



**PREDICTION OF LATER INTELLIGENCE BEYOND 2 YEARS OF
AGE FROM PREFERENCE FOR NOVELTY DURING INFANCY**

Reza Tasbihsazan Mashhadi

B.A. (Hons.) Tehran University (Iran)

M.A. Tehran University (Iran)

Thesis submitted in fulfilment of the requirements
for the Degree of Doctor of Philosophy

Department of Psychology

University of Adelaide

July 1997

TABLE OF CONTENTS

Title	Page
LIST OF FIGURES	ix
LIST OF TABLES	x
ABSTRACT	xiv
STATEMENT	xix
ACKNOWLEDGMENTS	xx
CHAPTER 1	
INTRODUCTION	1
1-1 BACKGROUND INFORMATION	1
1-2 PURPOSE OF THE RESEARCH	5
1-3 SAMPLES AND PROCEDURE	6
1-3-1 Samples	6
1-3-2 Measures and procedure	7
1-4 OVERVIEW OF THESIS	8
CHAPTER 2	
INFANT DEVELOPMENTAL SCALES AND PREDICTION OF LATER IQ	11
2-1 INTRODUCTION	11
2-2 THE CONSTRUCT OF INTELLIGENCE DURING INFANCY	12
2-2-1 Infant competencies	12
2-2-2 Theories of infant intelligence and cognitive development	13
(i) <i>Piaget's theory of cognitive development</i>	13
(ii) <i>A single general factor during infancy</i>	15
(iii) <i>Fagan's theory of infant intelligence</i>	16
2-3 INFANT DEVELOPMENTAL SCALES	18
(i) <i>The Mental Growth of the Preschool Child</i>	18

(ii) <i>Cattell Infant Intelligence Scale</i>	19
(iii) <i>Griffiths Mental Developmental Scales</i>	19
(iv) <i>Bayley Scales of Infant Development</i>	20
(v) <i>Neonatal Behavioural Assessment</i>	21
2-4 PREDICTION OF CHILDHOOD IQ FROM EARLY INFANCY	22
2-5 PREDICTABILITY OF INFANT DEVELOPMENTAL SCALES TO LATER IQ IN CHILDHOOD	24
2-5-1 Explanation of the low predictive validity of developmental scales	28

CHAPTER 3

INFANTS INFORMATION PROCESSING TASKS	30
3-1 INTRODUCTION	30
3-2 MEMORY IN THE INFANT	31
3-3 HABITUATION	33
3-3-1 Construct and procedure	33
3-3-2 Dependent measures of the habituation procedure	37
3-3-3 Test-retest reliability and long-term stability of habituation measures	38
3-3-4 Concurrent and longitudinal predictive validity of habituation	40
(i) <i>Concurrent validity</i>	40
(ii) <i>Predictive validity</i>	41
3-3-5 Criticism of habituation (theoretical concept and practical issues)	42
3-4 VISUAL RECOGNITION MEMORY	45
3-4-1 Theory and procedure	45
3-4-2 Psychometric properties of visual recognition memory tasks	47
3-4-3 Concurrent validity for measures of visual recognition memory	48
3-4-4 Prediction of later IQ from visual recognition memory tasks	49
<i>Studies from the Rose Laboratory</i>	50
3-4-5 Review of meta-analyses of the predictive validity of visual recognition memory	52
3-4-6 Critique of the literature and meta-analyses of the predictive validity of novelty tasks	53
3-5 UNDERLYING MECHANISMS	57

3-6 THE FAGAN TEST OF INFANT INTELLIGENCE	60
3-6-1 Standardization and statistical information on the	62
3-7 DEVELOPMENT OF RESEARCH PROGRAM	63
3-7-1 Research questions	64
(i) <i>Study 1</i>	64
(ii) <i>Study 2</i>	65
(iii) <i>Study 3 and Study 4</i>	66
(iv) <i>The predictive validity of the Bayley Scales</i>	67
(v) <i>Rising MDI in young children</i>	67
3-7-2 Pilot study	68
3-8 AIMS OF THE CURRENT RESEARCH PROJECT	69

CHAPTER 4

THE RELIABILITY AND STABILITY OF INFANT VISUAL ATTENTION MEASURES FROM 6 TO 12 MONTHS OF AGE	
4-1 INTRODUCTION	71
4-1-1 Internal consistency and test-retest reliability of the FTII	73
4-1-2 The reliability of other attentional measures	76
4-2 METHOD	78
4-2-1 Subjects	78
(i) <i>Birth status</i>	79
(ii) <i>Socioeconomic status</i>	80
4-2-2 Measures	81
(i) <i>The Fagan Test of Infant Intelligence</i>	81
(ii) <i>Bayley Scales of Infant Development</i>	82
4-2-3 Apparatus	82
4-2-4 Procedure	83
4-3 RESULTS	88
4-3-1 Descriptive statistic and preliminary analysis of the measures of interest	88
(i) <i>Novelty preference scores</i>	88
(ii) <i>Number of fixations during familiarization</i>	93

(iii) <i>Exposure time</i>	94
(iv) <i>Developmental measures (Bayley Scales)</i>	95
4-3-2 Internal consistency of the FTII	96
(i) <i>27 Week Test</i>	96
(ii) <i>29 Week Test</i>	96
(iii) <i>39 Week Test</i>	97
(iv) <i>52 Week Test</i>	97
4-3-3 The test-retest reliability of the FTII	98
(i) <i>Test-retest reliability of novelty scores</i>	98
(ii) <i>Test-retest reliability of number fixations and exposure time</i>	98
4-3-4 Intercorrelations between attentional measures	101
4-3-5 Correlation between information processing measures and demographic characteristics	102
4-3-6 Analyses of novelty trials data	103
4-4 DISCUSSION	105
(i) <i>Internal consistency of the FTII</i>	107
(ii) <i>Test-retest reliability</i>	109
(iii) <i>The relationship between attentional measures</i>	111

CHAPTER 5

THE PREDICTIVE VALIDATION OF THE FAGAN TEST OF INFANT INTELLIGENCE	113
5-1 INTRODUCTION	113
5-1-1 Concurrent validity of the Fagan test	114
5-1-2 The predictive validity of the FTII	116
5-2 METHOD	119
5-2-1 Subjects	119
5-2-2 Measures at follow-up testing	120
(i) <i>Bayley Scales of Infant Development (Revised Version)</i>	120
(ii) <i>The stanford-Binet Intelligence Scale</i>	120
5-2-3 Procedure	121
5-2-4 Attrition rate	122

5-3 RESULTS	123
5-3-1 Preliminary analysis	123
(i) <i>Childhood measures</i>	123
5-3-2 Relationship between information processing measures and the Bayley Scales	124
5-3-3 Predictive validity of the FTII	125
(i) <i>The relation between infancy novelty scores and later MDI</i>	125
(ii) <i>Bivariate correlations between novelty scores and IQ</i>	128
5-3-4 The sensitivity and specificity of the Fagan test	128
5-3-5 Predictive validity of exposure time and number of fixations to later mental outcomes	130
5-4 DISCUSSION	132
5-4-1 Explanations for low predictability of the FTII	134
5-4-2 Critique of the clinical utility of the FTII	136

CHAPTER 6
(PART I)

PREFERENCE FOR NOVELTY AT 16 MONTHS AS A PREDICTOR OF LATER MDI AND IQ IN EARLY CHILDHOOD	139
6-1 INTRODUCTION	139
6-1-1 Predictive validity of novelty task beyond 12 months of age	141
6-1-2 Time-fixed procedure for measuring novelty preference scores	143
6-2 METHOD	144
6-2-1 Subjects	144
(i) <i>Birth status</i>	145
(ii) <i>Socioeconomic status</i>	145
6-2-2 Measures during infancy	146
(i) <i>Visual recognition memory task</i>	146
6-2-3 Apparatus and procedure	150
6-2-4 Follow-up measures	152
6-3 RESULTS	153
6-3-1 Infancy measures	153

