliminary nature of the norms presented cannot be over-emphasised.

Preliminary decile rank norms were derived for the five age groups and are presented in Table 6.5.1. Decile rank norms were chosen for three reasons. First, the sample sizes of 73 and 90 for the two youngest age groups were not considered satisfactory for norms requiring finer point distinctions (e.g. percentile ranks). Second, decile rank norms can be readily understood by persons with little psychometric sophistication, and third, norms of this type while discouraging unrealistic beliefs on the precision of intellectual assessment would hopefully render information of sufficient discriminative capacity to assist decision-making.

TABLE 6.5.1 QT Decile Rank Norms for Five New Zealand Age Groups

QT Decile Rank Norms: New Zealand Subjects Aged 8:6 - 9:5

Raw Score	Decile Ranks
42+	10
40 - 41	9
37 - 39	8
35 - 36	7
33 - 34	6
32	5
30 - 31	4
26 - 29	3
21 - 25	2
- 20	1

QT Decile Rank Norms: New Zealand Subjects Aged 9:6 - 10:5

Raw Score	Decile Ranks
44+	10
42 - 43	9
39 - 41	8
37 - 38	7
36	6
34 - 35	5
33	4
31 - 32	3
29 - 30	2
- 28	1

TABLE 6.5.1 (Contd.)

QT Decile Rank Norms: New Zealand Subjects Aged 10:6 - 11:5

Raw Score	Decile Ranks
46+	10
44 - 45	9
43	8
41 - 42	7
40	6
38 - 39	5
37	4
35 - 36	3
33 - 34	2
- 32	1

QT Decile Rank Norms: New Zealand Subjects Aged 11:6 - 12:5

Raw Score	Decile Ranks
47+	10
45 - 46	9
43 - 44	8
41 - 42	7
39 - 40	6
38	5
37	4
35 - 36	3
32 - 34	2
- 31	1

TABLE 6.5.1 (Contd.)

QT Decile Rank Norms: New Zealand Subjects Aged 12:6 - 13:5

Raw Score	Decile Ranks
50+	10
48 - 49	9
47	8
46	7
44 - 45	6
43	5
42	4
40 - 41	3
36 - 39	2
- 35	1

CHAPTER 7

DISCUSSION OF RESULTS AND CONCLUSIONS

The discussion of results and the drawing of conclusions from this research is presented in three parts.

PART 1 summarises and discusses the major results from the studies of the QT with New Zealand European and Maori subjects.

PART 2 discusses these results in relation to previous QT research undertaken in Australia and the South Pacific.

PART 3 focuses upon a number of more general issues. The research reported is related to the study of intelligence and to cross-cultural considerations in this domain. Attention is given to the relationship of this research to similar New Zealand studies. Some general implications arising from this study are also noted.

PART 1 SUMMARY AND DISCUSSION OF MAJOR FINDINGS

Considerable statistical evidence has been presented in the preceding chapter on the QT performances of the ethnic, gender and age group samples, on QT reliability and validity, on QT item characteristics and on preliminary QT norms. While discussion has accompanied the presentation of the above data it is appropriate to summarize and comment upon the major findings of this research in order that the pattern and implications of the findings can be more clearly appreciated in relation to the research proposals.

The analysis of the QT Total Score performances showed that while gender of subject had no bearing on performance both age and ethnic group did. The increase in QT performance with increasing age for the samples tested would be expected on the grounds that any hypothetical construct concerning intelligence underlying QT performance would be related to maturational and experiential factors. A drop in the level of performance by the oldest age group sample was observed and it was concluded that this had resulted from sampling limitations.

The ethnic group difference was studied extensively. Analyses showed that the estimated average QT Total Score difference favouring European subjects over Maori subjects was in the order of 2 raw score points. However, the direct and independent variance accounted for by ethnic group was found to be very small. It was concluded that ethnic group membership was of little practical significance in explaining QT performance.

The analyses of the five QT subtests showed that in general performance increased with age. Ethnic group differences on the subtests were limited to the Form Assembly and Pattern Matching subtests. These findings might suggest an ethnically based difference in cognitive organisation, a question which will be commented upon

further when considering the factor analytic data. It was also found that gender of subject had a bearing upon performance on the Pattern Matching Subtest with males out-performing females. Again, the possible cognitive organisation implications will be taken up later.

The score distributions for the separate and combined ethnic group samples were found to display a similar degree of negative skew. The subtest score distributions for the same samples also displayed considerable similarity. The score distribution for the Pattern Matching Subtest clearly indicated the ease of the great majority of items for most subjects. Modification of this subtest would seem desirable.

The intercorrelation analyses between QT subtests and QT Total Scores for the separate and combined ethnic groups at the six age group levels showed that, in general, the correlations were of a moderate positive order and statistically significant. Subtest intercorrelations were of a lower order. The evidence would support the classical test theory requirements of low subtest intercorrelations and high subtest-total score correlations for an instrument of this nature.

In the case of two Maori age group samples the relationship between Knox Subtest scores and the remaining subtest scores was found to be almost non-existent. This may again suggest the operation of an ethnically based cognitive difference.

The reliability studies found that the estimates of QT reliability by the K-R 20 internal consistency method were satisfactory for both separate and combined ethnic group samples at each age group level. As expected, the internal consistency estimates were somewhat lower for the subtests. In general, the estimates for the Form Assembly Subtest were lowest, whereas those for the Pattern Matching Subtest were in a good many instances similar to full QT estimates.

The test-retest reliability studies did not provide particularly strong evidence on the stability of QT performances over the relatively short six-month interval. The estimates obtained from the pooled class group data for the separate and combined samples were all of a similar order. While these test-retest coefficients achieved statistical significance, they were below the level of relationship considered desirable for individual decision making.

The standard error of measurement estimates for the QT and its subtests were found to be, on the whole, similar for the two ethnic groups. The estimates for the QT tended to decrease with age. With most standard error of measurement estimates being in the order of 4.5 points, rather wide confidence bands are suggested.

The question of the validity of the QT was approached by considering the evidence on content validity and through studies into aspects of criterion-related and construct validity.

Evidence on the development of the QT, the nature of its items and the steps taken to facilitate its application to culturally different subjects was examined. It was argued that its content was in keeping with the test authors' intentions to construct a measure of general intelligence subject to minimal cultural influences. On these grounds a measure of support could be drawn for its suitability for use with both European and Maori subjects.

The criterion-related validity studies showed that the correlations between QT performance and the various criteria measures across class and ethnic groups were not distinctively different in any systematic way. The ethnic group samples were, however, small and this does mean that the data should be treated with caution.

The criterion-related validity data from the analysis of each class group disregarding ethnic group membership showed reasonably consistent and statistically significant relationships with the criteria

of school mathematics achievement and the teacher ranking of 'General Ability'. Correlations with criteria emphasising English language skills were, in general, lower and non-significant. In view of the content of the QT this is not surprising.

Performance on the QT was found to be related to performance on a group non-verbal reasoning test in a positive and statistically significant manner for the majority of class group samples, whereas the correlations with a group verbal intelligence test achieved statistical significance less frequently. These findings are in accord with propositions concerning the nature of the construct measured by the QT which will be commented upon shortly. What was apparent, however, from the intercorrelational data amongst these tests, the QT and the other criteria was the fact that the correlations between the group verbal intelligence test and the criteria were, on the whole, of a higher order than the correlations with the QT or the non-verbal reasoning test.

It would appear then that while some evidence could be deduced from the present studies in support of the criterion-related validity of the QT, the measure does not compare favourably in this regard with other simpler and more economic measures of ability.

The question of construct validity was approached in three ways. Firstly, it was briefly argued that test content should be in accord with the hypothesized construct tapped by the measure. It was suggested that a case could be made favouring the content of the QT in relation to the hypothesised construct of general intelligence. It was pointed out, however, that the judgement of the test user on this matter remains an important consideration.

The second approach concerned the study of the relationship between QT performance and age. It was found that QT performance increased with age in both the European and Maori samples and that the

relationship was almost identical for the two ethnic groups. This data supports a general intelligence construct that is influenced by maturational and experiential factors.

The third approach to the question of construct validity of the QT involved the factor analysis of subsets of the research data. Factor analyses were conducted on the QT subtest performances of the combined ethnic group samples at each of the six age group levels. The results suggest an important first factor which has been identified as a general intelligence or 'g' type factor. The subtests which appeared to be principally involved in this factor were Pattern Matching, Passalong and Form Assembly. The second factor in these analyses was thought to be of a spatial-perceptual nature tapping principally the perceptual discrimination aspects of the Pattern Matching, Beads and Form Assembly Subtests.

It must be pointed out that these interpretations are tentative rather than conclusive and are based on the more common features of the age group analyses. The interpretations are also limited by the absence of additional data that may have helped define the factors more clearly.

Separate ethnic group factor analyses of the QT subtests were undertaken. For both the European and Maori samples an important first factor was obtained. It was interpreted as a general factor, but it was noted that the subtest loadings were different for the ethnic groups. For both ethnic groups the Passalong and Pattern Matching Subtests are strongly implicated in the first factor along with age. In the case of European subjects, however, both the Beads and Form Assembly Subtests loaded clearly into the first factor but this was not the case for the Maori sample.

The second factor differences are also of interest. With the European sample the principal loading was on the Knox Subtest. With the Maori sample the Beads Subtest and age loaded on this factor.

While these second factors were of minor statistical significance, the findings might again point to an ethnic group difference in aspects of cognitive organisation even though the levels of QT performance are very similar.

Pooling data across class groups enabled a factor analysis of the QT subtests and a number of additional variables to be undertaken. The rotated factor solution evidenced a first factor involving the four group measures of ability and educational attainment. This was identified as a verbal-education factor reflecting English language skills in particular.

The second factor involved the variables of age, three QT subtests and three group measures. It was argued that the analysis pointed to this being a reasoning-ability factor somewhat distinct from the English language skill and educational attainment aspects of the first factor.

The third factor in the analysis appeared to involve variables with spatial/perceptual content and was interpreted as a spatial factor.

Overall, the results provided support for the view that the QT is primarily a measure of a general intelligence construct which is relatively independent of content of a verbal-educational nature. The items of the QT were studied by four methods.

Item difficulty was ascertained for the combined and separate ethnic group samples for each subtest. The proportions passing each item in the ethnic group samples were very similar. The data suggests that the item difficulties of the Knox, Beads and Passalong Subtests are acceptable. In the case of the Form Assembly Subtest the expected gradient of increasing item difficulty was found up until the last two items where the proportion of subjects passing each

item increased. This does not follow the desired pattern of steadily increasing item difficulty although it is realised that the mandatory administration of the last item is likely to have artificially increased the proportion solving it correctly. The item difficulty data for the Pattern Matching Subtest leads to the conclusion that some modification is needed. The first eleven items were extremely easy, suggesting that some should be excluded. There also appears to be a need to include some new and more difficult items.

Item discrimination indices were obtained for each QT item from an analysis based on the Total Sample. Approximately one third of the items had low discrimination indices. It was apparent that some items could be discarded or replaced in the Knox, Form Assembly and Pattern Matching Subtests. The most discriminating items were found to come from the Beads Subtest. The latter items in the Pattern Matching and Passalong Subtests also contributed to the discriminative power of the QT.

Point biserial correlations between each item and its subtest score and also between each item and the QT Total Score were reported for separate and combined ethnic groups. The item-QT Total Score point-biserial correlations approximate the discrimination indices and it was noted that evidence has shown these values to be linearly correlated. The pattern and level of correlations for each ethnic group was very similar and supported the conclusions drawn from the item discrimination index data.

The item factor analyses presented some difficulties. It was necessary to employ the product-moment correlation method to establish an item intercorrelation matrix and Cattell's Scree Test to limit the extraction of factors. The rotated factor analysis provided reasonable evidence of the clustering of items within subtests for all but the Form Assembly Subtest which did not feature clearly or consistently in the analysis.

The question of developing norms for the QT on the basis of the data to hand was considered. Both the small ethnic group difference and the question of sample sizes favoured the production of a single set of norms. Decile rank norms were prepared for five age groups. The norms are clearly preliminary and tentative. In addition, the evidence on QT reliability suggests caution in the use of both raw score or norm data in situations involving individual decision making.

By and large the original research proposals have been met by these studies. The question of the comparative QT performances of European and Maori subjects between the ages of approximately eight to fifteen years has been given considerable attention. Sufficient data was obtained to allow for the investigation of various psychometric properties of the QT with both European and Maori subjects and a set of preliminary decile rank norms were developed for five age group samples of New Zealand children.

The New Zealand studies have also provided much data that can be compared with the results of studies with the QT in other cultural and social contexts.

PART 2 THE RESULTS IN RELATION TO OTHER QT STUDIES

A number of instructive comparisons can be made between the results of these New Zealand QT studies and the results from the earlier and similar Australian studies (Kearney, 1966b; McElwain and Kearney, 1970).

Inspection of the mean QT performance data from Australian European and Australian Aboriginal samples reported in McElwain and Kearney (1970) indicate that, for subjects of the same age, performances of New Zealand European and Maori subjects are on a par with those

of Australian European subjects. The Australian Aboriginal data is reported for groups considered to be in medium or low contact with the dominant European Australian society. Mean QT score differences in performance favour Australian European subjects over medium contact Aboriginal subjects, who in turn out-perform low contact Aboriginal subjects. These results are seen to be "dependent to a considerable degree upon contact or some variable related to contact" (McElwain and Kearney, 1973, p.47).

The contact and acculturation situations in Australia and New Zealand between the European and indigenous populations are somewhat different (see Geddes, 1961; Sinclair, 1971). The New Zealand data might suggest a considerably greater degree of similarity between European and Maori in the sorts of experiences and learning contributing to QT performance than between Australian European and Aboriginal subjects.

Acculturation models have, however, been criticized in terms of the explicit and implicit value and life-style preferences reflected in the demarcation criteria (Fitzgerald, 1972) and because the reality of contemporary social development in culturally heterogeneous societies may be more of a cultural pluralistic nature than an assimilative movement towards a particular single culture system (Havighurst, 1974). Nonetheless, the differences in the social positions of the indigenous peoples of Australia and New Zealand are reflected, albeit crudely, in the pattern of QT results. This data would seem to support the contention that QT performance is influenced by 'contact related variables'.

Comparisons between the levels of performance on the QT subtests by the Australian European and New Zealand European and Maori subjects of comparable ages show the mean scores to be similar. The mean subtest performances of the medium and low contact Aboriginal samples are somewhat lower than those obtained with the above samples. The QT

subtest and Total Score intercorrelation analyses have been undertaken on a different basis in the Australian and New Zealand QT research. In the New Zealand studies this data is presented for the separate and combined ethnic groups at each age level. In the Australian studies it is presented in terms of sample groups from particular geographic locations disregarding age range. Inspection of the QT subtest and Total Score correlations from the two Australian European samples (i.e. McElwain and Kearney's, 1970, Milton and Taringa samples) does, however, show that the correlations are generally of the same moderate-positive order as those pertaining to the New Zealand age group samples. Subtest intercorrelations in both studies tended to be of a low positive order. The subtest-Total Score correlation data and the subtest intercorrelation data from both the Australian and New Zealand studies lends support to the contention that each of the QT subtests do make an important and partially independent contribution to overall QT performance. McElwain and Kearney (1970) have not presented data relating to the distribution of scores of QT Total Scores or subtest scores for the various Australian samples, therefore no comparisons of score distributions can be made.

The reliability studies undertaken in the New Zealand context point to the QT being somewhat less reliable than anticipated when compared with Australian findings. While the internal consistency estimates of reliability were essentially comparable, the New Zealand test-retest estimates were below those obtained in the Australian study despite the shorter time interval involved in the New Zealand studies. The standard error of measurement estimates were also higher for the New Zealand samples which would lead to the use of wider confidence bands.

The lower reliability of the QT in the New Zealand context must, for the present, be taken into account in any anticipated application of

the instrument. Additional research into QT reliability with New Zealand samples might also be in order.

The New Zealand validity studies are to a large extent in agreement with the findings and interpretations of McElwain and Kearney (1970). Content validity considerations have already been discussed in Chapter 6. It was concluded that the item types employed in the QT reflected the test authors' intentions of constructing a 'culture-reduced' measure of general ability and that there were therefore reasonable grounds for assuming content validity in both the Australian and New Zealand contexts.

The criterion-related validity studies in Australia and New Zealand evidence similar findings. In both situations moderate-positive correlations were obtained with teacher assessments of ability. Correlations with scholastic criteria reflecting English language skills were generally low. There is, however, a suggestion in the data from both situations that the QT correlates a little better with scholastic criteria of a mathematics nature.

From both the Australian and New Zealand studies it is apparent that the QT correlates less well with educational attainment criteria than do group verbal ability measures such as the Otis Test or group non-verbal ability measures such as the Raven's Progressive Matrices Test or Pacific Reasoning Series Test. Both Smith (1966) and McElwain and Kearney (1970) have discussed the implications of such findings. They have argued that the QT has a place in culturally extreme or diverse situations where instruments requiring a degree of schooling to facilitate administration and containing content of an educational attainment nature could not be applied. They suggested that demonstrating the criterion-related validity of the QT in such situations is likely to require criteria more directly reflecting the

acquisition of specified new skills.

Consideration of available data bearing upon QT construct validity has led to the drawing of similar conclusions in both Australia and New Zealand. The evidence favours a general ability construct akin to Spearman's 'g', but ambiguities in the factor analytic evidence cannot be entirely overlooked.

No changes were made to the content of the QT in its adoption for the New Zealand studies. Consideration of its content supported the conclusion that the item types were known to be good measures of a general ability construct.

Studies of the relationships between QT performance and age in both Australia and New Zealand evidence increasing performance with age which is in accord with the proposition that general ability is influenced by maturational processes during the period of general biological development.

The Australian QT subtest factor analyses were conducted upon the separate sample groups of low and medium contact Aboriginal subjects and European subjects from differing geographical locations. The first rotated factor was general and interpreted as a 'g' type factor. The second factor loaded primarily on the Form Assembly and Pattern Matching subtests and was interpreted as a spatial or 'k' type factor.

The New Zealand age group factor analyses of the QT subtests all evidenced an important first factor and a minor second factor. However, the pattern of loadings by subtests showed some inconsistency. General ability and spatial ability interpretations were suggested for these two factors but reservations were also expressed.

The New Zealand ethnic group factor analyses suggested differences in the organisation of the intellectual abilities tapped by the QT even though age for age ethnic group performances differed little.

The Australian data did not suggest differences in ethnic group factorial structure. McElwain and Kearney (1970) state that "The factorial structure of the test appears to be fairly similar over the groups" (p.143).