(i) <i>Novelty scores</i>	153
(ii) <i>Number of fixations and exposure time</i>	155
6-3-2 Childhood measures	158
6-3-3 Intercorrelations between infant information processing measures	159
6-3-4 Concurrent validity of information processing measures with the mental development scale	160
6-3-5 Predictive validity of the 16-month novelty score to the 27-month mental development scale	161
6-3-6 Prediction of later IQ from novelty tasks in infancy	161
6-3-7 Prediction of later MDI and IQ from other measures obtained from the information processing task	163
6-4 DISCUSSION	164
6-4-1 Explanations for low predictive validity of novelty task	166

(PART II)

THE EFFECT OF EXPIRING FAMILIARISATION TIME IN 16 MONTHS ON THE PREDICTION OF LATER IQ AT 27 MONTHS OF AGE	169
6-5 INTRODUCTION	169
6-6 METHOD	171
6-6-1 Subjects	171
6-6-2 Measures and stimuli during the infancy test and follow-up testing	172
6-6-3 Apparatus	173
6-6-4 Procedure	174
6-7 RESULTS	175
6-7-1 Preliminary analyses	175
(i) <i>Infancy data</i>	175
(ii) <i>Childhood cognitive performance</i>	177
6-7-2 Prediction of later MDI and IQ from the visual recognition memory task in infancy	178
6-8 DISCUSSION	178

CHAPTER 7

PREDICTION AT 16 MONTHS OF AGE OF COGNITIVE DEVELOPMENT AT 27 MONTHS OF AGE FROM INFANT SENSORIMOTOR SCALES AND SOCIOECONOMIC FACTORS	182
7-1 INTRODUCTION	182
7-1-1 The predictive power of developmental scales	183
7-1-2 Predicting later IQ from socioenvironmental factors	185
7-2 METHOD	186
7-2-1 Subjects	186
7-2-2 Measures and procedure	187
(i) <i>The Standard Progressive Matrices</i>	187
7-3 RESULTS	189
7-3-1 Preliminary analysis	189
(i) <i>Infancy measures</i>	189
(ii) <i>Childhood measures</i>	191
7-3-2 Intercorrelations between the MDI and SES	192
7-3-3 Reliability of mental and psychomotor indices	194
7-3-4 Bivariate correlations between infancy MDI and later MDI at childhood	195
7-3-5 Correlation between MDI at infancy and later IQ	195
7-3-6 Multivariate analyses	198
7-3-7 Correlational analyses of socioenvironmental indices and later mental outcomes	199
7-4 DISCUSSION	200

CHAPTER 8

INCREASING MENTAL DEVELOPMENT INDEX IN AUSTRALIAN CHILDREN	206
8-1 INTRODUCTION	206
8-1-1 Reliability of the Bayley Scales of Infant Development	206
8-1-2 MDI gains	208
8-2 METHOD	209

8-2-1 Subjects	209
8-2-2 Measures and procedure	211
8-3 RESULTS	211
8-3-1 IQ gains	211
8-3-2 Comparison between two mental scales (MDI)	213
8-3-3 Reliability and stability of the MDI	214
8-3-4 Comparison of the Bayley Scales with the Stanford-Binet Test	214
8-4 DISCUSSION	215

CHAPTER 9

GENERAL DISCUSSION AND CONCLUSION	220
9-1 THEORETICAL BACKGROUND	220
9-2 SUMMARY OF STUDIES AND RESULTS	222
9-2-1 Fagan Test Of Infant Intelligence (Study 1 & Study 2)	222
9-2-2 New Visual Recognition Memory Task (Study 3 & Study 4)	224
9-2-3 Bayley Scales of Infant Development (two complementary examinations of data)	226
9-3 DEVELOPMENTAL TRENDS IN COGNITIVE SKILLS AND FIXATION MEASURES	226
9-3-1 Intercorrelation between fixation measures	229
9-4 RELIABILITY OF VISUAL RECOGNITION MEMORY TASKS	230
9-5 THE RELATIONSHIP BETWEEN NOVELTY SCORES AND LATER MDI/IQ	233
9-6 THE PREDICTABILITY OF THE BAYLEY SCALES	238
9-7 CONTINUITY AND DISCONTINUITY ISSUE	240
9-8 LIMITATIONS OF THE RESEARCH PROGRAM	242
9-9 CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH	243
9-10 A FINAL COMMENT	246
Appendices	248
References	273

LIST OF FIGURES

Title	Page
Figure 1-1 The Longitudinal Design for Study 1, Study 2, Study 3 and Study 4 and the Measures Used for Each Age Group	9
Figure 4-1 Portable Viewing Stage Used in the FTII	83
Figure 4-2 Representation of Trials of the FTII for Ten Novelty Items	85
Figure 4-3 Percentage of Preference for Novelty Across All 10 Novelty Problems	91
Figure 4-4 Boxplot of Mean Novelty Preference Scores for Each Age Group	92
Figure 4-5 Normal Probability Plot of Mean Novelty Scores for Total Sample	92
Figure 4-6 Simple Bar Chart of Number of Fixations by Age Group	95
Figure 4-7 Scatterplot of Novelty Scores by Age Group	100
Figure 4-8 Comparison of Time Spent Looking at Novel Stimuli and Remaining Time	104
Figure 6-1 Representation of Novelty Items of the visual recognition memory Task	148
Figure 6-2a Bar Graph of Percentage of Preference for Novelty	156
Figure 6-2b Preference for Novelty in Normal and Intellectually Disabled Subjects	156
Figure 6-3 Developmental Change of Exposure Time and Number of Fixations	157
Figure 6-4 The Portable Stimulus Presentation Stage	173
Figure 7-1 Some Test Material from the Bayley Scale of Infant DEVELOPMENT	188
Figure 9-1 Intercorrelations Between Infants Measures and later MDI/IQ	223

LIST OF TABLES

Title	Page
Table 3-1 The Number of Samples in Four Meta Analyses Reviewed Involving Visual Recognition Memory Tasks	55
Table 4-1 Birth Status and Demographic Variables of Sample	80
Table 4-2 The Percentage of Preference for Novelty for Ten Novelty Items and Means of Novelty Score in Each Age Group	89
Table 4-3 Means and Standard Deviations of Exposure Time and the Number of Fixations for Each Age Group	94
Table 4-4 Means and Standard Deviations of MDI and PDI	95
Table 4-5 Internal Consistency (Cronbach's α) for Different Measures Obtained from the FTII	96
Table 4-6 Test-retest Stability of Attentional Measures	99
Table 4-7 Intercorrelation Between Attentional Measures	102
Table 4-8 Correlations Between Attentional Measures and Demographic Variables	102
Table 4-9 Means and Standard Deviations for Number and Time of Fixations During Novelty Trials	104
Table 5-1 Attrition Rate for Sample	122
Table 5-2 Means and Standard Deviations of Infancy and Childhood Measures	124
Table 5-3 Correlations Between MDI and PDI and Measures Obtained from Information Processing Task	125
Table 5-4 Bivariate Correlations Between the FTII and Later Bayley Scales at 18 to 24 Months and the (Number of Subjects) in Each Analysis	126
Table 5-5 Correlations Between the Mean of Novelty Scores at Two Ages and Later Bayley Scales at 18, 21 and 24 Months	127
Table 5-6 Correlations Between Novelty Preference Scores and Later Stanford-Binet Subtests at 24 Months of Age and the (Number of Subjects) in Each Correlation	129