McElwain and Kearney (1970) did recommend further exploration of the factorial structure of the QT through studies employing other psychological tests. A limited New Zealand study of this nature employing group measures of ability and educational attainment evidenced a first factor of a verbal-educational nature without substantial QT subtest loadings. The QT subtests were strongly implicated in the second factor along with age and higher loadings on the group ability measures. This factor was interpreted as a reasoning factor. The third small factor appeared to be of a spatial nature. While comparable Australian data has not come to hand as yet, these New Zealand results are not unlike those obtained by Richardson (1970) with Papua New Guinean armed service recruits. He found a first factor loading on measures of a verbal-educational nature, a second factor loading on three QT subtests and two group ability tests, a third factor with secondary loadings on some QT subtests and an obscure fourth factor. His first factor he labelled an "acculturation factor". Be that as it may, its content was clearly of a verbal-educational skills nature. The second factor was seen as being intelligence independent of culture and the third factor as speed of performance.

Together these results lend support to the proposition that the QT does measure a reasoning or general intelligence type construct that is largely independent of verbal and educational type skills.

Comparison of the data on QT item characteristics points to some important differences between the Australian and New Zealand findings. Examination of McElwain and Kearney's (1970, pp.88-89) data on the

proportion of subjects passing each item suggests that nearly all items were found to be more difficult by the subjects involved in the Australian studies. The differences in the difficulty of items in the Pattern Matching Subtest are considerable. It will be recalled that with the New Zealand subjects some ninety percent passed Items 1 to 11 inclusive. The proportion of subjects passing these items drops steadily to some fifty percent by Item 11 in the Australian case.

While the age ranges studied in Australia and New Zealand are similar, it would appear that the pooling of QT data from more heterogeneous samples in the Australian case may have contributed to this difference. Substantial QT mean score differences were found between low contact Aboriginal, medium contact Aboriginal and European samples. However, data from all Australian samples was combined when determining item difficulty. It may have been more appropriate for McElwain and Kearney to have also presented item difficulty data for their samples in terms of contact-level and ethnic group. The higher mean performances of their European samples does suggest that the items were easier for these subjects and a separate item difficulty analysis might have led to a restructuring of the subtest during the development of the instrument.

Modification of the Pattern Matching Subtest is certainly suggested by the New Zealand item difficulty data, including that from the separate European and Maori samples. Perhaps it should not be forgotten that Ord's (1968c) Pacific Design Construction Test, which was a forerunner to this subtest, contained additional more difficult items as well as employing speed of performance credits in the scoring system. This test would need to be considered in any proposed restructuring of the QT Pattern Matching Subtest.

Discrimination indices were not obtained from the Australian data. However, item correlations with subtest score and QT Total Score have been reported. The Australian data is presented for all the sub-

samples studied and for combined samples. Detailed comparison is not easy and probably not particularly meaningful. Nonetheless, it can be observed that in comparing item correlations with QT Total Score for the pooled Australian and New Zealand samples the correlations are generally higher in the Australian case. This would be in accord with the observation that the Australian sample, on the whole, found the QT items more difficult. Given Bridgman's (1974) evidence on the relationship between the discrimination index D and biserial item-score correlations, it would also be the case that discriminative power of the QT items could be judged superior with the Australian samples.

Obtaining item intercorrelations proved to be difficult with the New Zealand data primarily, it is thought, because of the relatively high proportion of subjects passing each item. While tetrachoric correlations were employed with the Australian data the product-moment method was applied in present study. Nevertheless, there is some common ground in the nature of the results from the factor analysis of the item intercorrelation matrices. The unrotated analysis for the Australian and New Zealand data both show a general first factor which loads on the majority of QT items. Both analyses also suggest factors more specific to the Knox and Beads Subtests.

A rotated analysis was not reported for the Australian data nor was any comment made on the necessity to restrict the extraction of factors. The similarities in the unrotated factor loadings are encouraging, but the differences in the initial data and analysis procedures should not be overlooked.

QT research elsewhere in the South Pacific has been on a much more limited scale. Nonetheless, the results lend weight to some conclusions arrived at above and also raise a number of questions.

McElwain and Kearney (1973) have suggested that variables related to contact with European society have a bearing upon the QT performances of Australian Aboriginal subjects. Bennett and Chandra

(1974) in studies in Fiji have reported QT performance differences favouring urban Fijian subjects over rural Fijian subjects. Their data suggests the operation of experiential factors related to the urban setting. Chandra (1975) continuing the QT studies in Fiji has incorporated urban and rural samples of Indian subjects finding the same pattern of results.

QT data from Cook Islands Maori subjects were considered by St. George and St. George (1974) in relation to that from the current New Zealand studies. It was observed that the samples of Cook Islands Maori subjects performed at a somewhat lower level than New Zealand Maori and European subjects of approximately the same age. It was noted that the results conformed with the suggestion by McElwain and Kearney (1973) to the effect that performance on the QT by ethnic or cultural minority groups would be influenced by the degree of contact with societies of an urban and technological nature.

These findings would appear to give some support to views of McElwain and Kearney on the question of ethnic or cultural group differences in QT performance.

A very clear and consistent gender difference in QT performance was found in the Cook Islands study by St. George and St. George (1974). Males out-performed females at the three age levels studied and the difference tended to increase with age. While gender effects were considered in the present New Zealand studies no investigation was made of such factors in the original Australian studies. McElwain and Kearney (1970) when describing the Australian norms comment in relation to possible male-female differences that "This has not been the subject of detailed examination although any differences are probably small and negligible" (p.56). This assumption would appear to have been unwarranted.

Already many studies based on Anglo-American samples have reported gender differences in performances on tasks concerning

intellectual skills and processes (Guilford, 1967, pp.403-406, has provided a compact summary). Furthermore, cognitive organisation differences related to both gender and ecological-cultural factors have been widely studied in research based upon Witkin's psychological differentiation concept of field dependence/independence. (Witkin, Dyk, Faterson, Goodenough and Karp, 1962). Witkin and Berry (1975) have recently reviewed the evidence on gender differences in psychological differentiation from cross-cultural research. They concluded that -

"In these cross-cultural data sex differences are sometimes present and sometimes not, their presence seeming to be more common in samples at the sedentary agricultural end of the eco-cultural dimension and to be associated with tight cultural variables such as stratification and pressure toward conformity" (pp.36-37).

These findings suggest that the question of gender differences cannot be lightly dismissed.

It would also appear that from the comments of McElwain and Kearney (1970) that gender factors do indeed have some bearing upon QT performance, especially in non-Western situations. They point out that Aboriginal children may handle the stress of the testing situation by giggling and add that "Girls have another habit of covering their mouths with their hands as they do this" (pp.34-35).

The test authors when discussing the use of a translator in QT administrations consider the presence of another member of the community desirable "particularly with low contact female adults" (p.38). Furthermore, test users are reminded of the need to take care in enlisting the aid of a male translator who may be placed "in an inappropriate sexual relationship with a particular female subject" (p.38) by the testing situation.

Both cognitive organisation differences and socio-cultural factors of the test situation could be advanced as 'explanations' of gender differences in QT performance. Further research is really

required to establish particular causal relationships. Nonetheless, the question is of sufficient moment to have warranted some preliminary investigation in the Australian context.

The question of cognitive organisation differences between various ethnic group samples involved in QT research has been raised on a number of occasions. McElwain and Kearney (1970) discussing ethnic group comparisons suggest that the question of cognitive style differences is of both theoretical and practical importance. They consider that the question "probably requires 'micro-clinical studies' of cognitive processes of the kind that Piaget has done with European children" (p.3). Nonetheless, McElwain and Kearney (1970) have advanced a few observations on cognitive organisation differences when discussing the QT subtest performances of Australian Aboriginal and European subjects.

In particular, they allude to possible differences in the cognitive strategies adopted in the solution of items in the Form Assembly Subtest. It was observed that Aboriginal subjects tended to make errors in the choice of solution pieces related to area whereas European subjects' errors were of a shape (angularity) nature. This was especially so on Item 13. McElwain and Kearney (1970) noted some concordance between this 'finding' and the general observation Aboriginal peoples in bush situations "can point unerringly in the direction of places but seem to have poor judgement of distances or of areas" (p.89).

From the separate ethnic group factor analyses of the subtests undertaken in the Australian studies, McElwain and Kearney (1970) concluded however that "The factorial structure of the test appears to be fairly similar over the groups" (p.143). The New Zealand factor analytic evidence from the ethnic group studies pointed to differences in the loadings of subtests for European and Maori subjects in the

factors extracted. This finding was suggestive of cognitive organisation or style differences.

The discussion above on gender differences in the level of QT performance also bears upon the question of cognitive organisation differences. This is particularly so in relation to the Pattern Matching Subtest. Statistically significant differences favouring males on this subtest have been reported by St. George and St. George (1974) with Cook Islands subjects, by Chandra (1975) with Fijian subjects and were also evident in the present study.

In the Witkin and Berry (1975) review of cross-cultural studies on psychological differentiation they note that block design tasks, of which the Pattern Matching Subtest is a good example, share "with the standard tests of field-dependence-independence the requirement of perceptual disembedding" (p.18). Witkin and Berry (1975) summarizing gender contrasts in field-dependence/independence note that the studies do indicate that "males more often than not score in a more field-independent direction than females" (p.30). In the case of block design tasks higher scores are indicative of field-independence which would support the view that the male subjects in the above studies displayed a more field independent cognitive style relative to females.

Witkin and Berry (1975) have, however, pointed out that ecological-cultural factors have an important bearing upon the occurrence of gender differences in non-Western settings and therefore the finding need not hold true for all comparisons. The evidence to date points to the fact that performance on one QT subtest might be linked to cognitive style considerations when considering gender differences.

The question of ethnic group differences in aspects of cognitive organisation remains more obscure. The data on this matter across various ethnic group samples is ambiguous. Restrictions in terms of variables entered into the factor analyses which may have thrown light on this question have presented interpretation difficulties. The

suggestions of cognitive organisation differences through the factor analysis of QT data do not readily fit into particular cognitive organisation models that have a measure of independent theoretical and empirical substantiation (i.e. Witkin's (1967) field-independence/dependence model; Cohen's (1969) relational/analytical thinking model).

The whole question of obtaining evidence concerning cognitive organisation data for ethnic and/or gender groups from the QT must be viewed with some caution. It should not be overlooked that the test authors do not altogether encourage the drawing of inferences from subtest scores and subtest score differences. Subtest reliabilities would also need to be taken into account in attributing importance to subtest score pattern differences. Furthermore, the whole question of the construct validity of the QT subtests within designated groups would require much more thorough investigation. It was argued earlier that the issue of cognitive styles data stemming from current QT research was in the nature of 'spin-off'. As 'spin-off' it really requires investigation in its own right employing appropriate methodologies and instrumentation in research projects directed specifically to the cognitive organisation question.

PART 3 GENERAL CONSIDERATIONS

Consideration of developments in the study and measurement of intelligence reviewed in Chapter 1 led to the conclusion that Vernon's (1960, 1961) hierarchical model of intelligence presented a useful framework for the interpretation of intelligence test data. The evidence from previous studies with the QT was to the effect that it was a measure of general intelligence with a small spatial-perceptual component. The New Zealand studies support this interpretation. The factor analyses, whether for combined or separate ethnic groups, point first and foremost to a 'g' or general intelligence factor. This is

followed by some evidence of a spatial-perceptual factor akin to Vernon's spatial-mechanical (k:m) major group factor. The QT's relative independence in these studies from the verbal-educational (v:ed) skills that form Vernon's other major group factor conforms with findings elsewhere and the test author's original views about the QT. McElwain and Kearney's (1970) interpretation of the general intelligence construct follow those suggested by Burt (1955) and Vernon (1961) and the present studies did not bring forth any evidence to the contrary.

The limitations of the factor analytic evidence must be borne in mind however. Any future research on the factor structure of the QT should employ an enlarged test battery including previously researched and well established marker tests. In addition, when studying communalities and differences in the factor structures across ethnic or culture groups the factor analysis methods currently being developed by Sörbom (1974) and Buss and Royce (1975) could be employed to advantage.

In defining what the QT measures the data and arguments tended towards the form of psychological definition discussed by Butcher (1968). Butcher noted that psychological definitions suggest a genetic substrate to intelligence and its development while still acknowledging the importance of environmental variables that interact in the development and measurement of intelligence. This is more restrictive than a general biological definition stressing adaptation to an environment which Butcher argued was not really of great practical use in the study of individual differences. However, such a psychological definition is a little broader than a purely operational definition which would insist on defining intelligence only in terms of a specific measurement operation.

Operational definitions are, of course, important but as Eysenck (1973) has argued it is quite in order to develop concepts. These stand

or fall on their usefulness and appropriateness in the search for invariances, the forming of generalizations and the prediction of future events. The concept of general intelligence continues to be useful and appropriate in the ways suggested by Eysenck.

McElwain and Kearney (1970) have throughout considered the QT to be a measure of 'basic cognitive capacity' and the factor analytic work to date supports a general ability interpretation. Furthermore, a most important feature of the factor analytic data is the fact that clear evidence of this general intelligence factor has been found in the studies undertaken with different cultural and ethnic group samples.

Support for this interpretation can also be found in other aspects of the QT studies. First, throughout the Australian and New Zealand research QT performance has been found to be related to age in a manner consistent with the view that the growth of intelligence is subject to maturational processes during the period of general biological development. The QT relationship with age is relatively consistent across all the culture and ethnic groups studied.

Second, the psychological definition and interpretation offered acknowledges the interaction and impact of environmental variables upon the growth and development of general intelligence. McElwain and Kearney (1970) while stressing the 'culture-reduced' nature of the QT still believe that performance on the test will be influenced by experiential and environmental factors.

As was noted earlier, they have argued that the mean differences in the levels of QT performance between various Australian sample groups appear to be related to contact with mainstream Australian European society (see McElwain and Kearney, 1973). McElwain and Kearney are not, however, specific about the actual environmental variables related to contact that might be associated with improved QT performance. It

would be reasonable to assume that the environmental variables influencing intellectual development reviewed by Vernon (1969) would all play a part in the culture-contact continuum alluded to by McElwain and Kearney. While the data does not allow more than broad speculation on this question it does seem that the small ethnic group differences in QT performances between Maori and European subjects obtained in these New Zealand studies reflect the higher levels of shared environmental factors and experiences bearing upon general intellectual development.

It is of interest to note that in a very recent study Fredrickson (1977), on finding no statistically significant differences in mean IQs between samples of U.S. Black and White children, has employed a shared environments interpretation. He has commented upon the relatively similar and stable nature of the Black and White communities from which the children in his sample were drawn and suggests that it is the absence of equivalence in cultural contexts that brought about the mean IQ differences between these populations.

While it is encouraging to see a number of strands of evidence coinciding in support of the suggested interpretation of the QT it would be as well to recall the warning given earlier about the reification of psychological concepts. The general intelligence concept to which the QT has been linked by way of statistical analyses and argument is not the only conceptualization or approach. This was clearly shown by the reviewing of other definitions and theories in Chapter 1. Adcock and Webberley's (1971) simplification of Guilford's Structure of Intellect model could, with the development of suitable independent measures of the proposed primary mental abilities, assist in the promotion of the multiple factor viewpoint. Alternatively, Jensen's (1970, a,b,c,d and 1972a) theory of associative and conceptual learning abilities or Level I and Level II abilities cannot be lightly dismissed. Jensen has taken into account a considerable amount of psychometric evidence in propounding his

theory and more than some other conceptualizations has attempted to link the theory to instructional and educational processes.

Bronowski (1973) has said that -

"Every theory is based on some analogy, and sooner or later the theory fails because the analogy turns out to be false. A theory in its day helps to solve the problems of the day" (p.140).

The analogy behind the general intelligence concept is obviously not beyond question. It has nonetheless proved to be one of the more durable psychological concepts and the tests of general intelligence have assisted in the solution of some of the problems of the day.

The question of whether or not the QT might produce any useful data on cognitive organisation has already been discussed in some detail. It was argued that the meaning of differences between some aspects of the factor structure of the QT with European and Maori subjects were ambiguous. Already other researchers (Chapman, 1973; Harker, 1976) have made direct studies on the issue of cognitive style differences between Maori and European children finding no consistent evidence of such differences. These studies have started from particular cognitive style models and employed the associated measuring techniques. Under these circumstances the likelihood of obtaining differences, if operative, would seem much more likely. Furthermore, any differences could have been understood in terms of the theory associated with the cognitive style models. The lack of evidence on Maori-European cognitive style differences in these recent and more appropriate studies cautions against treating the Maori-European QT factor structure differences as fundamental.

Turning to gender differences in cognitive organisation, it would appear that here the Pattern Matching Subtest of the QT is tapping a reasonably well established dimension of difference, namely field-dependence versus independence. The type of item employed in this

subtest has been independently validated as a useful measure of this dimension. This suggests first that gender differences on QT performance, and especially on this subtest should always be considered when using this instrument. Second, it would be instructive to extract the existing data on male and female performances on this subtest from all the QT research in the South West Pacific region. Given appropriate supporting data a test could be made of the relationship between gender differences across cultures and the ecological, social and biological antecedents of Witkin's psychological differentiation dimension of field dependence-independence (see Witkin and Berry, 1975).

Along with questions concerning cognitive organisation it might also be asked whether or not the QT data reported in this study has any bearing upon the processes of intellectual growth as described by developmental theorists such as Piaget (1950). The QT's relationship with the psychometric approach to the study of intelligence would initially suggest that any connection is unlikely. However, a feature of the New Zealand data for both ethnic group samples was the absence of any real increment in QT performance between the age levels 10:6-11:5 and 11:6-12:5. This plateau was evident in all five QT subtests. An empirically substantiated explanation of this occurrence cannot be arrived at from present data. Nonetheless, it is interesting to note that in Piagetian terms the age range involved in this plateau in QT performance covers the transition from the stage of concrete operational thought to the stage of formal operations. Research exploring the relationships between the performances of subjects varying in age and stage of cognitive growth on particular QT items might be in order. Attention could be given to the differing strategies employed by subjects in arriving at a solution to the items. The cognitive development theory associated with such a study need not, of course, be that of Piaget.

The discussion of both cognitive organisation and cognitive development questions in relation to QT performance is admittedly speculative. Nevertheless, it should serve to indicate some of the more complex issues associated with the study of intelligence that may lie behind what appears to be a relatively simple and straightforward measure. In coming to any conclusions about the usefulness of the QT the data and issues pertaining to the theory of intelligence underlying the test must be viewed in conjunction with data relating to practical aspects of its operation.

From a practical point of view a most important feature of the New Zealand QT data was the evidence on only moderate reliability. Apart from having a bearing upon the validity data the evidence on reliability could not be considered satisfactory. It is difficult to know whether or not it was a feature peculiar to these studies. The evidence on non-discriminating items suggests, however, that the difficulties lie with the test as constituted rather than the research itself. While it might be in order to seek further data on QT reliability the need to restructure items within a number of subtests would appear to be the more important first step.

On the question of QT validity it is the data on criterion-relatedness which has most to do with the practical issues of test use. There was some evidence in this research of QT performance being significantly related to mathematics attainment and teacher rankings of ability. The test results did not relate significantly to scholastic criteria of a more linguistic nature. The data on criterion-related validity will have been influenced by the reliability of the QT and the reliability and validity of the criterion measures themselves. Inasmuch as it is apparent that the QT reliability was not altogether satisfactory, the reliability and validity of the criterion measures could also be questioned. It is accepted that the criteria employed in the validity studies can be criticised on these grounds. There is a real need for

psychologists and educators to undertake research upon criterion measures on a scale equal to the efforts put into their psychological and educational tests.

It would seem too that the range of criteria could also be extended. With the exception of Kearney's (1966a) study relating a productivity index (ownership of cash crop coconut trees) to performances on a test which was a forerunner to the QT, criteria have in general been of a scholastic nature. This reflects no doubt the fact that the great majority of QT research has been with school aged samples where such data is readily available.

McElwain and Kearney (1970), in discussing the development of the QT, suggested that it would be an appropriate instrument for the selection of individuals with a capacity to rapidly learn complex skills. This suggests that QT criterion-related validity research should begin to focus upon learning itself. Ferguson (1954, 1956) has suggested that overlearning and positive transfer are important aspects of a person's performance in new learning situations. What we call 'ability' is overlearned skills available for use in familiar situations and for transfer to new situations. Claims that the QT predicts rapid learning could be investigated employing Ferguson's concepts of overlearning and positive transfer.

Studies such as those by Duncanson (1966) and Schroth (1976) do show measured abilities to be positively related to various learning tasks and Estes (1974) has provided a useful theoretical discussion on learning theory and intelligence. He sees a need for a much closer relationship between these fields with test behaviour being interpreted in terms of concepts drawn directly from learning theory. Research of this nature could lead to a much clearer understanding of the criterion-related validity of general ability measures.

Measuring intellectual abilities across cultures was shown in Chapter 1 to be an exceptionally complex issue. Within social scientific

literature there are theories both for and against the existence of basic cognitive universals between cultural and ethnic groups. At present the combined weight of evidence from linguistics, anthropology and psychology tends to support the 'underlying cognitive universals' point of view. This position recognises, however, that social, cultural and environmental factors greatly influence the manifestation of particular intellectual abilities.

The long history of applying measures of intelligence to subjects from differing cultural backgrounds has been accompanied by many changes in psychological theory and practice. It has become slowly recognised that the attempt to classify and rank humanity on measures of intelligence has often been motivated more by politics and ideology than by any sound scientific reasoning (see Kamin, 1974). The culture-bound nature of the theories of intelligence and the methods of measurement have led to their radical reappraisal. The theories have become much more conditional in nature and the methods of measurement subject to considerable change and adaptation in order to facilitate cross-cultural studies. The relativist position that has come to the fore of late suggests that even this is not enough. This view has been important in cueing psychologists to the within-culture perspective and to an appreciation of indigenous conceptualizations of intelligent behaviour. However, a number of logical problems with relativistic arguments were noted. These have a bearing upon cross-cultural study of intelligence.

So long as general theory building remains an important objective of psychology it will be necessary to advance theories that transcend particular culture systems. They may be theories of learning, of social interaction, of motivation or even intelligence. In this regard it can be noted that both Price-Williams (1975) and Triandis (1976) have suggested that cross-cultural psychological research, including that of a comparative

nature, will contribute much to the development and testing of our general psychological theories. The theories may with the passage of time be found wanting.

From a more practical point of view we have also seen that Vernon (1969) believes that a convergence of social, economic and technological forces across culture systems is having a moulding effect upon psychological characteristics of populations. Consequently, Vernon sees the cross-cultural application of psychological theories and methods (admittedly Western in origin) as less of a problem than some others might suggest. Price-Williams (1975), for instance, accepts that a pervasive cultural diffusion process is taking place, most probably via school systems, which is leading to greater degrees of psychological congruence across cultural and ethnically diverse groups. It does seem that this cultural diffusion process has encouraged the development of Western-type abilities in what were originally non-Western populations (see Irvine, 1969a,b). To the extent that this has occurred then Vernon sees the use of measures of ability developed from Western psychological theory and experience as justified.

Without detracting from the important and continuing cultural differences that exist between Maori and European in New Zealand, it remains fair to suggest that cultural diffusion has been considerable. Maori commentators such as Walker (1973) acknowledge this especially in relation to the adaptations the Maori people have made to the dominant European society. One of the most influential institutions in this process of cultural diffusion has been the school. It has been both a mediator and model to many Maori children and parents of the intellectual, educational, social, attitudinal and behavioural correlates of a Western-type technological society.

In addition to arguments that might be made concerning the extent of various cultural diffusion processes justifying the cross-cultural use

of an intelligence test, there are also certain empirical steps that can be taken to evaluate the cross-cultural suitability of a particular test. The psychological equivalence of the QT for Maori and European subjects has been an important issue in this research. It would seem reasonable to argue that cultural diffusion processes have created a measure of psychological congruence between Maori and European (perhaps especially in the school populations in recent years) thus providing some justification for the use of the QT. The next important issue is then to consider what conclusions might be reached concerning the psychological equivalence of the QT with the groups concerned on the basis of the empirical evidence to hand.

From the comparisons of mean QT performances it was found that European subjects obtained slightly higher scores than their Maori counterparts. However, ethnic group membership was not in itself a major factor in determining QT performance and it was concluded that the small ethnic group difference was of little practical importance.

While the absence of large ethnic group difference in QT performance is an important finding, it does not directly address the issue of psychological equivalence. Cleary, Humphreys, Kendrick and Wesman (1975), Drenth and van der Flier (1976) and others have discussed this issue in relation to such matters as test reliability, test validity and item statistics across cultural groups.

QT reliability, validity and item statistics were fully investigated for both European and Maori samples. The reliability data for the two ethnic groups were comparable although, as previously noted, the levels of reliability attained by some estimates were not encouraging. This applied equally to both ethnic groups.

The criterion-related validity data were similar for the ethnic groups with correlations tending to be of a moderate positive order with the teacher rankings of ability and mathematics assessments, but lower

with criteria associated with language skills.

The data bearing upon construct validity tended to support the use of the same general intelligence construct as an interpretative framework for both ethnic groups. Differences in the pattern and loadings of QT subtests were considered as possible indicators of construct differences, but the present data do not really allow any firm conclusions to be drawn on this. For both Maori and European samples the QT performances related to age of subjects in a manner consistent with the general intelligence theory originally associated with the test.

The item statistics derived for the two ethnic groups did not differ markedly one from the other.

Overall the weight of the empirical evidence is in support of the conclusion that the QT operates in a psychologically equivalent manner with regards Maori and European subjects of the ages studied in this research. Encouraging though this may be, the evidence on psychological equivalence would need to be weighed against the actual levels of reliability and the limited nature of the validity data in making a decision on future use.

Unfortunately, empirical data on just one instrument is insufficient. Ideally, evidence on the equivalence or otherwise of psychological and educational measures employed in New Zealand on any large scale should be sought for the major ethnic and cultural groupings. Apart from this work with the QT these issues have until recently been of little concern to others engaged upon either test development studies or even cross-cultural research on intelligence itself within New Zealand.

Since this comparative research with the QT focused on older subjects than those in the studies of Brooks (1973), Klippel (1973) or A. St. George (1972) direct comparisons of the results are not possible. Nevertheless, it remains to be seen if collectively the results from

these recent studies form anything approaching a coherent pattern.

In reviewing in Chapter 3 the many comparative studies undertaken on the levels of European and Maori intellectual ability, it was found that the results were not as consistent as had often been assumed. Maori subjects did not always perform less well than their European counterparts, differences in performance were not always less on 'non-verbal' measures and differences were at times assumed to be significant when statistically they were not so. It was found, however, that so far as the three recent studies mentioned above were concerned, ethnic group differences were small and few in number. To this we can now add the results from the current studies with older subjects. In the absence of dramatic ethnic group differences in these studies all of the authors concerned have been drawn to the conclusion that sweeping assertions about differences in the level of measured intelligence between Maori and European children cannot be empirically sustained. It is interesting to add to this the conclusions reached by researchers working in the cognitive styles domain where in the recent studies there has been a similar absence of consistent or marked ethnic group differences.

The short review on Maori educational underachievement presented in Chapter 2 showed that both differences in the level of intellectual ability and in the organisation of intellectual abilities have been suggested from time to time as reasons for the lower scholastic attainments of Maori pupils in comparison with European pupils. Collectively, the data from the recent studies on both the levels of measured intelligence and the organisation of intellectual abilities suggests that the reasons will have to be sought as vigorously elsewhere as in the 'heads of the children'. If the measures employed in these various studies are accepted as reasonable indicators of general ability then we do not have evidence of any pervasive ethnic group differences. Nor do we have any strong or consistent evidence on cognitive organisation differences.

It will be recalled that Harker (1973) has argued for an interaction model in the study of Maori educational underachievement. Amongst the sets of major factors suggested to be at least partially responsible for the ethnic group educational attainment differences lay the issue of intellectual ability and functioning. The evidence from this research when viewed in conjunction with other recent findings clearly indicates that intellectual ability factors per se cannot be invoked as adequate or sufficient 'explanations' of the educational attainment disparity. Reasons will have to be sought elsewhere in the complex of interacting cultural, social and educational factors that Harker and many others have suggested.

Whatever direction future research on the Maori educational attainment issue might take, an important outcome of this research and the related studies will have been the provision of sets of baseline data on a number of measures of ability and cognitive organisation. For some of these measures the process of establishing sets of psychometric data essential to their interpretation within and across specific New Zealand culture groups has begun.

The data on the QT presented in this thesis is one such set - probably the most comprehensive to date. The findings can be addressed to many specific questions concerning the QT and the performances of Maori and European children. Answers can now be given to some of the questions concerning QT levels of performance, reliability, validity and item characteristics. They are not the final answers, there are no final answers. Further research with the QT, if undertaken, will add to the findings and probably alter the answers. It may even begin to address some of the issues this research has inevitably left unresolved.

APPENDIX A

Queensland Test Record Form.

(Photographically reduced, Passalong
Subtest starting position illustrations
originally in colour)

1. KNOX CUBES

Item	Sequence	Response	α/β
PR I	1 2 3 4		
1	1 4 3 2		
2	4 1 2 3		
3	2 1 4 3		
4	3 1 4 2		
5	2 4 1 3		
6	3 2 1 4		
FR II	1 2 3 4 1		
7	1 3 1 2 4		
8	4 3 2 3 1		
9	3 2 4 3 1		
10	2 1 3 4 2		
11	1 2 3 2 4 1		
12	2 1 4 2 1 3		
13	3 1 2 4 1 3		
14	1 4 2 3 2 4 1		
15	3 1 4 2 4 1 3		

Scoring: One point for each test item correct.
Discontinue after 3 consecutive failures.

2. BEADS

Item	Exposure Time	Sequence	Response	α/β
PR I	unlimited	= 0 +		
PR II	5"	+ = 0		
1	10"	+ 0 = +		
2	10"	= + 0 +		
3	10"	0 + = 0 =		
4	10"	+ 0 + = 0		
5	15"	+ = 0 = + =		
6	15"	= 0 + 0 = +		
7	20"	0 = + = 0 + =		
8	20"	+ 0 = + = 0 =		
9	25"	= + = 0 = + = 0		
10	25"	0 + = + 0 + = 0 +		

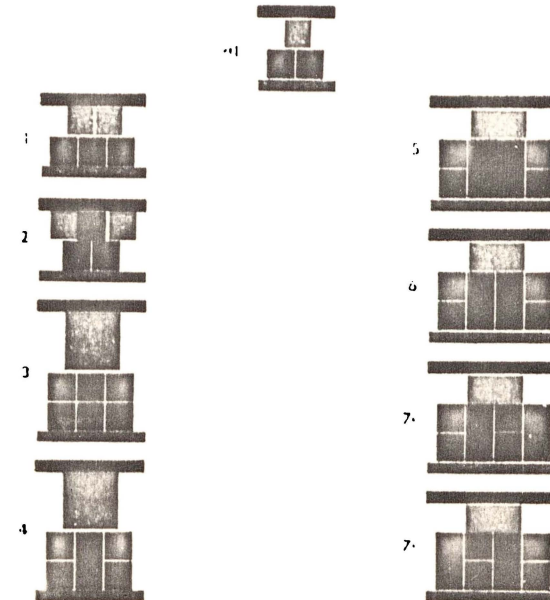
Scoring:
One point for each test item correct. Discontinue after 3 consecutive failures.

3. PASSALONG

Item	Time Limit	α/β
1	60"	
2	90"	
3	90"	
4	90"	
5	90"	
6	120"	
7	120"	

Scoring: One point for each test item correct within the time limit. Discontinue after failure of both items of a pair, i.e., 1 and 2, 3 and 4, or 5 and 6. There are alternative seventh items, 7a and 7b. If the red tile is moved to the subject's right in item 6, administer 7a; if it is moved to his left, administer 7b.

STARTING POSITION



Note: The blue end of the frame is towards the subject.

APPENDIX B

QT summary statistics and
intercorrelation analyses for
class group samples in test-
retest reliability studies.

;

TABLE B1 Means and Standard Deviations for Age, QT Subtests
and QT Total Score at Test and Retest Condition -
Class Group 1 (N=27), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	116.85	6.73	123.81	6.58
Knox	6.81	1.84	8.81	1.73
Beads	3.30	1.71	4.78	2.56
Passalong	4.04	1.40	5.30	1.35
Form Assembly	8.81	2.00	9.00	2.02
Pattern Matching	12.07	2.29	13.22	2.08
QT Total	35.04	6.09	41.11	6.57

TABLE B2 Means and Standard Deviations for Age, QT Subtests
and QT Total Score at Test and Retest Condition -
Class Group 2 (N=35), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	126.11	4.82	132.97	4.82
Knox	7.31	2.25	7.80	2.39
Beads	4.06	2.04	5.40	2.50
Passalong	4.89	1.23	6.09	0.85
Form Assembly	9.26	1.50	9.80	1.62
Pattern Matching	13.77	1.24	14.34	1.00
QT Total	39.29	4.93	43.43	4.97

TABLE B3 Means and Standard Deviations for Age, QT Subtests
and QT Total Score at Test and Retest Condition -
Class Group 3 (N=34), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	139.00	4.81	145.76	4.81
Knox	8.03	2.04	8.65	1.59
Beads	4.32	2.24	6.06	2.55
Passalong	5.03	1.29	5.38	0.99
Form Assembly	9.35	1.99	10.00	1.23
Pattern Matching	13.44	1.97	14.47	0.93
QT Total	40.18	4.71	44.56	4.91

TABLE B4 Means and Standard Deviations for Age, QT Subtests
and QT Total Score at Test and Retest Condition -
Class Group 4 (N=31), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	148.81	4.75	155.81	4.87
Knox	7.84	2.13	8.23	2.08
Beads	5.19	2.30	6.13	2.53
Passalong	5.13	1.20	5.68	0.94
Form Assembly	9.32	1.49	9.58	2.32
Pattern Matching	14.32	0.94	14.71	0.78
QT Total	41.81	5.33	44.32	5.29

TABLE B5 Means and Standard Deviations for Age, QT Subtests
and QT Total Score at Test and Retest Condition -
Class Groups 1 to 4 Combined (N=127),
Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	133.13	12.93	140.02	12.93
Knox	7.53	2.11	8.35	2.00
Beads	4.24	2.17	5.62	2.56
Passalong	4.80	1.33	5.63	1.07
Form Assembly	9.20	1.59	9.63	1.62
Pattern Matching	13.46	1.83	14.23	1.35
QT Total	39.24	5.69	43.46	5.50

TABLE B6 Means and Standard Deviations for Age, QT Subtests and QT Total Score at Test and Retest Condition -
Class Group 1 Europeans (N=17), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	114.47	6.17	121.53	6.08
Knox	6.76	2.11	8.88	1.87
Beads	3.47	1.62	4.88	2.32
Passalong	4.12	1.27	5.35	1.32
Form Assembly	8.76	2.22	9.29	1.90
Pattern Matching	12.47	2.12	13.76	1.25
QT Total	35.59	6.08	42.18	5.74

TABLE B7 Means and Standard Deviations for Age, QT Subtests and QT Total Score at Test and Retest Condition -
Class Group 1 Maori (N=10), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	120.90	5.86	127.70	5.72
Knox	6.90	1.37	8.70	1.57
Beads	3.00	1.89	4.60	3.06
Passalong	3.90	1.66	5.20	1.48
Form Assembly	8.90	1.66	8.50	2.22
Pattern Matching	11.40	2.50	12.30	2.87
QT Total	34.10	6.31	39.30	7.78

TABLE B8 Means and Standard Deviations for Age, QT Subtests and QT Total Score at Test and Retest Condition -
Class Group 2 European (N=27), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	125.89	4.42	132.74	4.46
Knox	7.48	2.31	7.70	2.40
Beads	4.22	1.95	5.70	2.28
Passalong	4.85	1.35	6.15	0.91
Form Assembly	8.96	1.51	9.70	1.75
Pattern Matching	13.74	1.26	14.30	0.87
QT Total	39.26	5.30	43.56	5.21

TABLE B9 Means and Standard Deviations for Age, QT Subtests and QT Total Score at Test and Retest Condition -
Class Group 2 Maori (N=8), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	126.88	6.23	133.75	6.18
Knox	6.75	2.05	8.13	2.47
Beads	3.50	2.39	4.38	3.07
Passalong	5.00	0.76	5.88	0.64
Form Assembly	10.25	1.04	10.13	1.13
Pattern Matching	13.88	1.25	14.50	1.41
QT Total	39.38	3.66	43.00	4.38

TABLE B10 Means and Standard Deviations for Age, QT Subtests
and QT Total Score at Test and Retest Condition -
Class Group 3 European (N=18), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	139.17	4.00	145.11	4.61
Knox	7.50	1.92	8.22	1.86
Beads	4.28	2.44	6.22	2.21
Passalong	5.28	1.02	5.44	0.70
Form Assembly	9.61	1.29	9.89	1.23
Pattern Matching	13.56	1.76	14.39	0.98
QT Total	40.22	4.45	44.17	4.02

TABLE B11 Means and Standard Deviations for Age, QT Subtests
and QT Total Score at Test and Retest Condition -
Class Group 3 Maori (N=16), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	138.81	5.71	146.50	5.09
Knox	8.63	2.06	9.13	1.09
Beads	4.38	2.06	5.88	2.94
Passalong	4.75	1.53	5.31	1.25
Form Assembly	9.06	1.53	10.13	1.26
Pattern Matching	13.31	2.24	14.56	0.89
QT Total	40.13	5.14	45.00	5.85

TABLE B12 Means and Standard Deviations for Age, QT Subtests and QT Total Score at Test and Retest Condition -
 Class Group 4 European (N=18), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	146.50	4.01	153.44	4.15
Knox	8.06	2.07	8.17	2.12
Beads	5.39	2.50	6.56	2.26
Passalong	5.44	1.10	5.89	0.47
Form Assembly	9.61	1.38	9.94	1.21
Pattern Matching	14.61	0.61	14.83	0.38
QT Total	43.11	4.85	45.39	3.73

TABLE B13 Means and Standard Deviations for Age, QT Subtests and QT Total Score at Test and Retest Condition -
 Class Group 4 Maori (N=13), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	152.00	3.81	159.08	3.86
Knox	7.54	2.26	8.31	2.10
Beads	4.92	2.06	5.54	2.85
Passalong	4.69	1.25	5.38	1.33
Form Assembly	8.92	1.61	9.08	1.80
Pattern Matching	13.92	1.19	14.54	1.13
QT Total	40.00	5.61	42.85	6.81

TABLE B14 Means and Standard Deviations for Age, QT Subtests and QT Total Score at Test and Retest Condition - European Subjects, Class Groups 1 to 4 Combined (N=80), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	131.09	12.60	137.80	12.49
Knox	7.46	2.13	8.18	2.12
Beads	4.34	2.20	5.84	2.30
Passalong	4.93	1.28	5.76	0.94
Form Assembly	9.21	1.63	9.71	1.56
Pattern Matching	13.63	1.64	14.33	0.96
QT Total	39.56	5.70	43.81	4.82

TABLE B15 Means and Standard Deviations for Age, QT Subtests and QT Total Score at Test and Retest Condition - Maori Subjects, Class Groups 1 to 4 Combined (N=47), Interval 6 Months

	Test		Retest	
	Mean	S.D.	Mean	S.D.
Age (in months)	136.62	12.88	143.81	12.92
Knox	7.64	2.08	8.64	1.76
Beads	4.09	2.15	5.26	2.93
Passalong	4.60	1.39	5.04	1.23
Form Assembly	9.19	1.54	9.49	1.73
Pattern Matching	13.17	2.10	14.06	1.83
QT Total	38.68	5.70	42.85	6.51

TABLE B16 Product-Moment Correlation Coefficients for Class Group 1
Between QT Subtests and QT Total Score at Test Condition (N=27)

	1	2	3	4	5
1. Knox					
2. Beads	.67				
3. Passalong	.29	.32			
4. Form Assembly	-.03	.48	.10		
5. Pattern Matching	.23	.61	.35	.41	
6. QT Total	.46	.76	.57	.63	.83

Values of .49 and above significant, $p < .01$, two-tailed test.