Table 5-7	Contingency Table of the Success of the FTII in Predicting Normality and Deficits in Intelligence (Using Cutoff Point \leq 53% for the FTII and \leq 70 for the MDI of the Bayley Scale at 24 Months of Age)	131
Table 5-8	Contingency Table of the Success of the FTII in Predicting Normality and Deficits in Intelligence (Using Cutoff Point \leq 54.5% for the FTII and \leq 85 for the MDI of the Bayley Scale at 24 Months of Age)	131
Table 5-9	Correlations Between Exposure Time and Later MDI and General IQ	132
Table 6-1	Birth Status and Demographic Characteristics of the Sample in Study 3	145
Table 6-2	Means, Standard Deviations and <i>t</i> Values of Percentage Novelty Scores for all Ten Novel Items	154
Table 6-3	Internal Consistency Coefficients (Alpha) for Attentional Measures at 16 Months	154
Table 6-4	Means and Standard Deviations for Information Processing Measures and Developmental Indices	158
Table 6-5	Means and Standard Deviations of the Bayley Scales and the Stanford-Binet Test at 27 Month of Age	159
Table 6-6	Intercorrelations Between Attentional Measures at 16 Months	160
Table 6-7	Intercorrelations Between Attentional Measures and Measures Obtained From the Developmental Scale	160
Table 6-8	Longitudinal Correlation Between Novelty Preference at 16 Months and MDI at 27 Months of Age	162
Table 6-9	Correlations Between Novelty Scores and 27 Months Stanford-Binet Test	162
Table 6-10	Correlations Between Exposure Time at 16 Months and Stanford-Binet at 27 Months	163
Table 6-11	Birth Status and Demographic Characteristic of the Sample	172
Table 6-12	Means and Standard Deviations of Novelty Scores for all Novel Items, with <i>t</i> Values for the Significance of Exceeding Chance	176

Table 6-13	Means and Standard Deviations for all Infancy and Childhood Measures	177
Table 6-14	Longitudinal Correlations Between Novelty Scores and Later Childhood Tests	178
Table 7-1	Means and Standard Deviations of the Bayley Scales During Infancy	190
Table 7-2	Means and Standard Deviations for Bayley Scales and Stanford-Binet Test	192
Table 7-3	Means and <i>t</i> Values for MDI Obtained at Infancy and Childhood	193
Table 7-4	Correlations Between MDI at Infancy and Parental Demographic Characteristics	193
Table 7-5	Test-retest Reliability of MDI and PDI During Infancy	194
Table 7-6	Bivariate Correlations Between Earlier Bayley Mental and Motor Scales (MDI and PDI) and Subsequent MDI in Childhood	196
Table 7-7	Correlations Between MDI at Infancy and Stanford-Binet Test at Childhood and the Number of Subjects in Each Correlation	197
Table 7-8	Multiple regression predicting later MDI and Stanford-Binet at two years	199
Table 7-9	Bivariate Correlations Between Birth Status and Parental Socioenvironmental Factors and Later Cognitive Outcome	200
Table 7-10	Correlations Between Children's Stanford-Binet Scores at 24 and 27 Months and Their Mothers' Standard Progressive Matrices Scores	201
Table 8-1	Demographic Characteristics and Birth Status of the Sample	210
Table 8-2	Summary Table of Sample, Mean Scores and Standard Deviations on Each Measure	212
Table 8-3	Comparison Between Two Versions of Bayley MDI (IQ Gains) for Total Sample and According to Gender and Ethnic Background	213

Table 8-4	Reliability Coefficients (a), Standard Error of Measurement (SEM) and Test-retest Stability of the Mental Scale of Four Age Groups	215
Table 8-5	Correlation Between MDI and Stanford-Binet Scores at 24 and 27 Months of Age	215

ABSTRACT

PREDICTION OF LATER INTELLIGENCE BEYOND 2 YEARS OF AGE FROM PREFERENCE FOR NOVELTY DURING INFANCY

The nature of infant intelligence, its measurement and the prediction of future intellectual functioning from early infancy measures have long been controversial issues, attracting considerable research. At least up until the early 1980s, empirical evidence convinced researchers that the nature of intelligence in infancy is qualitatively different from that in childhood and adulthood. The theory of discontinuity holds that either a conception of intelligence in terms of a *g* factor is not applicable to infancy or that the nature of intelligence changes from early infancy to the preschool period and later childhood (Lewis, Jaskir & Enright, 1986; McCall, Hogarty & Hurlburt, 1972). This theory therefore predicts no cross-age correlation among diverse cognitive behaviours from infancy to childhood and beyond.

Fagan (1984) and Sternberg (Sternberg, 1981; Berg & Sternberg, 1985) challenged the assumption that the development of intelligence was discontinuous and proposed that some aspects of intelligence - for example, an infant's predilection to prefer a novel object as opposed to a familiar object - are continuous throughout life. This proposition led researchers to reconceptualize the development of intelligence during the early years of life and to extend the information processing model to include the development of intellectual competence in infants. The information processing model assumes that some aspects of mental functioning are basic components of cognitive development. From a theoretical point of view, novelty preference in infancy is held to indicate the efficiency of brain

structures responsible for information processing. Preference for novelty should therefore predict later individual differences in intelligence.

Four studies were undertaken to test this hypothesis. Studies 1 and 2 were designed to test exclusively the reliability and predictive validity of the Fagan Test Of Infant Intelligence (FTII; Fagan & Shepherd, 1991) . Studies 3 and 4 investigated the possibility of extending the Fagan procedure to older infants using different types of familiarization time and scoring systems suggested by the author.

In terms of classical test theory, Study 1 found that the commercially available FTII is not a reliable test and it does not measure a single dimension as claimed. There was insufficient homogeneity between novelty items. For most ages, the model of test-retest reliability was also violated. Item analysis of the test confirmed that some of the novelty items are not discriminable, and several more items are not related to the total scale, suggesting that not all items measure the single construct "preference for novelty".

The low test-retest reliability observed in Study 1 placed in doubt the predictive validity of the task. Study 2 confirmed the poor predictive validity of the FTII. Only the FTII version administered at 52-weeks was, to a limited extent, predictive of the Mental Development Index (MDI) from the Bayley Scales and some subtests from the Stanford-Binet (4th. edition) at 24 months of age. These results therefore supported those studies that have found low predictability from a single administration of the FTII (e.g., Anderson, 1996; DiLalla et al., 1990). It is suggested that little reliance should be placed on single, or even double administrations of the FTII, as predictive measures of later IQ, at least in normal infants. In addition, there was obvious error when classifying infants in terms of normality. However, it is recognized that the small sample size and homogeneity of the sample recruited preclude generalization of the findings.

Study 3 and Study 4 were designed to investigate the relationship between novelty scores and later IQ in older infants. No significant correlation between the novelty task and later MDI/IQ was found. Instead, Study 3 clearly showed that the predictability of novelty scores to later IQ is strongly influenced by outlier subjects. When data from just two Down's Syndrome infants were included, the test predicted later outcome well; but when these infants were excluded, prediction was reduced to chance.

Study 3 also revealed that procedural exposure time and the number of fixations on the target figures which infants make both undergo a marked qualitative change in older infants (i.e., 16-month-old infants). Specifically, exposure time becomes shorter and the number of fixations become fewer, demonstrating that older infants fixate on the targets more steadily than younger babies, possibly due to maturation of the brain.

This finding resulted in a working hypothesis that infants at 16 months of age may be more self-regulated and better able to monitor their attention to a task, than younger infants. This suggested that it may not be necessary for the older infants to accumulate Fagan's measure of familiarization time. As recommended for the FTII procedure, familiarization time is essentially an "on-off" measurement, whereby the time of separate fixations is accumulated across a longer period of time. Instead, Study 4 tested whether an expiring familiarization time, starting from the initial fixation, would result in preference for novelty. It was further considered that if the expiring familiarization time did not result in novelty preference, it might produce familiarity preference. This possibility was based on evidence that, the type of preference (familiarity or novelty) depends on the stage at which a stimulus is being processed. That is, novelty preference is a substitute for familiarity preference when the processing becomes more complete (Richards, 1997; Wagner & Sakovits, 1986). However, the results did not support the usefulness of expiring

familiarization time. Novelty scores for most novelty items did not exceed chance level, thereby indicating insufficient familiarization time.