TABLE B17 Product-Moment Correlation Coefficients for Class Group 1
Between QT Subtests and QT Total Score at Retest Condition (N=27)

	1	2	3	4	5
1. Knox					
2. Beads	.26				
3. Passalong	.22	.57			
4. Form Assembly	-.05	.32	.25		
5. Pattern Matching	.23	.55	.43	.22	
6. QT Total	.46	.85	.70	.54	.75

Values of .49 and above significant, $p < .01$, two-tailed test.

TABLE B18 Product-Moment Correlation Coefficients for Class Group 2
Between QT Subtests and QT Total Score at Test Condition (N=35)

	1	2	3	4	5
1. Knox					
2. Beads	.17				
3. Passalong	.33	.22			
4. Form Assembly	.04	.09	.13		
5. Pattern Matching	.14	.30	.35	.16	
6. QT Total	.66	.65	.62	.43	.57

Values of .43 and above significant, $p < .01$, two-tailed test.

TABLE B19 Product-Moment Correlation Coefficients for Class Group 2
Between QT Subtests and QT Total Score at Retest Condition (N=35)

	1	2	3	4	5
1. Knox					
2. Beads	.07				
3. Passalong	.24	.11			
4. Form Assembly	.25	.10	.10		
5. Pattern Matching	.17	.13	.34	.33	
6. QT Total	.67	.62	.44	.58	.51

Values of .43 and above significant, $p < .01$, two-tailed test.

TABLE B20 Product-Moment Correlation Coefficients for Class Group 3
Between QT Subtests and QT Total Score at Test Condition (N=34)

	1	2	3	4	5
1. Knox					
2. Beads	.12				
3. Passalong	.02	-.17			
4. Form Assembly	-.19	.22	.43		
5. Pattern Matching	.30	.06	.32	.19	
6. QT Total	.45	.47	.46	.52	.72

Values of .43 and above significant, $p < .01$, two-tailed test.

TABLE B21 Product-Moment Correlation Coefficients for Class Group 3
Between QT Subtests and QT Total Score at Retest Condition (N=34)

	1	2	3	4	5
1. Knox					
2. Beads	.27				
3. Passalong	.32	.26			
4. Form Assembly	.05	.35	.22		
5. Pattern Matching	.30	.44	.46	.21	
6. QT Total	.60	.83	.58	.53	.66

Values of .43 and above significant, $p < .01$, two-tailed test.

TABLE B22 Product-Moment Correlation Coefficients for Class Group 4
Between QT Subtests and QT Total Score at Test Condition (N=31)

	1	2	3	4	5
1. Knox					
2. Beads	.25				
3. Passalong	.19	.28			
4. Form Assembly	.28	.43	.22		
5. Pattern Matching	.13	.23	.31	.44	
6. QT Total	.65	.76	.54	.70	.52

Values of .46 and above significant, $p < .01$, two-tailed test.

TABLE B23 Product-Moment Correlation Coefficients for Class Group 4
Between QT Subtests and QT Total Score at Retest Condition (N=31)

	1	2	3	4	5
1. Knox					
2. Beads	.35				
3. Passalong	.26	.06			
4. Form Assembly	.18	.35	.27		
5. Pattern Matching	.29	.32	.27	.54	
6. QT Total	.72	.77	.43	.65	.62

Values of .46 and above significant, $p < .01$, two-tailed test.

TABLE B24 Product-Moment Correlation Coefficients for Class Groups
1 to 4 Combined Between QT Subtests and QT
Total Score at Test Condition (N=127)

	1	2	3	4	5
1. Knox					
2. Beads	.14				
3. Passalong	.25	.21			
4. Form Assembly	.05	.30	.24		
5. Pattern Matching	.24	.34	.40	.31	
6. QT Total	.57	.68	.60	.57	.72

Values of .23 and above significant, $p < .01$, two-tailed test.

TABLE B25 Product-Moment Correlation Coefficients for Class Groups
1 to 4 Combined Between QT Subtests and QT Total
Score at Retest Condition (N=127)

	1	2	3	4	5
1. Knox					
2. Beads	.21				
3. Passalong	.17	.25			
4. Form Assembly	.10	.29	.22		
5. Pattern Matching	.14	.40	.39	.32	
6. QT Total	.56	.77	.53	.59	.66

Values of .23 and above significant, $p < .01$, two-tailed test.

TABLE B26 Product-Moment Correlation Coefficients for Class Group 1
European Subjects Between QT Subtests and QT
Total Score at Test Condition (N=17)

	1	2	3	4	5
1. Knox					
2. Beads	-.04				
3. Passalong	.38	.15			
4. Form Assembly	.07	.60	.14		
5. Pattern Matching	.38	.40	.14	.40	
6. QT Total	.57	.65	.48	.72	.76

Values of .61 and above significant, $p < .01$, two-tailed test.

TABLE B27 Product-Moment Correlation Coefficients for Class Group 1
Maori Subjects Between QT Subtests and QT
Total Score at Test Condition (N=10)

	1	2	3	4	5
1. Knox					
2. Beads	.34				
3. Passalong	.14	.50			
4. Form Assembly	-.39	.28	.04		
5. Pattern Matching	-.02	.85	.57	.52	
6. QT Total	.25	.91	.68	.48	.93

Values of .77 and above significant, $p < .01$, two-tailed test.

TABLE B28 Product-Moment Correlation Coefficients for Class Group 1
European Subjects Between QT Subtests and QT
Total Score at Retest Condition (N=17)

	1	2	3	4	5
1. Knox					
2. Beads	.43				
3. Passalong	.37	.42			
4. Form Assembly	-.24	.25	.11		
5. Pattern Matching	.28	.79	.24	.27	
6. QT Total	.57	.90	.61	.44	.77

Values of .61 and above significant, $p < .01$, two-tailed test.

TABLE B29 Product-Moment Correlation Coefficients for Class Group 1
Maori Subjects Between QT Subtests and QT
Total Score at Retest Condition (N=10)

	1	2	3	4	5
1. Knox					
2. Beads	.00				
3. Passalong	-.07	.76			
4. Form Assembly	.24	.39	.44		
5. Pattern Matching	.22	.47	.61	.11	
6. QT Total	.34	.82	.83	.61	.75

Values of .77 and above significant, $p < .01$, two-tailed test.

TABLE B30 Product-Moment Correlation Coefficients for Class Group 2
European Subjects Between QT Subtests and QT Total
Score at Test Condition (N=27)

	1	2	3	4	5
1. Knox					
2. Beads	.16				
3. Passalong	.50	.29			
4. Form Assembly	.06	.15	.17		
5. Pattern Matching	.34	.31	.34	.16	
6. QT Total	.72	.63	.71	.44	.63

Values of .49 and above significant, $p < .01$, two-tailed test.

TABLE B31 Product-Moment Correlation Coefficients for Class Group 2
Maori Subjects Between QT Subtests and QT Total
Score at Test Condition (N=8)

	1	2	3	4	5
1. Knox					
2. Beads	.12				
3. Passalong	-.83	.00			
4. Form Assembly	.30	.23	-.37		
5. Pattern Matching	-.63	.31	.45	.14	
6. QT Total	.34	.89	-.21	.57	.32

Values of .83 and above significant, $p < .01$, two-tailed test.

TABLE B32 Product-Moment Correlation Coefficients for Class Group 2
European Subjects Between QT Subtests and QT Total
Score at Retest Condition (N=27)

	1	2	3	4	5
1. Knox					
2. Beads	.08				
3. Passalong	.45	.17			
4. Form Assembly	.31	.12	.13		
5. Pattern Matching	.38	.14	.53	.34	
6. QT Total	.74	.57	.58	.61	.61

Values of .49 and above significant, $p < .01$, two-tailed test.

TABLE B33 Product-Moment Correlation Coefficients for Class Group 2
Maori Subjects Between QT Subtests and QT
Total Score at Retest Condition (N=8)

	1	2	3	4	5
1. Knox					
2. Beads	.12				
3. Passalong	-.71	-.26			
4. Form Assembly	-.11	.19	.02		
5. Pattern Matching	-.31	.18	-.08	.40	
6. QT Total	.42	.84	-.45	.46	.37

Values of .83 and above significant, $p < .01$, two-tailed test.

TABLE B34 Product-Moment Correlation Coefficients for Class Group 3
European Subjects Between QT Subtests and QT
Total Score at Test Condition (N=18)

	1	2	3	4	5
1. Knox					
2. Beads	-.18				
3. Passalong	-.05	-.10			
4. Form Assembly	-.23	.30	.36		
5. Pattern Matching	.26	.03	.50	.28	
6. QT Total	.36	.55	.45	.55	.72

Values of .59 and above significant, $p < .01$, two-tailed test.

TABLE B35 Product-Moment Correlation Coefficients for Class Group 3
Maori Subjects Between QT Subtests and QT
Total Score at Test Condition (N=16)

	1	2	3	4	5
1. Knox					
2. Beads	-.07				
3. Passalong	.18	-.24			
4. Form Assembly	-.08	.16	.44		
5. Pattern Matching	.39	.10	.20	.11	
6. QT Total	.57	.39	.49	.51	.73

Values of .62 and above significant, $p < .01$, two-tailed test.

TABLE B36 Product-Moment Correlation Coefficients for Class Group 3
European Subjects Between QT Subtests and QT
Total Score at Retest Condition (N=18)

	1	2	3	4	5
1. Knox					
2. Beads	.30				
3. Passalong	.10	-.29			
4. Form Assembly	-.14	.36	-.21		
5. Pattern Matching	.14	.42	-.01	-.11	
6. QT Total	.64	.85	-.01	.37	.51

Values of .59 and above significant, $p < .01$, two-tailed test.

TABLE B37 Product-Moment Correlation Coefficients For Class Group 3
Maori Subjects Between QT Subtests and QT
Total Score at Retest Condition (N=16)

	1	2	3	4	5
1. Knox					
2. Beads	.38				
3. Passalong	.75	.52			
4. Form Assembly	.33	.36	.52		
5. Pattern Matching	.61	.49	.85	.59	
6. QT Total	.70	.84	.86	.66	.82

Values of .62 and above significant, $p < .01$, two-tailed test.

TABLE B38 Product-Moment Correlation Coefficients for Class Group 4
European Subjects Between QT Subtests and QT
Total Score at Test Condition (N=18)

	1	2	3	4	5
1. Knox					
2. Beads	.40				
3. Passalong	-.09	.45			
4. Form Assembly	.19	.27	.08		
5. Pattern Matching	.06	-.01	.19	-.19	
6. QT Total	.68	.86	.47	.50	.14

Values of .59 and above significant, $p < .01$, two-tailed test.

TABLE B39 Product-Moment Correlation Coefficients for Class Group 4
Maori Subjects Between QT Subtests and QT
Total Score at Test Condition (N=13)

	1	2	3	4	5
1. Knox					
2. Beads	-.01				
3. Passalong	.45	-.01			
4. Form Assembly	.33	.65	.24		
5. Pattern Matching	.11	.44	.26	.78	
6. QT Total	.62	.64	.52	.88	.70

Values of .68 and above significant, $p < .01$, two-tailed test.

TABLE B40 Product-Moment Correlation Coefficients for Class Group 4
European Subjects Between QT Subtests and QT
Total Score at Retest Condition (N=18)

	1	2	3	4	5
1. Knox					
2. Beads	.08				
3. Passalong	.25	.12			
4. Form Assembly	-.23	.18	-.32		
5. Pattern Matching	.25	.52	-.11	.23	
6. QT Total	.60	.78	.23	.29	.62

Values of .59 and above significant, $p < .01$, two-tailed test.

TABLE B41 Product-Moment Correlation Coefficients for Class Group 4
Maori Subjects Between QT Subtests and QT
Total Score at Retest Condition (N=13)

	1	2	3	4	5
1. Knox					
2. Beads	.70				
3. Passalong	.34	-.04			
4. Form Assembly	.61	.41	.41		
5. Pattern Matching	.38	.24	.30	.64	
6. QT Total	.89	.77	.44	.81	.61

Values of .68 and above significant, $p < .01$, two-tailed test.

TABLE B42 Product-Moment Correlation Coefficients for European Subjects, Class Groups 1 to 4 Combined, Between QT Subtests and QT Total Score at Test Condition (N=80)

	1	2	3	4	5
1. Knox					
2. Beads	.16				
3. Passalong	.31	.28			
4. Form Assembly	.07	.33	.23		
5. Pattern Matching	.34	.28	.38	.29	
6. QT Total	.62	.68	.63	.57	.69

Values of .28 and above significant, $p < .01$, two-tailed test.

TABLE B43 Product-Moment Correlation Coefficients for European Subjects, Class Groups 1 to 4 Combined, Between QT Subtests and QT Total Score at Retest Condition (N=80)

	1	2	3	4	5
1. Knox					
2. Beads	.15				
3. Passalong	.23	.17			
4. Form Assembly	.00	.23	.05		
5. Pattern Matching	.20	.47	.29	.25	
6. QT Total	.60	.75	.45	.49	.65

Values of .28 and above significant, $p < .01$, two-tailed test.

TABLE B44 Product-Moment Correlation Coefficients for Maori Subjects, Class Groups 1 to 4 Combined, Between QT Subtests and QT Total Score at Test Condition (N=47)

	1	2	3	4	5
1. Knox					
2. Beads	.11				
3. Passalong	.17	.07			
4. Form Assembly	.00	.26	.25		
5. Pattern Matching	.14	.42	.41	.35	
6. QT Total	.50	.66	.55	.56	.77

Values of .37 and above significant, $p < .01$, two-tailed test.

TABLE B45 Product-Moment Correlation Coefficients for Maori Subjects, Class Groups 1 to 4 Combined, Between QT Subtests and QT Total Score at Retest Condition (N=47)

	1	2	3	4	5
1. Knox					
2. Beads	.34				
3. Passalong	.15	.31			
4. Form Assembly	.31	.35	.42		
5. Pattern Matching	.14	.35	.45	.40	
6. QT Total	.57	.79	.61	.70	.67

Values of .37 and above significant, $p < .01$, two-tailed test.

APPENDIX C

QT and criteria data
summary statistics for
class group samples in
criterion-related validity
studies.

TABLE C1 Means and Standard Deviations for Age, QT Subtests, QT Total Score and Criteria Data: Class Group 1, Total Sample (N=27), European Sample (N=17) and Maori Sample (N=10)

	Total Sample		European Sample		Maori Sample	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (in months)	116.85	6.73	114.47	6.17	120.90	5.86
Knox	6.81	1.84	6.76	2.11	6.90	1.37
Beads	3.30	1.71	3.47	1.62	3.00	1.89
Passalong	4.04	1.40	4.12	1.27	3.90	1.66
Form Assembly	8.81	2.00	8.76	2.22	8.90	1.66
Pattern Matching	12.07	2.29	12.47	2.12	11.40	2.50
QT Total	35.04	6.09	35.59	6.08	34.10	6.31
P.R.S.T.	19.07	9.33	21.71	8.38	14.60	9.56
Otis Int.	24.81	12.08	29.88	10.55	16.20	9.61
PAT RV	23.11	11.57	27.59	11.53	15.50	6.95
PAT RC	17.11	8.73	21.47	7.02	9.70	5.96
Mathematics	57.41	24.06	65.41	20.98	43.80	23.72

TABLE C2 Means and Standard Deviations for Age, QT Subtests, QT Total Score and Criteria Data: Class Group 2, Total Sample (N=35), European Sample (N=27) and Maori Sample (N=8)

	Total Sample		European Sample		Maori Sample	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (in months)	126.11	4.82	125.88	4.42	126.87	6.27
Knox	7.31	2.25	7.48	2.31	6.75	2.05
Beads	4.06	2.04	4.22	1.95	3.50	2.39
Passalong	4.89	1.23	4.85	1.35	5.00	0.76
Form Assembly	9.26	1.50	8.96	1.51	10.25	1.03
Pattern Matching	13.77	1.24	13.74	1.26	13.88	1.25
QT Total	39.29	4.93	39.26	5.30	39.38	3.66
P.R.S.T.	25.77	11.80	26.70	12.94	22.63	6.30
Otis Int.	31.91	10.60	33.74	10.48	25.75	9.00
PAT RV	31.89	9.67	33.37	10.09	26.88	6.20
PAT RC	18.20	5.80	19.37	5.74	14.25	4.27
Mathematics	65.17	17.54	67.52	14.78	57.25	24.28

TABLE C3 Means and Standard Deviations for Age, QT Subtests, QT Total Score and Criteria Data: Class Group 3, Total Sample (N=34), European Sample (N=18) and Maori Sample (N=16)

	Total Sample		European Sample		Maori Sample	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (in months)	139.00	4.81	138.55	4.66	139.50	5.07
Knox . .	8.03	2.04	7.50	1.92	8.63	2.06
Beads	4.32	2.24	4.06	2.36	4.63	2.13
Passalong	5.03	1.29	5.11	1.13	4.94	1.48
Form Assembly	9.35	1.99	9.50	1.25	9.19	1.60
Pattern Matching	13.44	1.97	13.50	1.72	13.38	2.28
QT Total	40.18	4.71	39.67	4.03	40.75	5.46
P.R.S.T.	26.53	10.05	24.67	10.57	28.63	9.32
Otis Int.	37.18	17.16	36.83	16.41	37.56	18.51
PAT RV	30.06	15.09	32.28	14.80	27.56	15.49
PAT RC	16.09	10.39	16.94	10.56	15.13	10.45
Language Skills	62.44	22.73	61.28	24.46	63.75	21.35
Mathematics	54.68	23.49	54.94	24.78	54.38	22.75
Nature-Social St.	53.82	18.28	57.94	13.81	49.19	21.81

TABLE C4 Means and Standard Deviations for Age, QT Subtests, QT Total Score and Criteria Data: Class Group 4, Total Sample (N=31), European Sample (N=18) and Maori Sample (N=13)

	Total Sample		European Sample		Maori Sample	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (in months)	148.81	4.75	146.50	4.02	150.00	3.81
Knox	7.84	2.13	8.06	2.07	7.54	2.26
Beads	5.19	2.30	5.39	2.50	4.92	2.06
Passalong	5.13	1.20	5.44	1.10	4.69	1.25
Form Assembly	9.32	1.49	9.61	1.38	8.92	1.61
Pattern Matching	14.32	0.94	14.61	0.61	13.92	1.19
QT Total	41.81	5.33	43.11	4.85	40.00	5.61
P.R.S.T.	28.13	10.27	32.56	7.43	22.00	10.75
Otis Int.	40.65	16.90	46.61	15.03	32.38	16.34
PAT RV	35.45	13.11	39.22	13.59	30.23	10.82
PAT RC	17.29	9.08	19.28	9.10	14.54	8.65
Language Skills	69.25	19.99	72.06	20.93	66.31	18.90
Mathematics	65.68	18.37	71.94	15.87	57.00	18.60
Nature-Social St.	59.65	16.03	65.44	15.14	51.62	14.06

APPENDIX D

Tables on the proportion of
age group sample passing each
item of the five QT Subtests:
Data presented for Total Age Group
samples and European and
Maori Age Group samples.

TABLE D1 Proportion of Each Age Group
Sample Passing Each Item of Knox Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
8:6- 9:5	73	75	68	82	82	68	93	32	60	15	37	18	07	00	00	00
9:6-10:5	90	84	81	88	91	83	99	37	68	26	33	22	08	00	01	00
10:6-11:5	194	84	86	92	92	86	96	53	72	28	39	43	15	04	05	01
11:6-12:5	122	82	83	95	93	85	98	45	65	24	42	41	15	03	03	01
12:6-13:5	138	89	84	96	91	91	99	60	73	29	49	50	18	05	09	02
13:6-14:5	83	86	89	89	92	86	98	48	64	31	37	36	19	06	05	01

Decimal points omitted

TABLE D2 Proportion of Each Age Group
Sample Passing Each Item of Beads Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10
8:6- 9:5	73	74	82	32	45	22	21	22	15	14	03
9:6-10:5	90	70	84	32	47	23	28	19	21	21	11
10:6-11:5	194	81	89	38	64	30	40	33	29	39	16
11:6-12:5	122	80	89	43	66	32	38	28	28	41	10
12:6-13:5	138	89	96	53	80	49	56	54	51	66	25
13:6-14:5	83	87	92	60	70	43	54	43	41	64	24

Decimal points omitted

TABLE D3 Proportion of Each Age Group
Sample Passing Each Item of Passalong Subtest

Sample	N	1	2	3	4	5	6	7
8:6- 9:5	73	100	99	75	55	33	30	08
9:6-10:5	90	98	100	80	57	49	36	11
10:6-11:5	194	100	99	87	59	62	49	20
11:6-12:5	122	100	98	93	70	60	58	20
12:6-13:5	138	100	100	93	78	76	65	20
13:6-14:5	83	100	100	92	70	73	55	24

Decimal points omitted

TABLE D4 Proportion of Each Age Group
Sample Passing Each Item of Form Assembly Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13
8:6- 9:5	73	88	93	93	95	81	96	49	78	48	37	03	26	55
9:6-10:5	90	97	93	98	96	84	98	62	79	49	39	07	29	62
10:6-11:5	194	96	97	99	98	91	97	70	79	55	46	03	32	66
11:6-12:5	122	97	98	99	96	89	99	68	79	56	55	05	34	57
12:6-13:5	138	99	96	99	99	93	96	75	86	57	64	12	50	62
13:6-14:5	83	95	99	99	100	94	98	75	87	54	60	06	43	60

Decimal points omitted

TABLE D5 Proportion of Each Age Group
Sample Passing Each Item of Pattern Matching Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
8:6- 9:5	73	99	92	93	96	95	95	99	71	70	70	74	56	45	29	36
9:6-10:5	90	100	99	98	99	99	98	99	90	89	88	93	66	58	32	54
10:6-11:5	194	99	99	99	100	99	99	100	93	91	95	95	80	74	48	73
11:6-12:5	122	99	98	98	98	99	100	100	94	92	95	94	80	75	53	70
12:6-13:5	138	100	100	100	100	100	100	100	96	99	97	100	91	86	71	86
13:6-14:5	83	99	100	99	100	100	100	100	95	98	100	99	84	81	73	88

Decimal points omitted

TABLE D6 Proportion of Each European Age Group
Sample Passing Each Item of Knox Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
8:6- 9:5	36	64	72	86	81	67	92	31	67	11	44	25	08	00	00	00
9:6-10:5	57	86	82	89	93	81	100	35	68	28	33	25	11	00	02	00
10:6-11:5	125	82	86	92	90	82	96	51	71	29	34	36	14	04	02	01
11:6-12:5	63	87	79	97	92	86	98	46	67	14	38	46	10	03	05	00
12:6-13:5	78	88	85	97	90	88	99	58	68	29	47	50	17	06	08	01
13:6-14:5	38	84	87	89	95	84	100	45	63	37	39	34	16	05	03	00

Decimal points omitted

TABLE D7 Proportion of Each European Age Group
Sample Passing Each Item of Beads Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10
8:6- 9:5	36	67	83	44	39	25	17	22	19	14	03
9:6-10:5	57	67	84	39	53	32	28	23	26	25	14
10:6-11:5	125	85	89	36	69	34	44	33	31	42	18
11:6-12:5	63	79	94	48	68	35	40	30	27	48	08
12:6-13:5	78	88	92	58	77	50	56	53	55	67	23
13:6-14:5	38	89	89	63	66	45	61	45	47	66	32

Decimal points omitted

TABLE D8 Proportion of Each European Age Group
Sample Passing Each Item of Passalong Subtest

Sample	N	1	2	3	4	5	6	7
8:6- 9:5	36	100	97	81	61	31	28	00
9:6-10:5	57	98	100	81	63	58	42	12
10:6-11:5	125	100	100	88	62	69	50	21
11:6-12:5	63	100	100	98	79	65	63	21
12:6-13:5	78	100	100	96	85	87	65	22
13:6-14:5	38	100	100	92	82	79	63	26

Decimal points omitted

TABLE D9 Proportion of Each European Age Group
Sample Passing Each Item of Form Assembly Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13
8:6-9:5	36	94	100	94	92	86	97	56	86	44	39	06	22	58
9:6-10:5	57	96	91	98	100	88	96	63	75	46	40	07	33	63
10:6-11:5	125	97	98	99	98	93	97	69	82	55	53	02	35	62
11:6-12:5	63	98	98	100	97	94	100	76	75	52	52	06	46	52
12:6-13:5	78	100	95	100	99	99	99	79	90	50	74	17	45	63
13:6-14:5	38	95	97	100	100	97	100	84	89	58	71	08	45	58

Decimal points omitted

TABLE D10 Proportion of Each European Age Group
Sample Passing Each Item of Pattern Matching Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
8:6-9:5	36	97	97	97	100	97	97	100	81	75	81	86	69	67	36	47
9:6-10:5	57	100	100	100	100	98	100	100	95	91	93	93	74	68	39	65
10:6-11:5	125	99	99	100	100	100	100	100	94	90	96	94	82	81	55	74
11:6-12:5	63	100	100	98	100	100	100	100	94	90	97	95	90	78	56	78
12:6-13:5	78	100	100	100	100	100	100	100	100	99	97	100	95	91	76	88
13:6-14:5	38	100	100	100	100	100	100	100	97	97	100	100	92	89	79	95

Decimal points omitted

TABLE D11 Proportion of Each Maori Age Group
Sample Passing Each Item of Knox Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
8:6- 9:5	37	86	65	78	84	70	95	32	54	19	30	11	05	00	00	00
9:6-10:5	33	82	79	85	88	88	97	40	67	21	33	18	03	00	00	00
10:6-11:5	69	86	84	91	96	94	97	57	72	28	48	55	17	03	10	01
11:6-12:5	59	76	86	93	95	85	98	44	64	34	46	36	20	03	02	02
12:6-13:5	60	90	83	93	92	93	98	63	80	28	50	50	20	03	12	03
13:6-14:5	45	87	91	89	89	87	96	51	64	27	36	38	22	07	07	02

Decimal points omitted

TABLE D12 Proportion of Each Maori Age Group
Sample Passing Each Item of Beads Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10
8:6- 9:5	37	81	81	19	51	19	24	22	11	14	03
9:6-10:5	33	76	85	21	36	09	27	12	12	15	06
10:6-11:5	69	75	88	42	57	23	33	33	25	32	13
11:6-12:5	59	80	85	37	64	29	36	25	29	34	12
12:6-13:5	60	90	100	47	83	47	55	57	47	65	27
13:6-14:5	45	84	93	58	73	42	49	42	36	62	18

Decimal points omitted

TABLE D13 Proportion of Each Maori Age Group
Sample Passing Each Item of Passalong Subtest

Sample	N	1	2	3	4	5	6	7
8:6- 9:5	37	100	100	70	49	35	32	16
9:6-10:5	33	97	100	79	45	33	24	09
10:6-11:5	69	100	99	86	55	51	48	17
11:6-12:5	59	100	97	86	61	54	53	19
12:6-13:5	60	100	100	90	70	62	65	18
13:6-14:5	45	100	100	91	60	69	49	22

Decimal points omitted

TABLE D14 Proportion of Each Maori Age Group
Sample Passing Each Item of Form Assembly Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13
8:6- 9:5	37	81	86	92	97	76	95	43	70	51	35	00	30	51
9:6-10:5	33	97	97	97	88	79	100	61	85	55	36	06	21	61
10:6-11:5	69	94	96	100	99	87	97	72	72	54	33	03	26	74
11:6-12:5	59	95	97	98	95	83	98	59	83	59	58	03	20	63
12:6-13:5	60	98	98	97	100	87	93	70	82	65	52	07	57	62
13:6-14:5	45	96	100	98	100	91	96	67	84	51	51	04	42	62

Decimal points omitted

TABLE D15 Proportion of Each Maori Age Group
Sample Passing Each Item of Pattern Matching Subtest

Sample	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
8:6- 9:5	37	100	86	89	92	92	92	97	62	65	59	62	43	24	22	24
9:6-10:5	33	100	97	94	97	100	94	97	82	85	79	94	52	39	21	36
10:6-11:5	69	100	100	99	100	99	99	100	90	93	93	97	77	61	36	70
11:6-12:5	59	98	97	97	97	98	100	100	95	93	93	93	69	71	51	63
12:6-13:5	60	100	100	100	100	100	100	100	92	98	97	100	87	78	65	83
13:6-14:5	45	98	100	98	100	100	100	100	93	98	100	98	78	73	69	82

Decimal points omitted

REFERENCES

- Adams, R.S. (1970) Maori versus Pakeha. In Education at the crossroads. Special edition of the Social Studies Observer. Hamilton: Waikato Social Studies Association, p.48-53.
- Adams, R.S. (1973) Differentiating Maori from Pakeha: Some educational and social implications. In D.H. Bray and C.G.N. Hill. (Eds.) Polynesian and Pakeha in New Zealand education. Vol.1. Auckland: Heinemann, p.63-67.
- Adams, R.S. (1975) On ethnic differences: A response to Geraldine McDonald. New Zealand Journal of Educational Studies, 10, 151-152.
- Adcock, C.J., McCreary, J.R., Ritchie, J.E. and Somerset, H.C.A. (1954) An analysis of Maori scores on the Wechsler-Bellevue. Australian Journal of Psychology, 6, 16-29.
- Adcock, C.J. and Webberley, M. (1971) Primary mental abilities. Journal of Genetic Psychology, 84, 229-243.
- Alexander, W.P. (1932) A new performance test of intelligence. British Journal of Psychology, 23, 52-63.
- American Psychological Association. (1966) Standards for educational and psychological tests and manuals. Washington: A.P.A.
- American Psychological Association. (1974) Standards for educational and psychological tests. Washington: A.P.A.
- Anastasi, A. (1961) Psychological testing. 2nd ed. New York: Macmillan.
- Anastasi, A. (1964) Culture-fair testing. Educational Horizons, 43, 26-30.
- Anastasi, A. (1968) Psychological testing. 3rd ed. New York: Macmillan.
- Archer, D. and Archer, M. (1970) Race, identity and the Maori people. Journal of the Polynesian Society, 79, 201-218.
- Archer, D., Oppenheim, R.S., Karetu, T.S. and St. George, R. (1971) Intelligence and the Pakeha child. National Education, 53, 258-260.
- Arthur, G.A. (1930) Point scale of performance tests. Form I. Chicago: Stoelting.
- Arthur, G.A. (1943) A point scale of performance tests. Revised Form II. New York: Psychological Corporation.
- Ashcroft, E.R. (1963) Some aspects of testing Maori children. Education, 12, 30-32.

- Ausubel, D.P. (1961) Maori youth. Wellington: Price-Milburn.
- Awatere, D. (1972) Ethnic awareness and attitudes in Auckland children. M.A. thesis, University of Auckland.
- Baird, F. (1952) A comparative study of Maori and European children attending a rural public school. Dip.Ed. Investigation. Auckland University College.
- Ball, D.G. (1940) Maori education. In I.L.G. Sutherland (Ed.) The Maori people today. Wellington: N.Z.C.E.R.
- Banton, M. (1967) Race relations. London: Tavistock.
- Baratz, S.S. and Baratz, J.C. (1970) Early childhood intervention: The social science base of institutional racism. Harvard Educational Review, 40, 29-50.
- Barham, I.H. (1965) The English vocabulary and sentence structure of Maori children. Wellington: N.Z.C.E.R.
- Barney, D. (1975) Who gets to pre-school? The availability of pre-school education in New Zealand. Wellington: N.Z.C.E.R.
- Barrington, J.M. and Beaglehole, T.H. (1974) Maori schools in a changing society: An historical review. Wellington: N.Z.C.E.R.
- Bartlett, F.C. (1923) Psychology and primitive culture. Cambridge: Cambridge University Press.
- Beaglehole, E. and Beaglehole, P. (1946) Some modern Maoris. Wellington: N.Z.C.E.R.
- Beaglehole, J.C. (Ed.) (1955-69) The journals of Captain James Cook on his voyages of discovery. Cambridge: Hakluyt Society.
- Beaglehole, T.H. (1955) The Maori Schools, 1814 - 1880. M.A. thesis, Victoria University College.
- Beattie, R. (1971) Racial intelligence theory resurrected. The Christchurch Press, 111, October 9, 13.
- Bennett, M.J. (n.d.) Notes on the use of the Queensland Test in Fiji. Suva: University of the South Pacific.
- Bennett, M.J. and Chandra, S. (1974) Some ecological factors in individual test performance. In J.L.M. Dawson and W.J. Lonner (Eds.) Readings in cross-cultural psychology. Hong Kong: University of Hong Kong Press, p.185-189.
- Benton, R.A. (1972) Should bilingual schooling be fostered in New Zealand? Wellington: N.Z.C.E.R.
- Bernal, E.M. (1975) A response to "Educational uses of tests with disadvantaged subjects." American Psychologist, 30, 93-95.

- Bernardoni, L.C. (1964) A culture fair intelligence test for the Ugh, No and Oo-La-La cultures. Personnel and Guidance Journal, 42, 554-557.
- Bernstein, B. (1961) Social structure, language and learning. Educational Research, 3, 163-176.
- Berry, J.W. (1969) On cross-cultural comparability. International Journal of Psychology, 4, 119-128.
- Berry, J.W. (1971) Ecological and cultural factors in spatial-perceptual development. Canadian Journal of Behavioural Science, 3, 324-336.
- Berry, J.W. (1972) Radical cultural relativism and the concept of intelligence. In L.J. Cronbach and P.J.D. Drenth (Eds.) Mental tests and cultural adaptation. The Hague: Mouton, p.77-88.
- Berry, J.W. and Annis, R.C. (1974) Ecology, culture and psychological differentiation. International Journal of Psychology, 9, 173-193.
- Bhatia, C.M. (1955) Performance tests of intelligence under Indian conditions. London: Oxford University Press.
- Biesheuvel, S. (1949) Psychological tests and their application to non-European peoples. In G.B. Jeffery (Ed.) The Yearbook of Education. London: Evans, p.87-126.
- Biesheuvel, S. (1952) The study of African ability. African Studies, 11, 47-57, and 105-117.
- Biesheuvel, S. (1969) Methods for the measurement of psychological performance. Oxford: Blackwell.
- Binet, A. and Simon, T. (1905a) Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. L'Année Psychologique, 11, 191-244.
- Binet, A. and Simon, T. (1905b) Application des méthodes nouvelles au diagnostic du niveau intellectuel chez de enfants normaux et anormaux d' hospice et d' école primaire. L'Année Psychologique, 11, 245-336.
- Boas, F. (1911) The mind of primitive man. New York: Free Press.
- Back, R.D. and Vandenberg, S.G. (1968) Components of heritable variation in mental test scores. In S.G. Vandenberg (Ed.) Progress in Human Behavior Genetics. Baltimore: John Hopkins Press, p.233-260.
- Bonnevie, J. (1967) Maori children and creative thought. Dip.Ed. Investigation, Massey University.
- Brace, C.L., Gamble, G.R. and Bond, J.T. (Eds.) (1971) Race and Intelligence. Washington: American Anthropological Association.