Despite the failure of Study 4 to confirm that expiring familiarization time results in preference for novelty, other changes might be made to improve the novelty task. For example, if the infant is tested by measuring the minimum familiarization time required for familiarity preference, instead of the time necessary for novelty preference, it is possible (in theoretical terms) to predict later intelligence from familiarity preference. Familiarity preference could then be used for older infants, where the establishment of novelty preference is difficult because of higher levels of activity and reduced cooperation for testing.

Two supplementary investigations using data collected in the previous four studies examined the predictive validity of a conventional infancy test, the Bayley Scales Of Infant Development (Bayley, 1969), providing Australian data on the reliability and validity of the Revised Version of the Bayley Scales (Bayley, 1993) and investigating the possibility of MDI gains in an Australian infant population. The analysis of data obtained from the Bayley Scales provided sufficient evidence to conclude that MDI scores obtained at 9 months, and thereafter, predict later mental outcome within the first 27 months of age. The number of significant correlations at 21 and 24 months of age from the 9-month MDI was fewer than significant predictions from 12 and 16 months of age, when most correlations were highly significant. However, in comparison with the novelty score procedure for the FTII, the Bayley Scales have considerably more predictive power within the first 27-month span of life.

Finally, the results of the second supplementary investigation, which compared MDIs obtained from the first version of the Bayley Scales (Bayley, 1969) with those

obtained from the second version (Bayley, 1993), established that Australian infants have gained 18 points on the MDI during the last 25 years. As a consequence therefore, the original Bayley Scales released in 1969 is no longer a reliable test of infant performance, at least for children aged 18 to 27 months and, possibly, for younger infants.

The failure of the FTII to predict later IQ, and the superior predictive value of the Bayley Scales during the first years of life are discussed.

Statement

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and to my best knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text.

I consent to this thesis being made available for photocopying and loan if accepted for the award of the degrees.

Reza Tasbihsazan Mashhadi

July, 1997

ACKNOWLEDGMENTS

I wish to express my appreciation to my principal supervisor Dr. Ted Nettelbeck and my co-supervisor Dr. Neil Kirby for their sound advice, encouragement and patience.

I would also like to thank several people in the Department of Psychology, University of South Australia for their support and motivation. In particular I take this opportunity to thank Dr. Chris Cooper for his financial support, and Ms. Jill Barlow for proof reading and earlier draft of this dissertation. I would also like to thank Carmen Rayner, Bob Wilson, Mark Brown, Allan Baldock, Theo Sudomlak, Stever Tupper and Geoff Matthews for their individual help in preparing equipment and the programmes for the tests used in the project. My gratitude also goes to the parents who volunteered their children for the project. In addition, I am indebted to the University of Urmia for its financial support. Without it this work would not have been possible.

Finally, I would like to thank and dedicate this work to my wife, Haideh, and to my children Samira, Saeed and Sama who helped me in many ways.



CHAPTER 1

INTRODUCTION

1-1 BACKGROUND INFORMATION

Spearman advanced his first theory of intelligence in England in 1904 and a year later the first psychometric intelligence test was published in France by Binet and his collaborator Simon. Nine decades after the work of these two innovative authors, both the concept of intelligence and the measuring intelligence are still controversial topics in psychology (Sternberg, 1995; Weinberg, 1989). It would appear that debate surrounding the concept of intelligence is a long way from being resolved (Myers, 1986).

The nature of infant intelligence also remains a controversial issue (Thompson, Plomin, & DeFries, 1985). Until the beginning of the 20th century, infants were commonly regarded as helpless and incompetent creatures with blank minds who were passively responsive to the environment. This followed views consistent with the empiricism associated with John Locke (1632-1704). However, by the mid 20th century, numerous studies that focussed on neonates and infant developmental status had found that infants were capable of a variety of sensory and motor activities and cognitive functions (Bayley, 1969; Graham, Matarazzo, & Caldwell, 1956; Griffiths, 1954). Given the competency of mental functioning and well-documented memory abilities (Cohen, 1979; Olson & Sherman, 1983), it appears that infants are sufficiently qualified to be called intelligent creatures. However, it is held by some researchers that terms like "intelligence" or "g factor" are inappropriately applied to infant mental abilities (Lewis & McGurk, 1972; McCall, Hogarty, & Hurlburt, 1972). McCall and his colleagues, for example, have

questioned the concept of a *g* factor and suggested that "... the term mental as applied to infant behaviour or tests should be abandoned in favor of some conceptually more neutral label, perhaps Piaget's "sensorimotor", "perceptual-motor..." (McCall et al., 1972, p. 746). Similarly, Escalona and Moriarty (1961) maintained that intelligence, as a group of functions that is measured by intelligence tests in older children, is not evident before the age of 18 months. They concluded that at the preverbal stages of development the term "intelligence test" as applied to preschool children is misleading.

The extent to which early development of intellectual capacities is relatively determined by inherited factors or by experience has been widely debated. Piagetian theory and the information processing approach exemplify different positions in this argument. Piaget maintained that the development of cognition is a process consisting of several different stages. That is, intellectual development has to pass through an invariant sequence of profound qualitative transitions within four different periods of age (Piaget, 1954). Conversely, the information processing perspective has emphasised the importance of the continuity of individual differences in mental activities. Recently, Anderson (1992) has made an attempt to accommodate the Piagetian theory of discontinuity within a theory which attributes stability of individual differences across age to different mechanisms to those responsible for cognitive development. He argues that the Piagetian discontinuity model is roughly consistent with the "maturation of new modules" which become available at different ages, thereby creating new "stages" in development. Discontinuity can be explained by the acquisition of new modules which are unrelated to individual differences in intelligence but probably dependent on chronological age. Anderson maintains that since modules are unconstrained by the speed of basic processing which provides the basic essential for the emergence of intelligence, developmental changes in modules which are

universal will be unrelated to individual differences in intelligence. Intelligence for information processing theorists might be measured by individual differences in some fundamental capacities, like the speed of a basic processing mechanism (Anderson, 1992; Eysenck, 1986; Miller & Vernon, 1992; Vernon, 1989, 1983). Intelligence is continuous and thus individual differences in mental abilities during childhood can, in principle, be predicted from early infancy (Fagan, 1992).

Since the 1920s and 1930s when the first infant developmental scales were published (Bayley, 1933; Gesell, 1925), the prediction of later intelligence and the early identification of non-normal infants has been the focus of attention for many researchers. A great deal of research has attempted to find an empirical relationship between an infant's performance on developmental scales and long-term intellectual outcomes in childhood, and to provide an adequate conceptual explanation for this.

A number of reviews based on longitudinal studies during the last five decades have acknowledged the low predictability of later IQ in childhood from developmental scales administered during infancy, at least for normal infants (Bayley, 1970; Fagan & Singer, 1983; McCall et al., 1972; McCall, Eichorn, & Hogarty, 1977; but see Honzik [1983] and Kopp & McCall [1982] for predictive validity of conventional infant tests for low scoring subjects). McCall (1982), for example, has concluded that, "... from a practical standpoint, infant tests administered before 18 months of age are useless in predicting later childhood IQ for normal samples." (p. 178). However, in the light of empirical evidence the contention that sensorimotor tests are useless in terms of predictability has been challenged (Humphreys & Davey, 1988; Siegel, 1989). It is therefore possible that this conclusion is premature, particularly during the first years of life.