- Bray, D.H. (1971) Maori adolescent temporal values: distance of goals perceived as important and of delayed gratification, as compared with Pakehas. New Zealand Journal of Educational Studies, 6, 62-77.
- Bray, D.H. and Hill, C.G.N. (Eds.) (1973, 1974) Polynesian and Pakeha in New Zealand education. Vols. 1 and 2. Auckland: Heinemann.
- Brazziel, W.F. (1969) A letter from the South. Harvard Educational Review, 39, 348-356.
- Brazziel, W.F. (1970) Beyond the sound and fury. Measurement and Evaluation in Guidance, 3, 7-9.
- Bridgman, C.S. (1964) The relation of the upper and lower item discrimination index, D, to the bivariate normal correlation coefficient. Educational and Psychological Measurement, 24, 85-90.
- Brimble, A.R. (1963) The construction of a non-verbal intelligence test in Northern Rhodesia. Rhodes-Livingstone Institute Journal, 34, 23-35.
- Brislin, R.W. (1976) Comparative research methodology: Cross-cultural studies. International Journal of Psychology, 11, 215-229.
- Brislin, R.W., Lonner, W.J. and Thorndike, R.M. (1973) Cross-Cultural Research Methods. New York: Wiley.
- Bronowski, J. (1973) The ascent of man. London: British Broadcasting Corporation.
- Brooks, I.R. (1973) A cross-cultural study of cognitive abilities in Maori and Pakeha four-year olds. M.Phil. thesis, University of Waikato.
- Brooks, I.R. (1975) A cross-cultural study of cognitive abilities in Maori and Pakeha four-year olds. Psychology Research Paper No. 3. University of Waikato.
- Brooks, I.R. and Cuttance, P.F. (1973) Socio-economic status in New Zealand. A critique of the Elley-Irving scale. New Zealand Journal of Educational Studies, 8, 154-159.
- Brophy, J.E. and Good, T.L. (1974) Teacher-student relationships: Causes and consequences. New York: Holt, Rinehart and Winston.
- Bruner, J.S. (1964) The course of cognitive growth. American Psychologist, 19, 1-15.
- Bruner, J.S. (1965) The growth of mind. American Psychologist, 20, 1007-1017.
- Bruner, J.S., Olver, R.R. and Greenfield, P.M. (1966) Studies in cognitive growth. New York: Wiley.

- Burch, W.R. (1967) Cross-cultural dialogues: Some trained incapacities of educators and some opportunities. New Zealand Journal of Educational Studies, 2, 113-126.
- Burt, C. (1949) The structure of the mind. A review of the results of factor analysis. British Journal of Educational Psychology, 19, 100-114 and 176-199.
- Burt, C. (1955) The evidence for the concept of intelligence. British Journal of Educational Psychology, 25, 158-177.
- Burt, C. (1958) The inheritance of mental ability. American Psychologist, 13, 1-15.
- Burt, C. (1963) Is intelligence distributed normally? British Journal of Statistical Psychology, 16, 175-190.
- Burt, C. (1969) Intelligence and heredity: Some common misconceptions. Irish Journal of Education, 3, 75-94.
- Burt, C. (1971) Heredity and environment. Bulletin of the British Psychological Society, 24, 9-15.
- Burt, C. (1972) Inheritance of general intelligence. American Psychologist, 27, 175-190.
- Buss, A.R. and Royce, J.R. (1975) Detecting cross-cultural communalities and differences: Intergroup factor analysis. Psychological Bulletin, 82, 128-136.
- Butcher, H.J. (1968) Human intelligence. Its nature and assessment. London: Methuen.
- Butchers, A.G. (1930) Education in New Zealand. Dunedin: Coulls, Somerville, Wilkie.
- Callander, D.J., Marsh, R.W. and Milne, K.A. (1967) Revised Tomlinson Junior School Test: Standardised for New Zealand schools. Wellington: Department of Education.
- Calvert, C.G. (1950) A preliminary enquiry into the language and other difficulties of Maori adolescents and the extent to which these are apparent in English attainment and intelligence test scores. M.A. thesis, Auckland University College.
- Campbell, D. (1958) The Rakau Studies. A review article.(B). A psychologist's view. Journal of the Polynesian Society, 67, 371-386.
- Cancro, R. (1971) Intelligence: Genetic and environmental influences. New York: Grune and Stratton.
- Cattell, J.McK. and Farrand, L. (1896) Physical and mental measurements of the students of Columbia University. Psychological Review, 3, 618-648.
- Cattell, R.B. (1940) A culture-free intelligence test. Journal of Educational Psychology, 31, 161-179.

- Cattell, R.B. (1944) Culture-Free Intelligence Test. Champaign: Institute for Personality and Ability Testing.
- Cattell, R.B. (1953) A guide to mental testing. 3rd ed. London: University of London Press.
- Cattell, R.B. (1963) Theory of fluid and crystallized intelligence. A critical experiment. Journal of Educational Psychology, 54, 1-22.
- Cattell, R.B. (1966) Handbook of multivariate experimental psychology. Chicago: Rand McNally.
- Cattell, R.B. (1971) Abilities: Their structure, growth and action. Boston: Houghton-Mifflin.
- Chandra, S. (1975) Some patterns of response on the Queensland Test. Australian Psychologist, 10, 185-192.
- Chapman, J.W. (1973) Identity status, occupational aspiration and field dependence in Maori and Pakeha boys. M.A. thesis, Victoria University of Wellington.
- Child, D. (1970) The essentials of factor analysis. London: Holt, Rinehart and Winston.
- Chomsky, N. (1968) Language and mind. New York: Harcourt, Brace and World.
- Cleary, T.A., Humphreys, L.G., Kendrick, S.A. and Wesman, A. (1975) Educational uses of tests with disadvantaged students. American Psychologist, 30, 15-41.
- Codd, J.A. (1972) Cultural factors in the cognitive development of Maori children. Delta, 11, 26-36.
- Cohen, J. (1959) The factorial structure of the WISC at ages 7-6, 10-6 and 13-6. Journal of Consulting Psychology, 23, 285-299.
- Cohen, R. (1969) Conceptual styles, culture conflict and non-verbal tests of intelligence. American Anthropologist, 71, 828-856.
- Cole, M. and Bruner, J.S. (1971) Cultural differences and inferences about psychological processes. American Psychologist, 26, 867-876.
- Cole, M., Gay, J., Glick, J.A. and Sharp, D.W. (1971) The cultural context of learning and thinking. London: Methuen.
- Cole, M. and Scribner, S. (1974) Culture and thought: A psychological approach. New York: Wiley.
- Colgan, D.M. (1972) The Maori: Integration or subjugation? In G.M. Vaughan (Ed.) Racial issues in New Zealand. Auckland: Akarana Press, p.19-29.
- Commission on Education (1962) The report of the Commission on Education in New Zealand. Wellington: Government Printer.

- Cooke, V.J. (1933) A comparative study of the attainment of Maori and European children in some aspects of English ability. M.A. thesis, Auckland University College.
- Countryman, E. (1971) Racial intelligence theory examined and rejected. The Christchurch Press, 111, October 16, 19.
- Countryman, E. (1971) Letter to the Editor - 'Racial intelligence'. The Christchurch Press, 111, October 22, 8.
- Cronbach, L.J. (1951) Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.
- Cronbach, L.J. (1960) Essentials of psychological testing. 2nd ed. New York: Harper and Row.
- Cronbach, L.J. (1969) Heredity, environment and educational policy. Harvard Educational Review, 39, 338-347.
- Cronbach, L.J. (1970) Essentials of psychological testing. 3rd ed. New York: Harper and Row.
- Cryns, A.G.J. (1972) African intelligence: A critical survey of cross-cultural intelligence research in Africa south of the Sahara. Journal of Social Psychology, 57, 283-301.
- Darlington, R.B. (1971) Another look at "culture-fairness". Journal of Educational Measurement, 8, 71-82.
- Darwin, C. (1859) The origin of species by natural selection. London: Murray.
- Dasen, P.R. (1972) Cross-cultural Piagetian research. A summary. Journal of Cross-Cultural Psychology, 3, 23-39.
- Department of Education (1924) Terman group tests of mental ability. Wellington: Department of Education.
- Department of Education (1971) Maori children and the teacher. Wellington: Government Printer.
- Department of Education (1973) Education statistics of New Zealand. Part 1, 1973. Wellington: Department of Education.
- Department of Education (n.d.) Register of progress and achievement (Form E 19/16). Wellington: Government Printer.
- Department of Industries and Commerce (1967) The Maori in the New Zealand economy. Wellington: Department of Industries and Commerce.
- Department of Statistics (1974) Population, Part A. New Zealand population and migration, 1972-73 and 1973-74. Wellington: Government Printer.
- Deutsch, M., Fishman, J.A., Kogan, L., North, R. and Whiteman, M. (1964) Guidelines for testing minority group children. Journal of Social Issues, 20, 129-145.

- Dewes, K.M. (1968) The place of Maori language in the education of Maoris. Te Kaunihera Maori, Spring, 51-61.
- Dockrell, W.B. (Ed.) (1970) On intelligence. London: Methuen.
- Doppelt, J.E. and Bennett, G.K. (1967) Testing job applicants from disadvantaged groups. Test Service Bulletin, 57, 1-5.
- Downes, H.E.P. (1948) Some aspects of aptitude and intelligence testing of New Zealanders: Maori and Pakeha. M.A. thesis, Victoria University College.
- Drenth, P.J.D. and van der Flier, H. (1976) Cultural differences and comparability of test scores. International Review of Applied Psychology, 25, 137-144.
- DuBois, P.H. (1970) A history of psychological testing. Boston: Allyn and Bacon.
- du Chateau, P. (1967) Ten-point gap in Maori aptitudes. National Education, 49, 157-158.
- Duncanson, J.P. (1966) Learning and measured abilities. Journal of Educational Psychology, 57, 220-229.
- Ebbinghaus, H. (1896) Ueber eine neue method zur prüfung geistige fähigkeiten und ihre anwendung bei Schulkindem. Z. fur Psychol. und Physiol. der Sinnesorgane, 13, 401-457.
- Edwards, A.L. (1972) Experimental design in psychological research. 4th ed. New York: Holt, Rinehart and Winston.
- Eells, K., Davis, A., Havighurst, R.J., Herrick, V.E. and Tyler, R. (1951) Intelligence and cultural differences. Chicago: University of Chicago Press.
- Elley, W.B. (1969) Changes in mental ability in New Zealand school children. New Zealand Journal of Educational Studies, 4, 140-155.
- Elley, W.B. and Irving, J.C. (1972) A socio-economic index for New Zealand based on levels of education and income from the 1966 census. New Zealand Journal of Educational Studies, 7, 153-167.
- Elley, W.B. and Irving, J.C. (1974) 'A socio-economic index for New Zealand'. A rejoinder to I.R. Brooks and P.F. Cuttance. New Zealand Journal of Educational Studies, 9, 85-87.
- Elley, W.B. and Reid, N.A. (1969) Progressive Achievement Tests. Teacher's manual, Reading Comprehension and Reading Vocabulary. Wellington: N.Z.C.E.R.
- Elley, W.B. and Reid, N.A. (1971) Progressive Achievement Tests: Teacher's manual, Listening Comprehension. Wellington: N.Z.C.E.R.

- Engelhart, M.D. (1965) A comparison of several item discrimination indices. Journal of Educational Measurement, 2, 69-71.
- Estes, W.K. (1974) Learning theory and intelligence. American Psychologist, 29, 740-749.
- Ewing, J.L. and Shallcross, J. (Eds.) (1970) An introduction to Maori education: Selected readings. Wellington: New Zealand Universities Press.
- Eysenck, H.J. (1971) Race, intelligence and education. London: Temple Smith.
- Eysenck, H.J. (1973) The measurement of intelligence. Lancaster: Medical and Technical Publications.
- Fehr, F.S. (1969) Critique of hereditarian accounts of "intelligence" and contrary findings: A reply to Jensen. Harvard Educational Review, 39, 571-580.
- Feldman, M.W. and Lewontin, R.C. (1975) The heritability hang-up. Science, 190, 1163-1168.
- Ferguson, G.A. (1954) On learning and human ability. Canadian Journal of Psychology, 8, 95-112.
- Ferguson, G.A. (1956) On transfer and the abilities of man. Canadian Journal of Psychology, 10, 121-131.
- Ferguson, G.A. (1971) Statistical analysis in psychology and education. 3rd ed. New York: McGraw-Hill.
- Ferron, O. (1965) The test performance of 'coloured' children. Educational Research, 8, 42-57.
- Findley, W.G. (1956) Rationale for evaluation of item discrimination statistics. Educational and Psychological Measurement, 16, 175-180.
- Fitt, A.B. (1952) The Stanford-Binet Scale: Its suitability for New Zealand. Wellington: N.Z.C.E.R.
- Fitzgerald, T.K. (1968) The social position of the Maori university graduate: A sample survey of seventy-five New Zealand Maori graduates. Polynesian Studies, 7, 1-42.
- Fitzgerald, T.K. (1972) Education and identity: A reconsideration of some models of acculturation and identity. New Zealand Journal of Educational Studies, 7, 45-58.
- Flugel, J.C. (1933) A hundred years of psychology. London: Duckworth.
- Forster, J. and Ramsay, P. (1969) Migration, education and occupation: The Maori population 1936-66. In J. Forster (Ed.) Social Process in New Zealand. Auckland: Longman Paul, p.198-232.

- Fowler, H.W. and Fowler, F.G. (Eds.) (1964) The concise Oxford Dictionary of current English. 5th ed., Oxford: Clarendon Press.
- Fredrickson, L.C. (1977) Measured intelligence: Species specific? Perhaps; Race specific? Perhaps not. Journal of Genetic Psychology, 130, 95-104.
- Gadd, B. (1976) Cultural difference in the classroom: Special needs of Maoris in Pakeha schools. Auckland: Heinemann.
- Gallimore, R. and Howard, A. (Eds.) (1968) Studies in a Hawaiian community: Na makomako o Nanakuli, Honolulu: Bishop Museum.
- Gardner, M. (1964) Mathematical games. Scientific American, 210, 122-130.
- Garrett, E.M. (1973) Social relationships. In D.H. Bray and C.G.N. Hill (Eds.) Polynesian and Pakeha in New Zealand Education. Vol. 1. Auckland: Heinemann, p. 29-43.
- Garrett, H.E. (1955) General psychology. New York: American Book Co.
- Garrett, H.E. (1962) The SPSSI and racial differences. American Psychologist, 17, 260-263.
- Garrett, H.E. (1964) I.Q. and school achievement of Negro and White children of comparable age and school status. Mankind Quarterly, 5, 45-49.
- Geddes, W.R. (1961) Maori and Aborigine: A comparison of attitudes and policies. The Australian Journal of Science, 24, 217-225.
- Gellner, E. (1973) The savage and modern mind. In R. Horton and R. Finnegan (Eds.) Modes of thought: Essays on thinking in Western and non-Western societies. London: Faber and Faber, p.162-181.
- George, H.V. (1974) Bernstein's generalization. Education, 23, 20-21.
- Ginsburg, H. (1972) The myth of the deprived child. Poor children's intellect and education. Englewood Cliffs: Prentice-Hall.
- Glass, G.V. and Stanley, J.C. (1970) Statistical methods in education and psychology. Englewood Cliffs: Prentice-Hall.
- Glazer, N. (1975) Ethnicity: A world phenomenon. Dialogue, 8, 34-46.
- Goodenough, F.L. (1926) Measurement of intelligence by drawings. Yonkers: World Book Co.
- Gordon, E. (1974) 'A convenient explanation'. Education, 23, 22-23.
- Graves, N.B. and Graves, T.D. (1974) Inclusive versus exclusive behavior in New Zealand school settings: Polynesian-Pakeha contrasts in learning styles. Auckland: Department of Anthropology, University of Auckland.

- Greenfield, P.M. and Bruner, J.S. (1969) Culture and cognitive growth. In D.A. Goslin (Ed.) Handbook of socialization theory and research. New York: Rand McNally, p.633-657.
- Gregory, K.B. (1974) Twin strands in the suburb and its primary school. In D.H. Bray and C.G.N. Hill (Eds.) Polynesian and Pakeha in New Zealand education. Vol. 2. Auckland: Heinemann, p.99-113.
- Gregson, R.A.M. (1971) Psychologists' comments on racial intelligence theories. The Christchurch Press, 111, October 19, 8.
- Gregson, R.A.M. (1971) Letter to the Editor - 'Racial intelligence' The Christchurch Press, 111, October 22, 8.
- Guilford, J.P. (1954) Psychometric methods. 2nd ed. New York: McGraw-Hill.
- Guilford, J.P. (1956a) The structure of intellect. Psychological Bulletin, 53, 267-293.
- Guilford, J.P. (1956b) Fundamental Statistics in Psychology and Education. New York: McGraw-Hill.
- Guilford, J.P. (1959) Three faces of intellect. American Psychologist, 14, 469-479.
- Guilford, J.P. (1967) The nature of human intelligence. New York: McGraw-Hill.
- Guilford, J.P. (1968) The structure of intelligence. In D.K. Whitla (Ed.) Handbook of measurement and assessment in the behavioral sciences. Reading: Addison-Wesley, p.215-260.
- Guilford, J.P. and Fruchter, B. (1973) Fundamental statistics in psychology and education. 5th ed. New York: McGraw-Hill.
- Guilford, J.P. and Hoepfner, R. (1971) The analysis of intelligence. New York: McGraw-Hill.
- Gustafson, B.S. (1967) A study of non-verbal intelligence and reading comprehension in Maori and non-Maori third form pupils considered in socio-economic status groups. Dip.Ed. Investigation. Massey University.
- Guthrie, G.M. (n.d.) A culture-fair information test of intelligence. Pittsburg: Pennsylvania State University.
- Haddon, A.C. (Ed.) (1901-35) Reports of the Cambridge Anthropological Expedition to Torres Straits. 6 vols. Cambridge: Cambridge University Press.
- Harker, R.K. (1971a) Maori education and research. Australian and New Zealand Journal of Sociology, 7, 46-57.
- Harker, R.K. (1971b) Socio-economic and cultural factors in Maori academic attainment. Journal of the Polynesian Society, 80, 20-41.