The invention of a simple new technique by Fantz (1958a) to measure looking behaviour in infancy, and introducing the information processing approach as a model of human memory (Shiffrin & Atkinson, 1969), resulted in new ways of investigating the cognitive development of infants. Fantz found that, from birth, infants are responsive to environmental stimulation, look longer at some stimuli than at others and prefer a new stimulus to a previously seen stimulus (Fantz, 1964). The infant's differential looking behaviour implies that babies are able, firstly, to discriminate between stimuli, and, secondly, to prefer some stimuli as opposed to others. Such procedures may represent a significant breakthrough in infant assessment. From a theoretical point of view, preference for novelty is held to indicate the efficiency of brain structures responsible for information processing in infants. Thus, it should reflect, to some extent, subsequent intellectual development.

The last two decades have seen increasing research interest in the possible operationalisation of attention, particularly visual attention, as a means for measuring an infant's cognitive integrity (Lewis, 1982). Using a variety of nonverbal measures, investigators have designed numerous experiments that have provided the field of infant assessment with evidence concerning perceptual/cognitive competencies during early infancy. Growing evidence supports the existence of a relationship between measures of looking behaviour in the first year of life and intelligence test scores in childhood (e.g., Bornstein, 1989a, b; Fagan, Singer, Montie, & Shepherd, 1986; Rose, Feldman, & Wallace, 1992; see also reviews by Bornstein & Sigman, 1986; McCall & Carriger, 1993). However, some researchers believe that the observed relationships are statistically spurious and the product of including outliers (Kagan, 1989), or due to small, atypical samples (Lecuyer 1989). Little is known about the rate at which fundamental attentional behaviours improve

and the stage of development at which preference for novelty could be of value for future programming.

1-2 PURPOSE OF THE RESEARCH

Several investigators have supported the notion that some aspects of an infant's information processing ability, for example response to novelty, are relatively stable throughout the early developmental period (Sternberg, 1985; Sternberg & Okagaki, 1989), and may therefore provide a moderate predictor of later intelligence (Fagan, 1992). However, the complexity of the structure of intelligence during infancy on the one hand, and some mixed and contradictory findings on the other, make it difficult at this stage to accept these conclusions with confidence.

Researchers who doubt the reliable predictability of later intelligence on the basis of such early infant attentional measures have raised a number of questions concerning them. The psychometric properties of new measures, including low reliability, low predictive validity, inadequate statistical documentation, the size and characteristics of samples, the error rates that inevitably occur with these measures, and the theory underlying habituation and recognition memory during infancy are some, but not all, of the critical points raised.

The program of research presented here is mainly concerned with the possibility of predicting mental outcomes during the first 27 months of life. A major aim has been to test whether measures of attention taken during the first 6½ to 16 months of age predict developmental and intelligence scores at later ages. Using the Bayley Scales of Infant Development (Bayley, 1969), the Fagan Test of Infant Intelligence (Fagan & Shepherd, 1991) and a newly developed visual recognition memory task, four studies were undertaken to address the research question. Despite the extensive research that has been carried out on

the construction of the Fagan test (Fagan & Detterman, 1992) the test has not been subject to a critical review (Benasich & Bejar, 1992). Literature in the field reveals that the Fagan test has neither been appraised nor administered in Australia. Therefore, the present research project was undertaken to verify the reliability and predictive validity of the Fagan test, and to compare the results with data collected simultaneously by conventional developmental scales. The upward extension of such cognitive measures, if possible, was the next objective of the project.

1-3 SAMPLES AND PROCEDURE

1-3-1 Samples

The total sample for Study 1 was 85 (41 female, 44 male) normal, middle-class, home-reared infants whose ages ranged between 27 and 52 weeks (postnatal age corrected for prematurity). The same sample was utilized in Study 2, in which the predictability of infancy measures was assessed. The participants for Study 3 were 16 infants (9 female, 7 male) with an average age of 69 weeks. Two of these subjects had Down's Syndrome and the remainder were healthy infants. A total of 16 subjects (7 female, 9 male) with an average age of 69 weeks participated in Study 4. Two additional investigations were also conducted using data obtained from all four samples. The comprehensive demographic characteristics for all samples will be given in relation to each study.

Sample selection for all studies was based on a volunteer procedure and participants may not therefore reflect a representative sample of the Australian infant population. All subjects resided in Adelaide, South Australia.

1-3-2 Measures and procedure

Two different type of tests were administered to infants. One was a traditional developmental scale - the Bayley Scales of Infant Development (BSID; Bayley, 1969). The infant's scores on the Bayley Scales were a Mental Development Index (MDI) and a Psychomotor Development Index (PDI). The second test was the first commercially available test of infant visual recognition memory - the Fagan Test of Infant Intelligence (FTII; Fagan & Shepherd, 1991) and/or a new Visual Recognition Memory (VRM) task for older infants (i.e., 16 months). This novelty task was developed by the author and was based on the Fagan procedure. The Fagan test (and/or new novelty task) provided a score which defined the infant's preference for novelty - the percentage of visual fixations directed towards a novel stimulus.

All infants were followed up one year after the first testing. In addition, infants whose follow-up tests were administered at 18 and 21 months of age were seen again at 24 months, when the final follow-up was scheduled. In the first follow-up session, the infants were tested with two versions of the Bayley Scales, each with different standardisation samples (Bayley, 1969, 1993). In the second follow-up session both versions of the Bayley Scales, as well as the Stanford-Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986, Fourth Edition; Australian adaptation; see also de Lemos, 1989) were administered. Each battery of tests, either during infancy or at childhood, was practiced on nonstudy samples before being used on infants involved in the experimental samples.

In order to explore the importance of birth status factors and environmental influences in predicting later IQ, demographic characteristics were obtained from hospital records and parents. In addition, an intelligence test (Standard Progressive Matrices, Australian edition [de Lemos, 1989]) was given to the mothers of all participating infants, to

assess if a mother's intelligence test score could be used as a proxy for her infant's potential IQ.

The longitudinal design of the project is illustrated in Figure 1-1. This figure also represents the hypothetical relationships within infancy measures and between infancy measures and the childhood MDIs at 18, 21, 24 and 27 months and IQ at 24 and 27 months of age. The expected significant correlations are shown in solid lines¹.

1-4 OVERVIEW OF THESIS

The research on which this thesis is based consisted of four studies and two complementary investigations of the data as described below. Details and design of the longitudinal studies and the results from each study will be reported in later chapters. Before reporting the studies a comprehensive and critical review of the relevant literature currently available will be presented in Chapters 2 and 3.

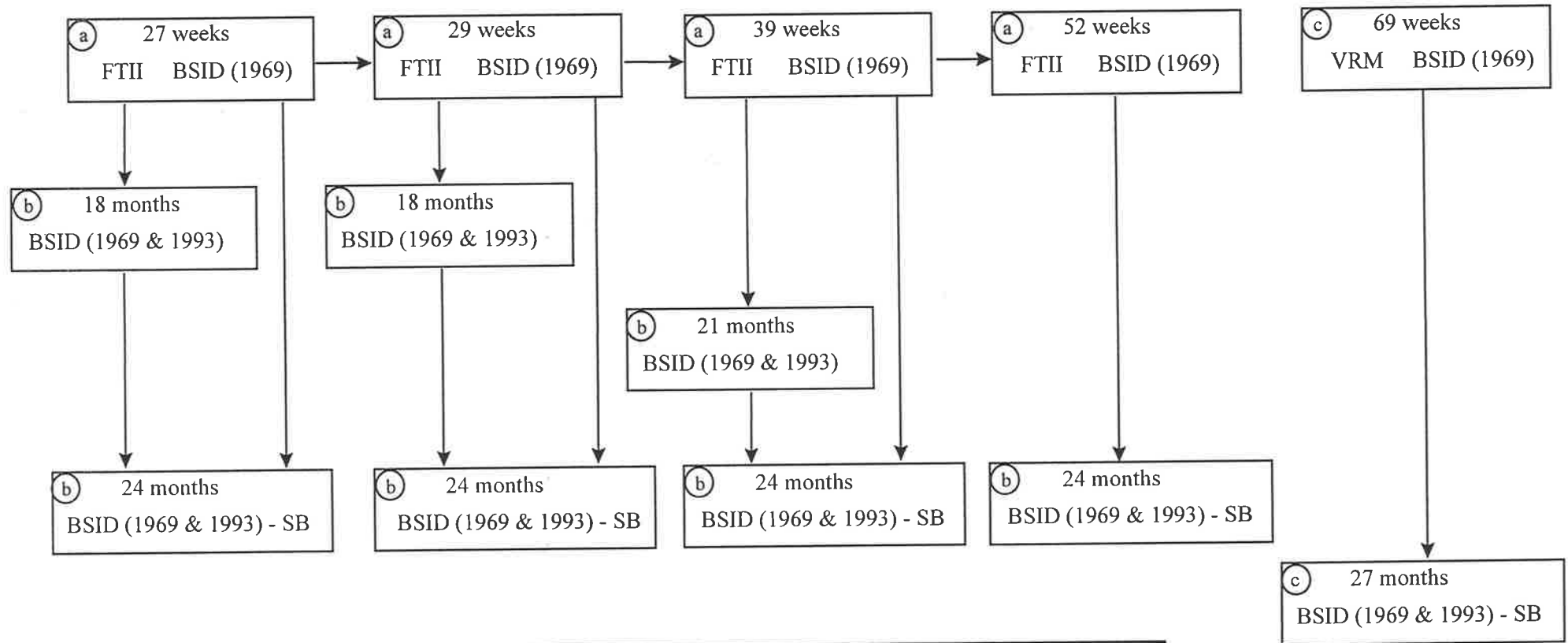
Chapter 4 is concerned with psychometric properties and item analysis of the Fagan test (Study 1). The internal consistency of the Fagan test was assessed using Cronbach's alpha. The longitudinal analysis also made it possible to estimate the test-retest reliability and stability of the FTII at two weeks, two months and three months intervals.

Chapter 5 reports the issue of predictive validity of the Fagan test (Study 2). Having used the FTII as a predictor at four different ages (27, 29, 39, 52 weeks of age), and the Bayley Scales and the Stanford Binet test as criteria, Study 2 followed up the same subjects in Study 1 and attempted to establish the first age in infancy on which the prediction could be based. This chapter also places the concurrent validity of novelty preference scores in relation to developmental indices.

¹ A record of all summary data collected for all subjects is included as Appendix A.

Figure 1-1

The Longitudinal Design for Study 1, Study 2, Study 3 and Study 4 and the Measures Used for Each Age Group



FTII is the Fagan Test of Infant Intelligence (Fagan & Shepherd, 1991).

BSID is the Bayley Scales of Infant Development (Edition 1; Bayley, 1969 & Edition 2; Bayley, 1993).

SB is the Stanford-Binet Intelligence Scale, Fourth Edition (Thorndike, Hagen & Sattler, 1986).

VRM is a new Visual Recognition Memory task, based on the Fagan test, and developed as a part of this thesis.

The cells designated (a) constitute Study 1, (b) constitute Study 2, cells labeled (c) indicate Study 3 and Study 4 using two different familiarization times.

Chapter 6 of this dissertation tests the time limit for the prediction of later IQ, by attempting to extend the usefulness of the visual recognition memory task beyond 12 months (as is used by the Fagan test) to 16 months of age. The efficiency of two different types of familiarization time within the Fagan procedure have also been assessed. Infants in Study 3 were assigned a predetermined time to accumulate information on which to make a preference response. Working on the assumption that older infants are more self-regulated than younger infants, the second group of infants operated with a preset familiarization time which began to expire with the first fixation (Study 4).

Chapter 7 examines the importance of the predictability of the Bayley Scales and socio-environmental factors in early childhood. It investigates if a combined score of an information processing task and a developmental scale during the first 16 months of infancy can predict later IQ more effectively than either test alone.

Using data from both versions of the Bayley Scales (Bayley, 1969, 1993) at 18, 21, 24 and 27 months of age, one aim of the project was to evaluate the reliability and validity of the new Bayley Scales (Bayley, 1993) in an Australian infant population, compared with the results of an American normative sample. In the light of recent evidence concerning the IQ gains in adult literature, it was of interest to investigate the possibility of an MDI gain for the infant population during the last 25 years. The results and discussion concerning the reliability of the new version of the Bayley scale and of an MDI gain will be introduced in Chapter 8.

In the last chapter, a general discussion combines the results from the different studies and presents conclusions concerning the productivity of each type of infancy test when predicting later mental outcome.

CHAPTER 2

INFANT DEVELOPMENTAL SCALES AND PREDICTION OF LATER IQ

2-1 INTRODUCTION

Unlike the long history of philosophical views on cognitive development during infancy, scientific research in this area has a relatively brief history. Beginning in the early twentieth century, the nature of infant intelligence, its measurement and the possibility of predicting later intelligence, were the main concerns for developmentalists. Towards the end of the century debate about these issues is still alive and the main question about the predictability of intelligence is still unanswered. This practical question is, of course, associated with the unresolved theoretical questions of whether the nature of intellectual behaviour is durable throughout the life span or undergoes a qualitative change. If intelligence is a durable construct, the prediction of later IQ from early infancy tests may be possible.

The need for prediction has its roots not only in human curiosity but also in the need for clinical utility. The predictability of later intelligence is important to reduce the anxiety of parents whose babies show some delay in early infancy, particularly parents of first-born babies. In the early months of life, the prediction of later cognitive functioning may be critical for their development since, except for extreme cases or infants who are born with a congenital disorder, other infants who are at risk may not be identified in the first year of life (Hunt, 1981; Jacobson, Fein, Jacobson, Schwartz, & Dowler, 1985). Advanced technology has increased the chance for life for many infants who are born with

complications (Millar, Weir, & Supramaniam, 1991), including infants with an extremely low birth weight (e.g., one-third that of normal infants, Charlton, 1996). These infants need to be monitored during their first two years of life to detect any risk factor. The practical advantage of reliable predictors is that they could be used as a screening device for the identification of intellectual disability¹ and for the implementation and evaluation of intervention programs, or even for adoption placement.

Some recent research that takes advantage of measures believed to be tests of cognitive abilities has found shared variance between tests measuring an infant's response to novelty and tests measuring mental activities later in childhood. However, the possibility that such correlations are low, and that some socioeconomic variables or the parent's IQ can predict the infant's later IQ to the same extent, or even better (McCall, 1982; Wohlwill, 1980), places the immediate practical utility of such processing measures in question, at least in the short-term.