- Harker, R.K. (1973) The scholastic attainments of Maori children: A review and synthesis. In D.H. Bray and C.G.N. Hill (Eds.) Polynesian and Pakeha in New Zealand education, Vol.1. Auckland: Heinemann, p.54-62.
- Harker, R.K. (1975) Streaming and social class. In P.D.K. Ramsay (Ed.) The family and the school in New Zealand society. Carlton: Pitman Pacific, p.330-336.
- Harker, R.K. (1976) A survey of the cognitive skills of Maori and Pakeha school children. Report to the Inter-departmental Committee on Polynesian Research, Wellington.
- Harvard Educational Review (1969) Environment, heredity and intelligence. Cambridge: Harvard Educational Review, Reprint Series No. 2.
- Harvard Educational Review (1969) Science, heritability and IQ. Cambridge: Harvard Educational Review, Reprint Series No. 4.
- Havighurst, R.J. (1974) Anthropology and cultural pluralism. Three case studies: Australia, New Zealand and U.S.A. Wellington: N.Z.C.E.R.
- Havighurst, R.J. (1976) The relative importance of social class and ethnicity in human development. Human Development, 19, 56-64.
- Hawkins, P. (1972) Restricted codes and Maori-English. New Zealand Journal of Educational Studies, 7, 59-68.
- Healy, W. and Fernald, G.M. (1911) Tests for practical mental classification. Psychological Monographs, 13, Whole No. 54.
- Hebb, D.O. (1949) The organisation of behavior. New York: Wiley.
- Hector, H. (1960) Results from a simple gestalt continuation test applied to illiterate Black mine workers. Journal of the National Institute of Personnel Research, 7, 156-161.
- Hermansson, G.L. (1974) Counselling Polynesian youth. In D.H. Bray and C.G.N. Hill (Eds.) Polynesian and Pakeha in New Zealand education. Vol.2. Auckland: Heinemann, p.141-147.
- Herrnstein, R.J. and Boring, E.G. (1966) A source book in the history of psychology. Cambridge: Harvard University Press.
- Herskovits, M. (1965) Forward In F. Boas, The mind of primitive man. New York: Free Press (revised edition) p.5-12.
- Herskovits, M. (1973) Cultural relativism. Toronto: Random House.
- Hicks, R.E. (1969) Some comments on test development in emerging countries with special reference to Papua and New Guinea. Papua New Guinea Journal of Education, 6, 29-43.

- Hirsch, J. (1968) Behavior-genetic analysis and the study of man. In M. Mead, T. Dobzhansky, E. Tobach and R. Light (Eds.) Science and the concept of race. New York: Columbia University Press, p.37-48.
- Horn, J.L. and Cattell, R.B. (1966) Refinement and test of the theory of fluid and crystallized general intelligence. Journal of Educational Psychology, 57, 253-270.
- Horn, J.L. and Cattell, R.B. (1967) Age differences in fluid and crystallized intelligence. Acta Psychologica, 26, 107-129.
- Horton, R. (1967) African traditional thought and western science. Africa, 37, 50-71 and 155-187.
- Humphreys, L.G. (1962) The organisation of mental abilities, American Psychologist, 17, 475-483.
- Humphreys, L.G. (1973) Statistical definitions of test validity for minority groups. Journal of Applied Psychology, 58, 1-4.
- Humphreys, L.G. (1975) Addendum (to 'Educational uses of tests with disadvantaged students'). American Psychologist, 30, 95-96.
- Humphreys, L.G. and Dachler, H.P. (1969a) Jensen's theory of intelligence. Journal of Educational Psychology, 60, 419-426.
- Humphreys, L.G. and Dachler, H.P. (1969b) Jensen's theory of intelligence; A rebuttal. Journal of Educational Psychology, 60, 432-433.
- Hunn, J.K. (1961) Report on the Department of Maori Affairs. Wellington: Government Printer.
- Hunt, J.McV. (1961) Intelligence and experience. New York: Ronald Press.
- Hunt, J.McV. (1969) Has compensatory education failed? Has it been attempted? Harvard Educational Review, 39, 278-300.
- Inhelder, B. and Piaget, J. (1958) The growth of logical thinking from childhood to adolescence. New York: Basic Books.
- Irvine, S.H. (1965) Adapting tests to the cultural setting: A comment. Occupational Psychology, 39, 13-23.
- Irvine, S.H. (1966) Towards a rationale for testing attainments and abilities in Africa. British Journal of Educational Psychology, 30, 24-32.
- Irvine, S.H. (1969a) Factor analysis of African abilities and attainments: Constructs across cultures. Psychological Bulletin, 71, 20-32.
- Irvine, S.H. (1969b) Contributions of ability and attainment testing in Africa to a general theory of intellect. Journal of Biosocial Science, Supplement 1, 91-102.

- Irvine, S.H. (1970) Affect and construct: A cross-cultural check on theories of intelligence. Journal of Social Psychology, 80, 23-30.
- Irvine, S.H. (1973) Tests as inadvertent sources of discrimination in personnel decisions. In P. Watson (Ed.) Psychology and Race. Harmondsworth: Penguin, p.453-466.
- Jackson, G.D. (1975) On the report of the ad hoc committee on educational uses of tests with disadvantaged students: Another psychological view from the Association of Black Psychologists. American Psychologist, 30, 88-93.
- Jackson, P.M. (1966) The use of the performance scale of the Wechsler Intelligence Scale for children in a New Zealand urban community. M.A. thesis, University of Auckland.
- Jarvie, I.C. (1975) Cultural relativism again. Philosophy of the Social Sciences, 5, 343-353.
- Jensen, A.R. (1969a) How much can we boost IQ and scholastic achievement? Harvard Educational Review, 39, 1-123.
- Jensen, A.R. (1969b) Jensen's theory of intelligence: A reply. Journal of Educational Psychology, 60, 427-431.
- Jensen, A.R. (1970a) A theory of primary and secondary familial mental retardation. In N.R. Ellis (Ed.) International review of research in mental retardation. Vol. 4, New York: Academic Press, p.33-105.
- Jensen, A.R. (1970b) Another look at culture-fair testing. In J. Hellmuth (Ed.) Disadvantaged child. Vol. 3, New York: Brunner/Mazel, p.53-101.
- Jensen, A.R. (1970c) Can we and should we study race differences? In J. Hellmuth (Ed.) Disadvantaged child. Vol.3, New York: Brunner/Mazel, p.124-157.
- Jensen, A.R. (1970d) Hierarchical theories of mental ability. In W.B. Dockrell (Ed.) On intelligence. London: Methuen, p. 119-190.
- Jensen, A.R. (1972a) Genetics and education. London: Methuen.
- Jensen, A.R. (1972b) Addendum to 'A theory of primary and secondary familial mental retardation'. In A.R. Jensen, Genetics and education. London: Methuen, p.281-293.
- Jensen, A.R. (1973a) Educational differences. London: Methuen.
- Jensen, A.R. (1973b) Educability and group differences. London: Methuen.
- Johnston, R.J. (1974) Urban patterns. In R.J. Johnston (Ed.) Society and environment in New Zealand. Christchurch: Whitcomb and Tombs, p.150-169.
- Journal of Educational Psychology (1921) Intelligence and its measurement: A symposium. Journal of Educational Psychology, 12, 123-147 and 195-216.

- Kagan, J., Moss, A.H. and Sigel, A.H. (1963) Psychological significance of styles of conceptualization. In J.C. Wright and J. Kagan (Eds.) Basic cognitive processes in children. Monograph of the Society for Research in Child Development, Serial No. 86, p.73-112.
- Kaiser, H.F. (1959) Computer program for varimax rotation in factor analysis. Educational and Psychological Measurement, 19, 413-420.
- Kamin, L.J. (1974) The science and politics of IQ. New York: Wiley.
- Kearney, G.E. (1962) P.I.R. IV - A performance test for indigenous peoples. B.Comm. thesis, University of Queensland.
- Kearney, G.E. (1966a) Cognitive capacity among the Orokaiva. In R.G. Crocombe (Ed.) Orokaiva papers. New Guinea Research Bulletin, No. 13, Port Moresby: New Guinea Research Unit, p.1-26.
- Kearney, G.E. (1966b) Some aspects of the general cognitive ability of various groups of Australian Aborigines as assessed by the Queensland Test. Ph.D. thesis, University of Queensland.
- Keddie, N. (Ed.) (1973) Tinker, tailor The myth of cultural deprivation. Harmondsworth: Penguin.
- Keeling, B. (1969) Developing a battery of tests for high school classification: A second progress report. Education Research Newsletter, 3, 7-9.
- Keeling, B. (1971) Developing a battery of tests for high school classification and guidance; Third progress report. Education Research Newsletter, 4, 4-5.
- Keeling, B. and Seddon, R. (1969) A test battery for Form III entrance classification. New Zealand Journal of Educational Studies, 4, 57-68.
- Kenworthy, L.M., Martindale, T.B. and Sadaraka, S.M. (1968) Some aspects of the Hunn Report: A measure of progress. Wellington: School of Political Science and Public Administration, Victoria University of Wellington.
- Kerlinger, F.N. and Pedhazur, E.J. (1973) Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.
- Kidd, A.H. (1962) The culture-free aspects of Cattell's test of g: Culture-free. Journal of Genetic Psychology, 101, 343-362.
- King, M. (Ed.) (1975) Te ao hurihuri: The world moves on. Aspects of Maoritanga. Wellington: Hicks Smith.
- Kipling, R. (1908) Kim. London: Macmillan.
- Kirk, S.A. and McCarthy, J.J. (1961) The Illinois Test of Psycholinguistic Abilities: An approach to differential diagnosis. American Journal of Mental Deficiency, 66, 399-412.

- Klippel, M.D. (1973) A cross-cultural study of two approaches to the measurement of intelligence. M.A. thesis, University of Auckland.
- Klippel, M.D. (1975) Measurement of intelligence among three New Zealand ethnic groups: Product versus process approaches. Journal Cross-Cultural Psychology, 6, 365-376.
- Knox, H.A. (1914) A scale based on the work at Ellis Island for estimating mental deficit. Journal of the American Medical Association, 62, 741-747.
- Köhler, W. (1961) Psychological remarks on some questions of anthropology. In M. Henle (Ed.) Documents of gestalt psychology. Berkeley: University of California Press. p.203-221.
- Kohs, S.C. (1920) The block-design tests. Journal of Experimental Psychology, 3, 357-376.
- Kohs, S.C. (1923) Measuring intelligence: A psychological statistical study based upon the block design tests. New York: Macmillan.
- Krejcie, R.V. and Morgan, D.W. (1970) Determining sample size for research activities. Educational and Psychological Measurement, 30, 607-610.
- Kuder, G.F. and Richardson, M.W. (1937) The theory of the estimation of test reliability. Psychometrika, 2, 151-160.
- Leiter, R.G. (1936) The Leiter International Performance Scale. Honolulu: University of Hawaii Bulletin, No. 15.
- Levine, M. (1976) The academic achievement test: Its historical context and social functions. American Psychologist, 31, 228-238.
- Levi-Strauss, C. (1963) Structural anthropology. New York: Basic Books.
- Levi-Strauss, C. (1966) The savage mind. Chicago: University of Chicago Press.
- Levy-Bruhl, L. (1910) How natives think. New York: Washington Square Press.
- Linn, R.L. and Werts, C.E. (1971) Considerations for studies of test bias. Journal of Educational Measurement, 8, 1-4.
- Loehlin, J.C., Lindzey, G. and Spuhler, J.N. (1975) Race differences in intelligence. San Francisco: Freeman and Co.
- Logan, J.L. (1952) An investigation of the language difficulties of Maori training college students and the significance of these difficulties in intelligence or general ability test scores. Dip.Ed. Investigation, Auckland University College.
- Love, R. (1943) The Maori, yesterday, to-day and tomorrow. New Zealand Journal of Public Administration, 6, 60-73.

- Lovegrove, M.N. (1963) It's hard to be objective: Social studies warning from the changing attitudes to Maoris. National Education, 45, 77-81.
- Lovegrove, M.N. (1964) A cross-cultural study of scholastic achievement and selected determiners. Ph.D. thesis, University of Auckland.
- Lovegrove, M.N. (1966) The scholastic achievement of European and Maori children. New Zealand Journal of Educational Studies, 1, 15-39.
- Lovegrove, M.N. (1972) What history tells us about the abilities of the Maori. Journal of the Australian and New Zealand History of Education Society, 1, 21-32.
- Lunzer, E.A., Dolan, T. and Wilkinson, J.E. (1971) The development of systematic thinking. London: Social Science Research Council, Interim Report.
- MacArthur, R.S. (1968) Some differential abilities of northern Canadian native youth. International Journal of Psychology, 3, 1968, 43-51.
- MacArthur, R.S. (1974) Construct validity of three New Guinea Performance Scale sub-tests: Central Eskimo and Nsenga Africans. In J.L.M. Dawson and W.J. Lonner (Eds.) Readings in Cross-Cultural Psychology. Hong Kong: University of Kong Kong Press, p.51-60.
- McClew, A. (1958) A comparative study of the reading achievement of Maori and European children in an urban school. M.A. thesis, Auckland University College.
- McCreary, J.R. (1966) Reading tests with Maori children. New Zealand Journal of Education Studies, 1, 40-50.
- McDonald, G. (1974a) Pre-school education and Maori communities: A matter of values. In D.H. Bray and C.G.N. Hill (Eds.) Polynesian and Pakeha in New Zealand education. Vol. 2, Auckland: Heinemann, p.58-73.
- McDonald, G. (1974b) The Maori concepts of Tapu and Noa in relation to pre-school groups. In D.H. Bray and C.G.N. Hill (Eds.) Polynesian and Pakeha in New Zealand education. Vol. 2, Auckland: Heinemann, p.74-78.
- McDonald, G. (1975) A comment on some of the concepts and methods used in studies of Maoris and education. New Zealand Journal of Educational Studies, 10, 75-83.
- McDonald, G. (1976) The categories Maori and Pakeha as defined by research workers and by self report. New Zealand Journal of Educational Studies, 11, 37-49.
- McElwain, D.W. (1969) Some aspects of the cognitive ability of Aboriginal children. In S.S. Dunn and C.M. Katz (Eds.) Aborigines and education. Melbourne: Sun Books, p.264-272.

- McElwain, D.W. (n.d.) Field study in the Gilbert and Ellice Islands.
Brisbane: Department of Psychology, University of Queensland.
- McElwain, D.W. and Griffiths, D.M. (1957) Report on the possibility of using psychological procedures as an aid to recruiting in the Pacific Islands Regiment.
Brisbane: Report for Headquarters Northern Command, Australian Military Forces.
- McElwain, D.W. and Kearney, G.E. (1970) Queensland Test Handbook.
Hawthorn: A.C.E.R.
- McElwain, D.W. and Kearney, G.E. (1973) Intellectual development. In G.E. Kearney, P.R. de Lacey and G.R. Davidson (Eds.) The psychology of Australian Aborigines. Sydney: Wiley, p.43-56.
- McKenzie, M. (1970) Maori education 1960-1969: A bibliography.
Wellington: N.Z.C.E.R.
- McKenzie, N.R. (1930) School surveys 1927 and 1929. New Plymouth, North and South Taranaki Branches of the New Zealand Educational Institute.
- McKenzie, N.R. (1931) The educability of the Maori. In P.M. Jackson (Ed.) Maori and education. New Plymouth: Avery, p.195-219.
- McNemar, Q. (1964) Lost: Our intelligence? Why? American Psychologist, 19, 871-882.
- McNemar, Q. (1975) On so-called test bias. American Psychologist, 30, 848-851.
- Mann, C.W. (1935) Fiji Test of General Ability: Manual of instructions.
Suva: Government Printer.
- Maori Affairs Amendment Act (1974) Maori Affairs Amendment Act, No. 73. In The Statutes of New Zealand 1974, Vol.2. Public Acts Nos. 26-128. Wellington: Government Printer.
- Maori Synod of the Presbyterian Church of New Zealand (1961) A Maori view of the "Hunn Report". Christchurch: Presbyterian Bookroom.
- Marquart, D.I. and Bailey, L.L. (1955) An evaluation of the Culture-Free Tests of Intelligence. Journal of Genetic Psychology, 86, 353-358.
- Marsden, E. (1924) Terman intelligence tests in New Zealand. In 1925 New Zealand Official Yearbook, Wellington: Government Printer, p.823-834.
- Martin, A.W. (1966) Word fluency: A comparative study of word fluency in European and Maori high school pupils. M.A. thesis, Victoria University of Wellington.
- Mehrens, W.A. and Lehmann, I.J. (1973) Measurement and evaluation in education and psychology. New York: Holt, Rinehart and Winston.

- Metge, J. (1958) The Rakau studies: A review article. (A). An anthropological assessment. Journal of the Polynesian Society, 67, 352-371.
- Metge, J. (1964) A new Maori migration. London: Athlone Press.
- Metge, J. (1976) The Maoris of New Zealand: Ruatahi. London: Routledge and Kegan Paul.
- Miles, T.R. (1957) Contributions to intelligence testing and the theory of intelligence. British Journal of Educational Psychology, 27, 153-165.
- Ministry of Defence (1960) An analysis of the abilities, attainments and occupations of eighteen-year old males in New Zealand. Wellington: Ministry of Defence.
- Mitchell, D.R. (1968) Cultural deprivation: A review of the literature and a comparison of deprived and non-deprived five-year old children. Dip.Ed. Investigation. University of Canterbury.
- Mitchell, D.R. (1970) Education and the disadvantaged child. In R.J. Bates (Ed.) Prospects in New Zealand education. Auckland: Hodder and Stoughton, p.181-198.
- Montagu, M.F.A. (Ed.) (1969) The concept of race. New York: Collier.
- Morin, J. (1962) G52, test d' intelligence pour Nord-Africains illetrés. Travail Humain, 25, 353-364.
- Myrdal, G. (1957) The rich lands and the poor. New York: Harper.
- Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. and Bent, D.H. (1975) Statistical package for the social sciences. 2nd ed. New York: McGraw-Hill.
- N.Z.C.E.R. (1969) Otis Tests of Mental Ability: Manual of directions. Wellington: N.Z.C.E.R.
- N.Z.C.E.R. (n.d.) Manual of directions for using the Metropolitan Readiness Tests. Wellington: N.Z.C.E.R.
- New Zealand Educational Institute (1967) Report and recommendations on Maori education. Wellington: New Zealand Educational Institute.
- New Zealand Official Yearbook - 1971. Wellington: Government Printer.
- New Zealand Post-Primary Teachers' Association (1970) Maori education. Wellington: New Zealand Post-Primary Teachers' Association.
- Odell, B.F. (1944) A preliminary survey of the new Revised Stanford-Binet Scale in New Zealand. M.A. thesis, Auckland University College.
- Oliver, R.A.C. (1933) The adaptation of intelligence tests to tropical Africa. Overseas Education, 4, 186-191; 5, 8-13.