2-2 THE CONSTRUCT OF INTELLIGENCE DURING INFANCY

2-2-1 Infant competencies

Using currently available technology for the assessment of very young infants (for example, looking behaviour), it has been demonstrated that even new-born infants have more capabilities than was once believed. They are endowed with a complex neural system that enables them to process information and perform a variety of mental activities (Cohen, 1979; Fagan, 1992; Nelson, Henschel, & Collins, 1993; Weiss & Zelazo, 1991; Zelazo, 1982). Even newborns are able to remember things (e.g., an auditory stimulation) that they have experienced during the last weeks before birth (DeCasper & Spence, 1986). Normal

¹ Intellectual disability is the recommended terminology for mental retardation in Australia and it means the same as mental retardation in the United States.

newborns are able to discriminate a new word from an old one after 24 hours (Swain, Zelazo, & Clifton, 1993) at 2 weeks of age or younger. They can imitate adult facial gestures such as tongue protrusion, mouth opening and lip protrusion at less than 3 weeks of age (Meltzoff & Moore, 1983, 1977) and happy, surprised facial expressions (Field, Woodson, Greenberg & Cohen, 1982) when they are younger than one week. At 3 months, infants are capable of visual discrimination and recognition of surprise from the happy and sometimes from the sad (Young-Browne, Rosenfeld, & Horowitz, 1977). Infants discriminate the picture of their mother from a stranger at 3 months (Barrera & Maurer, 1981), and at 18 weeks of age can discriminate between color photographs of relatively unfamiliar people (Cohen, DeLoache, & Pearl, 1977) or perceive a similarity between a live person and that person's photograph (Dirks & Gibson, 1977). At 5 to 9 weeks of age, infants show recognition memory after 24 hours for colour and for form (Bushnell, McCutcheon, Sinclair, & Tweedlie, 1984) and at 8 months demonstrate robust haptic memory (Catherwood, 1993). Cumulatively, the evidence indicates that even newborns are active information processors. As Flavell has pointed out, the infant is a "constructive thinker and learner" who is equipped with cognitive structures and processing strategies (Flavell, 1992, p. 998).

2-2-2 Theories of infant intelligence and cognitive development

(i) *Piaget's theory of cognitive development*: The Piagetian theory is the most comprehensive theory in relation to cognitive development (Flavell, 1982). Although controversial, Piagetian cognitive theory has had an enormous influence on developmental cognitive psychology (Flavell, 1982; Hull, 1994) and "as yet, nothing has replaced Piagetian theory" (Anderson, 1992, p. 115). According to Piaget (1954), the process of cognitive development embodies an invariant sequence of transformations or changes, beginning at

birth with the simple sensorimotor activities, and concluding in adolescence when teenagers are able to think symbolically and reason logically. For cognitive development to be completed, each individual must pass through four distinct periods of development, in sequential order. Each new stage of development, while qualitatively different from the preceding stage, is based upon the content of the previous stage. Piaget constructed his theory on the basis of a biological model in which the same basic mechanism of development underlies each stage. Piaget named this mechanism “adaptation”. Adaptation is a function that remains invariant and consists of two cognitive processes, assimilation and accommodation. Assimilation enables the infant to interpret new information and incorporate it into an existing pattern of behaviour, termed a “schema”. Accommodation allows an infant to change the existing schema to fit a new environmental experience. The interaction of these two processes results in a new balance called “equilibration”, and prepares the infant for the next stage of development.

The four stages of cognitive development are, in order, the sensorimotor stage, the preoperational stage, the concrete-operational stage and the formal operational stage. The first period of cognitive development begins from birth and lasts until 2 years of age. During this period the cognitive activities are generally restricted to sensory and motor abilities. The major intellectual function during this period is the concept of object permanency. At approximately 8 months of age an infant becomes able to retain a mental representation of an object and gradually comes to understand that the object may still exist even though it is not within view. Before developing this schema, a hidden object is held to no longer exist for the infant if it cannot be seen.

The preoperational stage takes about five years to be completed. Having established object permanency, children start to think about objects but their thoughts are still not

logical. The development of language and verbal communication are other aspects of mental development during this period.

The preoperational stage is followed by the concrete-operational stage. Children are able to reason logically but only in the presence of physical (not symbolic) events. One of the major achievements during this period is “conservation”, when children are able to recognize that equivalent quantities of objects are identical, even though they are presented in different forms. This period spans the four years from 7 to 11 years of age. Finally, the formal-operational stage, which begins around 12 years of age, enables teenagers to think logically and abstractly in the absence of any physical form.

Piaget's theory is more concerned with the understanding of mental processes and functions of intelligence than in individual differences in intelligence (Anderson, 1992; Ginsburg & Opper, 1988; Phillips, 1981; Uzgiris, 1989). Despite Piaget's theory making a major contribution to the field of cognitive development, it has been criticised as being unclear, incomplete and placing inappropriate emphasis on stage-like development (Flavel, 1982, 1992; Halford, 1989). Flavel (1982), for example, challenged the theory and maintained that human cognitive growth is generally not as stage-like as suggested by Piaget. Similarly, Halford (1989), Siegler and Richards (1982) and Wade and Tavriss (1990) maintained that, within a specific stage, many young children can demonstrate more cognitive skills or attain various aspects of cognition at earlier ages than Piaget thought.

(ii) *A single general factor during infancy:* Considering the difficulties that researchers confront with the conceptualization of *g* (Detterman, 1982; Jenson & Weng, 1994; Kranzler & Jensen, 1991; Thorndike, 1994; Vernon, 1989) and definition of intelligence in adulthood, the nature and definition of intelligence during infancy is even more difficult. Consequently, there is a variety of theoretical approaches, some of them

supported by empirical evidence as to the construct of infant intelligence. The first controversy is related to the conceptualization of a single factor equivalent to Spearman's *g* during infancy. Bayley (1955, p. 808) has acknowledged that "Something akin to *g*, or a high first-factor loading, must appear soon after the second or third year". McCall (McCall et al., 1972, 1977) have also questioned the existence of a common *g* factor throughout infancy. Based on principle component analysis, they suggested that the nature of *g* changes during the first two years of age. He and his colleagues analysed Gesell test scores obtained at 6, 12, 18 and 24 months in Fels' study population, and reported that the main developmental trend at 6 months is manipulation of objects and gross and fine motor behaviour. At 12 months, however, items requiring the imitation of fine motor behaviour, learning social behaviour and simple verbal skills seemed to define the main developmental trend. Further change was identified at 18 and 24 months, when verbal items appeared as the main component in the analysis.

In a similar vein, Lewis (Lewis, 1983; Lewis, Jaskir, & Enright, 1986; Lewis & Sullivan, 1985) suggested that there is little consistency across different ages and over different measures of infant mental functioning, suggesting little support for a common single factor in infant intelligence. They concluded that "... intelligence at any age is a set of separate mental abilities at each age rather than a single, general one... (Lewis et al., 1986, p. 350). Lewis postulated that since intelligence undergoes qualitative transformation during infancy, a multidimensional model of intelligence may provide a better picture of infant intelligence.

(iii) *Fagan's theory of infant intelligence:* Fagan's theory is based on the premise that, contrary to the positions of McCall et al. (1972) and of Lewis and Sullivan (1985), as outlined in the previous section, some biological basis equivalent to *g* exists from birth.

Information processing theory assumes that infants are intelligent because, from birth, they are capable of processing information and acquiring knowledge about their environmental context. Individual differences in intellectual functions during infancy are basically interpreted in terms of variation in infants' abilities to process available information. One implication of the theory is that it should provide evidence to explain the neural bases of intelligence. For example, an infant's response to novelty is held to be a fundamental behavioural manifestation of the speed of encoding and the retaining and retrieving of information from memory, in the absence of the original stimulus. Speed of such processing is held to be a biologically determined characteristic of intelligence (Anderson, 1992; Eysenck, 1986, 1988; Vernon & Mori, 1992).

Information processing theory regards intelligence as a function which consists of a small set of basic processes dedicated to the acquisition of knowledge (Fagan, 1984a). These processes are innate, relatively automatic and critical aspects of them do not change with age. According to Fagan, "Knowledge is gained as the result of assimilation, over time, of information by the intellectual processes" (Fagan, 1992, p. 82); and individual differences in the speed of knowledge processing underlies variation in the *g* factor. Consequently infants who encode information faster are capable of better memory performance and are expected to be more intelligent. Differences in the processing of information will be associated with later IQ because the intellectual processes necessary for a child to resolve an item in a psychometric intelligence test are held to be essentially the same as the processes an infant employs to differentiate between novel and familiar objects (Fagan, 1990; Fagan & Detterman, 1992). In contrast to a Piagetian stage-like theory, intelligence, as indexed by an infant's ability to encode, abstract, retain and retrieve information, is held to be constant over the life span (Fagan, 1981). Fagan's theory of the

constancy or continuity of intelligence from infancy is predominantly based on the study of visual fixation, with several investigators concluding that development in infancy, as indexed by fixation measures, is constant across childhood. Thus, intelligence, as measured by psychometric IQ tests, will also be constant. This theory therefore challenges the widely held contention that the nature of intelligence is not continuous and that "... a simple conception of a constant and pervasive *g* factor is probably not tenable as a model for 'mental' development, especially for the infancy period" (McCall et al., 1972, p. 746).