- Oppenheim, R.S. (1972) (Review of) Maori children and the teacher. Journal of the Polynesian Society, 81, 134-135.
- Ord, I.G. (1959) Development of a test of cognitive capacity for indigenes of Papua New Guinea. M.A. thesis, University of Queensland.
- Ord, I.G. (1967a) The New Guinea Performance Scale and its educational uses. Papua New Guinea Journal of Education, 5, 7-16.
- Ord, I.G. (1967b) Manual for use with the New Guinea Performance Scale. Port Moresby: Department of the Public Service Commissioner.
- Ord, I.G. (1968a) Manual for the Pacific Reasoning Series Test. Hawthorn: A.C.E.R.
- Ord, I.G. (1968b) The P.I.R. Test and derivatives. Australian Psychologist, 2, 137-146.
- Ord, I.G. (1968c) Manual for Pacific Design Construction Test. Hawthorn: A.C.E.R.
- Ord, I.G. (1970) Mental tests for pre-literates. London: Ginn
- Ord, I.G. (1972a) Assessing cognitive capacities of non-literate New Guinea adults and children. In L.J. Cronbach and P. Drenth (Eds.) Mental tests and cultural adaptation. The Hague: Mouton, p. 183-195.
- Ord, I.G. (1972b) Testing for educational and occupational selection in developing countries. Occupational Psychology, 43, Monograph Issue (Whole No. 3).
- Ord, I.G. and Schofield, J. (1970) Manual - Pacific Infants Performance Scale. Milton: Jacaranda.
- Ortar, G. (1963a) Is a verbal test cross-cultural? Scripta Hierosolymitana, 13, 219-235.
- Ortar, G. (1963b) The transfer of psychological diagnostic measures from one culture to another. Acta Psychologica, 21, 218-230.
- Peterson, J. (1926) Early conceptions and tests of intelligence. New York: World Book Co.
- Piaget, J. (1950) The psychology of intelligence. London: Routledge and Kegan Paul.
- Piaget, J. (1972) Intellectual evolution from adolescence to adulthood. Human Development, 15, 1-12.
- Piddington, R. (1968) Emergent development and integration. In E. Schwimmer (Ed.) The Maori people in the nineteen-sixties. Auckland: Blackwood and Janet Paul, p.257-269.
- Pintner, R. and Paterson, D.G. (1917) A scale of performance tests. New York: Appleton.

- Pitt, D. and Macpherson, C. (1974) Emerging pluralism: The Samoan community in New Zealand. Auckland: Longman Paul.
- Polack, J.S. (1840) The manners and customs of the New Zealanders. 2 Vols. London: Madden.
- Pool, I. (1963) When is a Maori a "Maori"? Journal of the Polynesian Society, 72, 206-210.
- Poortinga, Y.H. (1971) Cross-cultural comparison of maximum performance tests: Some methodological aspects and some experiments with simple auditory and visual stimuli. Psychologia Africana, Monograph Supplement, No. 6 (Whole No.).
- Popham, W.J. (1967) Educational statistics. New York: Harper and Row.
- Popper, K.P. (1966) The open society and its enemies. Vol.2, 5th ed.rev. London: Routledge and Kegan Paul.
- Preston, R., St. George, A. and St. George, R. (1974a) Pacific Reasoning Series Test. Test Information Report No. 1. Port Moresby: Psychological Services Branch.
- Preston, R., St. George, A. and St. George, R. (1974b) Pacific Design Construction Test. Test Information Report No. 3. Port Moresby: Psychological Services Branch.
- Price-Williams, D.R. (1975) Explorations in cross-cultural psychology. San Francisco: Chandler and Sharp.
- Pride, J.B. (1974) Maori children and Bernstein: A linguistic appraisal. Education, 23, 17-20.
- Ramsay, P.D.K. (1975) The New Zealand Maori and the school. In P.D.K. Ramsay (Ed.) The family and the school in New Zealand society. Carlton: Pitman Pacific, p.288-301.
- Rangihau, J. (1975) Being Maori. In M. King (Ed.) Te ao hurihuri: The world moves on: Aspects of Maoritanga. Wellington: Hicks Smith, p.221-233.
- Raven, J.C. (1938) Guide to using Progressive Matrices. London: Lewis and Co.
- Redmond, M. and Davies, F.R.J. (1940) The standardization of two intelligence tests. Wellington: N.Z.C.E.R.
- Reid, N.A. and Hughes, D.C. (1974) Progressive Achievement Tests, Teacher's manual, Mathematics. Wellington: N.Z.C.E.R.
- Renwick, W.L. (1963) Another Victorian view of the Maori. National Education, 45, 365-368.
- Richardson, J.F. (1970) An analysis of some data from an application of the Queensland Test in Papua New Guinea. Psychology Research Unit, Report No. 45, Port Moresby: Australian Military Forces.

- Richardson, K. and Spears, D. (Eds.) (1972) Race, culture and intelligence. Harmondsworth: Penguin.
- Riessman, F. (1962) The culturally deprived child. New York: Harper and Row.
- Riske, M. (1950) The Maoris' IQ. National Education, 32, 284-285.
- Ritchie, J. (1958) Childhood in Rakau. Publications in Psychology, No. 10. Wellington: Victoria University of Wellington.
- Ritchie, J.E. (1957a) Some observations on Maori and Pakeha intelligence test performance. Journal of the Polynesian Society, 66, 351-356.
- Ritchie, J.E. (1957b) Maori background data schedule. Department of Psychology, Working Paper No. 1. Wellington: Victoria University of Wellington.
- Ritchie, J.E. (1958) Basic Personality in Rakau. Publications in Psychology, No. 8. Wellington: Victoria University of Wellington.
- Ritchie, J.E. (1963) The making of a Maori. Wellington: Reed.
- Ritchie, J.E. (1968) Workers. In E. Schwimmer (Ed.) The Maori people in the nineteen-sixties. Auckland: Blackwood and Janet Paul, p.288-310.
- Ritchie, J.E. (n.d.) Educational testing: Pilot project 1963 at Panguru High School, Hokianga. Report to the Maori Education Foundation. Wellington: Victoria University of Wellington.
- Ritchie, J.E. and Beaglehole, E. (1958) The Rakau studies. Journal of the Polynesian Society, 67, 132-154.
- Ritchie, J.E. and Ritchie, J. (1968) Children. In E. Schwimmer (Ed.) The Maori people in the nineteen-sixties. Auckland: Blackwood and Janet Paul, p.311-327.
- Rogers, C. (1956) Measuring intelligence in New Zealand. Auckland University College Monograph Series No. 2, Auckland: Pilgram Press.
- Rogers, D.G. (1964) A standardization for New Zealand use of the Raven Coloured Progressive Matrices Test (Sets A, Ab, B, Revised Order, 1956) M.A. thesis, University of Canterbury.
- Rutishauser, H. (1957) Der quotienten - differenzen - algorithmus. Basel/Stuttgart: Birkhaeuser.
- St. George, A. (1972) The Pacific Infants Performance Scale: Some preliminary and comparative New Zealand studies. B.Phil. thesis, University of Waikato.
- St. George, A. (1974) The Pacific Infants Performance Scale: Some preliminary and comparative New Zealand studies. Psychology Research Paper No.1, Hamilton: University of Waikato.

- St. George, R. and Preston, R. (1976) Psychological testing in Papua New Guinea; A critical appraisal. Unpublished paper, Massey University.
- St. George, R. and St. George, A. (1974) The measurement of cognitive abilities in the Cook Islands: A research report. Report to the Interdepartmental Committee on Polynesian Research, Wellington.
- St. George, R. and St. George, A. (1975) The intellectual assessment of Maori and European school children. In P.D.K. Ramsay (Ed.) The family and the school in New Zealand society. Carlton: Pitman Pacific, p.302-321.
- Salmond, A. (1974) 'Maori-English' and the 'Restricted Code'. Education, 23, 21-22.
- Samuda, R.J. (1975) Psychological testing of American minorities: Issues and consequences. New York: Dodd, Mead.
- Sattler, J.M. (1965) Analysis of functions of the 1960 Stanford-Binet Intelligence Scale, Form L-M. Journal of Clinical Psychology, 21, 173-179.
- Scheffe, H. (1953) A method for judging all contrasts in the analysis of variance. Biometrika, 40, 87-104.
- Schmidt, F.L. and Hunter, J.E. (1974) Racial and ethnic bias in psychological tests. American Psychologist, 29, 1-8.
- Schroth, M.L. (1976) The use of IQ as a measure of learning rate with minority children. Journal of Genetic Psychology, 128, 101-108.
- Schwarz, P.A. (1963) Adapting tests to the cultural setting. Educational and Psychological Measurement, 13, 673-686.
- Schwarz, P.A. and Krug, R.E. (1972) Ability testing in developing countries. A handbook of principles and techniques. New York: Praeger.
- Schwimmer, E. (Ed.) (1968) The Maori people in the nineteen-sixties. Auckland: Blackwood and Janet Paul.
- Seddon, R. and Keeling, B. (1968) Developing a battery of tests for high school classification. Education Research Newsletter, 1, 1-3.
- Semeonoff, B. and Trist, E. (1958) Diagnostic performance tests. London: Tavistock.
- Serpell, R. (1972) Intelligence tests in the third world. Race Today, 4, 46-49.
- Serpell, R. (1974) Estimates of intelligence in a rural community in Eastern Zambia. Human Development Research Unit Report No. 25, Lusaka: University of Zambia.

- Shockley, W. (1972) Dysgenic, geneticity, raceology: A challenge to the intellectual responsibility of educators. Phi Delta Kappan, 53, 297-307.
- Shortland, E. (1851) The southern districts of New Zealand: A journal with passing notices of the customs of the aborigines. London: Longman, Brown, Green and Longmans.
- Shouksmith, G.A. (1960) A manual for the Canterbury Reasoning Test. Christchurch: Caxton.
- Shouksmith, G.A. (n.d.) New Zealand norms for the D.A.T. Battery. Unpublished paper. University of Canterbury.
- Shuey, A.M. (1966) The testing of Negro intelligence. 2nd ed. New York: Social Science Press.
- Siegel, S. (1956) Non-parametric statistics for the behavioral sciences. New York: McGraw-Hill.
- Silvey, J. (1963) Aptitude testing and educational selection in Africa. Rhodes-Livingstone Institute Journal, 34, 9-22.
- Simpson, G.M. (1955) A study of the intelligence and scholastic progress of Maori pupils in the secondary department of a district high school. Dip.Ed. Investigation, Auckland University College.
- Sinclair, K. (1971) Why are race-relations in New Zealand better than in South Africa, South Australia or South Dakota? New Zealand Journal of History, 5, 121-127.
- Smith, K.K. (1966) A validation study of the Queensland Test. B.A. (Hons.) thesis, University of Queensland.
- Smith, L.M. (1956) The educational assessments of Maori children. M.A. thesis, Auckland University College.
- Smith, M.W. (1974) Alfred Binet's remarkable questions: A cross-national and cross-temporal analysis of the cultural biases built into the Stanford-Binet Intelligence Scale and other Binet tests. Genetic Psychology Monographs, 89, 307-334.
- Sörbom, D. (1974) A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 24, 229-239.
- Spearman, C.E. (1904) "General Intelligence", objectively determined and measured. American Journal of Psychology, 15, 201-293.
- Spearman, C.E. (1927) The abilities of man. London: Macmillan.
- Spearman, C.E. and Wynn-Jones, L.L. (1950) Human ability. London: Macmillan.

- Stanley, J.C. and Hopkins, K.D. (1972) Educational and psychological measurement and evaluation. Englewood-Cliffs: Prentice-Hall.
- Stern, W. (1911) Die differentielle psychologie in ihren methodischen grundlagen. Leipzig: Barth.
- Stinchcombe, A.L. (1969) Environment: The cumulation of events. Harvard Educational Review, 39, 511-522.
- Stodolsky, S.S. and Lesser, G. (1967) Learning patterns in the disadvantaged. Harvard Educational Review, 37, 546-593.
- Sutherland, I.L.G. (1929) The study of Maori mind. Journal of the Polynesian Society, 38, 127-147.
- Sutherland, I.L.G. (1931) Maori mentality. In P.M. Jackson (Ed.) Maori and education. New Plymouth: Avery, p.73-87.
- Sutherland, I.L.G. (1940) The Maori people to-day. Wellington: N.Z.C.E.R.
- Taylor, C.R.H. (1965) A Pacific bibliography. Printed matter relating to the native peoples of Polynesia, Melanesia and Micronesia. Oxford: Clarendon Press.
- Taylor, C.R.H. (1972) A bibliography of publications on the New Zealand Maori and the Moriori of the Chatham Islands. Oxford: Clarendon Press.
- Terman, L.M. (1916) The measurement of intelligence. Boston: Houghton Mifflin.
- Terman, L.M. (1921) Intelligence and its measurement. Journal of Educational Psychology, 12, 127-133.
- Terman, L.M. and Merrill, M.A. (1937) Measuring intelligence: A guide to the administration of the new Revised Stanford-Binet Tests of Intelligence. London: Harrap.
- Thomas, D.R. (1975) Cooperation and competition among Polynesian and European children. Child Development, 46, 948-953.
- Thomson, A.S. (1859) The story of New Zealand. Past and present. Savage and civilised. Vol.1. London: Murray.
- Thomson, G. (1950) The factor analysis of human ability. London: University of London Press.
- Thomson, R. (1968) The Pelican history of psychology. Harmondsworth: Penguin.
- Thorndike, E.L. (1927) The measurement of intelligence. Columbia: Bureau of Publications, Teachers College, Columbia University.
- Thorndike, R.L. (1963) The concepts of over- and under-achievement. Columbia: Bureau of Publications, Teachers College, Columbia University.
- Thorndike, R.L. (1971) Concepts of culture-fairness. Journal of Educational Measurement, 8, 63-70.

- Thornton, J. (1892) A present day view of the Maori race. Church Missionary Intelligencer, 43, 201-206.
- Thurstone, L.L. (1938) Primary mental abilities. Psychometric Monographs, No.1, (Whole No.).
- Travers, W.T.L. (1872) Stirring times of Te Rauparaha. Christchurch: Whitcombe and Tombs.
- Triandis, H.C. (1976) Methodological problems of comparative research. International Journal of Psychology, 11, 155-159.
- Tuck, B.F., Hanson, A.L. and Zimmerman, M. (1975) The WISC-R: A New Zealand study of norms and validity. New Zealand Journal of Educational Studies, 10, 52-58.
- Vandenberg, S.G. (1962) The hereditary abilities study: Hereditary components in a psychological test battery. American Journal of Human Genetics, 14, 220-237.
- Valentine, C.A. (1971) Deficit, difference and bicultural models of Afro-American behavior. Harvard Educational Review, 41, 137-157.
- Vaughan, G.M. (1964a) The development of ethnic attitudes in New Zealand school children. Genetic Psychology Monographs, 70, 135-173.
- Vaughan, G.M. (1964b) Ethnic awareness and attitudes in New Zealand. Publications in Psychology, No.17, Wellington: Victoria University of Wellington.
- Vaughan, G.M. (Ed.) (1972) Racial issues in New Zealand. Auckland: Akarana Press.
- Vernon, P.E. (1955) The assessment of abilities. Studies in Education, 7, 189-215.
- Vernon, P.E. (1960) Intelligence and attainment tests. London: University of London Press.
- Vernon, P.E. (1961) The structure of human abilities. 2nd ed. London: Methuen.
- Vernon, P.E. (1962) Intellectual development in non-technological societies. In G. Nielsen (Ed.) Child and education. Proceedings of the 14th International Congress of Applied Psychology. Vol.3, Copenhagen: Murksgaard, p.94-105.
- Vernon, P.E. (1969) Intelligence and cultural environment. London: Methuen.
- Voyat, G. (1970) IQ: God given or man-made? A discussion of "Jensenism". In J. Hellmuth (Ed.) Disadvantaged child. Vol.3. New York: Bruner/Mazel, p.158-162.
- Walker, R.J. (1973) Biculturalism and education. In D.H. Bray and C.G.N. Hill (Eds.) Polynesian and Pakeha in New Zealand education. Vol.1. Auckland: Heinemann, p.110-122.

- Walsh, A.C. (1973) Developments since the Hunn Report and their bearing on education. In D.H. Bray and C.G.N. Hill (Eds.) Polynesian and Pakeha in New Zealand education, Vol.1. Auckland: Heinemann, p.18-28.
- Walters, R.H. (1953) Wechsler-Bellevue Test results of prison inmates. Australian Journal of Psychology, 5, 46-54.
- Walters, R.H. (1956) A comparison of the primary mental abilities of Maori children with those of children of European descent. New Zealand Bulletin of Psychology, 2, 75-86.
- Walters, R.H. (1958) The intelligence test performance of Maori children. A cross-cultural study. Journal of Abnormal and Social Psychology, 57, 107-114.
- Warburton, F.W. (1951) The ability of the Gurkha recruit. British Journal of Psychology, 42, 114-133.
- Watson, J.E. (1967) Horizons of unknown power: Some issues of Maori schooling. Wellington: N.Z.C.E.R.
- Watson, J.E. (1968) Accommodating the Polynesian heritage of the Maori child: A New Zealand problem. Wellington: N.Z.C.E.R.
- Wechsler, D. (1949) The Wechsler Intelligence Scale for Children - Manual. New York: Psychological Corporation.
- Wechsler, D. (1955) Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corporation.
- Wechsler, D. (1974) Manual for the Wechsler Intelligence Scale for Children - Revised. New York: Psychological Corporation.
- Werner, H. (1948) Comparative psychology of mental development. New York: Science Editions.
- Werner, H. and Kaplan, B. (1956) The developmental approach to cognition: Its relevance to the psychological interpretation of anthropological and ethnolinguistic data. American Anthropologist, 58, 866-880.
- Wesman, A.G. (1968) Intelligent testing. American Psychologist, 27, 267-274.
- Wilkinson, J.H. (1959) The evaluation of the zeros of ill-conditioned polynomials. Numerische Mathematik, 1, 150-180.
- Williams, J.S. (1960) Maori achievement motivation. Publications in Psychology, No. 13, Wellington: Victoria University of Wellington.
- Winer, B.J. (1962) Statistical principles in experimental design. New York: McGraw-Hill.
- Winiata, M. (1944) Preliminary investigation to the Ranginui revision of the Fiji Test of General Ability. M.A. thesis, Auckland University College.

- Wissler, C. (1901) The correlation of mental and physical traits. Psychological Review, Monograph Supplement, Vol. 3, No.6 (Whole No.).
- Witkin, H.A. (1967) Cognitive styles across cultures. International Journal of Psychology, 2, 233-250.
- Witkin, H.A. and Berry, J.W. (1975) Psychological differentiation in cross-cultural perspective. Journal of Cross-Cultural Psychology, 6, 4-87.
- Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R. and Karp, S.A. (1962) Psychological differentiation. New York: Wiley.
- Wober, M. (1968) Notes on administering psychological tests in Africa. Bulletin of the British Psychological Society, 20, 25-34.
- Wober, M. (1974) Towards an understanding of the Kiganda concept of intelligence. In J.W. Berry and P.R. Dasen (Eds.) Culture and cognition: Readings in cross-cultural psychology. London: Methuen, p.261-280.
- Woodworth, R.S. (1910) Race differences in mental traits. Science, 31, 171-186.
- Zimmerman, M.L. (1971) An assessment of the cognitive ability and attitude to rehabilitation of European and Maori trainees at a borstal institution. B.Phil. thesis, University of Waikato.