2-3 INFANT DEVELOPMENTAL SCALES

The first psychometric infant assessment scale was published in the mid 1920s (Gesell, 1925). The test focused on pragmatic and clinical utility and was largely based on items indicating the development of sensorimotor skills. The other pioneer in this field was Nancy Bayley, whose extensive work after 40 years of research resulted in the construction of the Bayley Scales of Infant Development (Bayley, 1969). Although it is recognized that several other participants (e.g., Buhler, Cattell, Griffiths, Hunt, Shirley and Uzgiris) have made extensive attempts to contribute to the field, the ideas underlying the infancy developmental scales have remained the same as those that were originally developed by these pioneers. That is, all infant developmental scales are based on several items measuring infant ability in performing gross and fine motor skills, sensory skills and perceptual/cognitive activities (Brooks-Gunn & Weinraub, 1983). In the following section some of the most frequently used infant developmental scales will be reviewed.

(i) *The Mental Growth of the Preschool Child*: The Mental Growth of the Preschool Child (Gesell; 1925) was the first mental scale developed for infants aged 2 years and younger. The scale was designed to identify infants for whom welfare and hygiene

were important (Yang, 1979). The scale covers six assessment points ranging between 3 months and 2 years of age. Gesell's scale is divided into four subscales measuring different areas of development - Motor, Adaptive, Language and Personal-Social behaviours. Many items in early months of life are related to the assessment of gross motor activities of the infant. The number of verbal items increases as an infant gets older. Although the test has been restandardized at different point in time, the majority of items in the revised versions (e.g., Knobloch & Pasmanick, 1974), remain the same as in the original version (see Self & Horowitz, 1979 for more comprehensive review of Gesell Developmental Schedules).

(ii) *Cattell Infant Intelligence Scale:* The Cattell Infant Intelligence Scale was developed by Cattell (1940). The test is claimed to be a downward extension of the Stanford-Binet test. However, the empirical data do not support that contention (see Yang, 1979 for a review of studies that have attempted to test correlations between the Cattell Scale and the Stanford-Binet). Many items, although modified, were borrowed from the Gesell test (Gesell, 1925) and were based on the measurement of sensorimotor activities.

(iii) *Griffiths Mental Developmental Scales:* On the basis of long experience in testing children, Griffiths published the first version of her scale in 1954. The Griffiths Mental Developmental Scales (Griffiths, 1954, 1986) consists of 260 items that have been classified into five groups of different items intended to measure different aspects of an infant's development. The five scales are the Locomotor Scale, the Personal-Social Scale, Hearing and Speech, Hand and Eye Development and Performance. Each scale involves 52 graded items arranged in order of difficulty for each month. The infant is scored on each item by pass or fail. Griffiths justified her work in terms of an urgent need for early diagnosis of mental conditions in some special cases. The inclusion in the test of a separate scale for assessing vocabulary was rationalized by the speed of development in language

skills during the first two years of life. The final sample for the test was 571 infants, and the age range encompasses infants from 1 to 24 months of age.

(iv) *Bayley Scales of Infant Development:* The Bayley Scales is one of the most researched infant tests. It was intended to measure an infant's competencies and observable behaviours, in order to provide an evaluation of an infant's current developmental status. Although the complete formal edition of the test was marketed in 1969 (*Bayley Scales of Infant Development*; Bayley, 1969), the work and data collection for the scales dated back to the early 1930s. That is, many items in the Mental and Motor Scales of the test were based on the earlier work of Bayley and her colleague on three California Scales - California First-Year Mental Scale (Bayley, 1933), California Preschool Mental Scale (Jaffa, 1934) and California Infant Scale of Motor Development (Bayley, 1936).

The Bayley Scales comprises three scales, each measuring some aspects of infant developmental status. The Mental Development Index (MDI) evaluates an infant's ability in perception, generalization, discrimination, problem solving, memory, vocalization and language. The development of motor activities such as sitting, standing, reaching, stepping movement and fine motor skills is assessed by means of a Psychomotor Development Index (PDI). Mental activities and motor abilities are substantially correlated with each other ($r = .50$ to $r = .60$; Bayley, 1969) during the first 12 months of life. An Infant Behaviour Record (IBR) form provides psychologists and paediatricians with a descriptive rating scale in which some characteristic behaviours during testing such as affect, motivation, personal social behaviour and attention are rated by the examiner.

Recently, a revision of the Bayley Scales was marketed (Bayley, 1993). Although the revision has the same structure as the earlier version, many test items and its basal and ceiling levels have been changed. In addition, the age range of the test has been increased

from 30 to 42 months in the second edition. The test-retest reliability and the correlation between the Bayley Scales and a psychometric intelligence test (i.e., McCarthy Scales of Children's Abilities; McCarthy, 1972), are reported to be relatively high.

All infant developmental scales, by and large, are based on a detailed description of the perceptual and motor development of babies, starting as early as the first month of life and finishing, for the most part, at approximately 2 years of age. Early test developers attempted to scale down the intelligence tests designed for older children (Lewis & Sullivan, 1985), and a close examination of items makes it clear that they are generally measuring three broad sets of activities - perceptual/cognitive behaviour, motor developmental milestones and personal social behaviour. There are many common items across the various scales (Brooks-Gunn & Weinraub, 1983; Honzik, 1983; Vietze, 1988). For example, if the relatively small age range of between 9 and 12 months is considered (an accurate comparison is not possible because of different age placements for the two tests), the following items are some, but not all, that are in common between the Griffith and Bayley Scales: React to mirror image, Vocalize different syllables, Uncover toy, Fine prehension, Stand up by furniture, Crawling, Sit well, Walk with help, Respond to verbal request, Plays pat-a-cake, Put cubes in cup, Imitate words, Retain two of three cubes offered, Rings bell, Jabbers expressively, Securing ring by string, Dangle ring, Pick up cup to find and secure cubes.

(v) ***Neonatal Behavioural Assessment:*** Apart from infant developmental scales that are generally designated for babies older than one month, during the 1950s and the 1960s some neonatal screening tests were developed by paediatricians. The initiator in the field was Graham who developed the first neonate test (Graham et al., 1956), but one of the most widely used tests is the Brazelton Neonatal Behavioural Assessment Scale (Brazelton,

1973). The main aims of most scales were to assess reflexive behaviour, muscle tonus, motor maturity, response to visual and auditory stimulation and to detect neurological symptoms (Worobey, 1990). However, these tests have sometimes been used as a predictive index of later intelligence (e.g., Field et al, 1978; Field, Dempsey, & Shuman, 1981; Rosenblith, 1979).

2-4 PREDICTION OF CHILDHOOD IQ FROM EARLY INFANCY

In response to the need for prediction of later IQ, psychologists and paediatricians have tried several procedures and instruments to predict later mental competencies from early infancy. Whereas paediatricians have employed physiologically based measurements, psychologists have used a variety of standard mental tests and other experimental observational and/or nonstructured observational procedures.

It has been suggested that several physical characteristics are related to later mental development. For example, it has been reported that a neurodevelopmental examination at 12 months correctly classified 82% of low birthweight subjects in terms of the Stanford-Binet IQ at 3 years of age (Ross, Lipper, & Auld, 1986, but see Bee et al. [1982] for poor predictive validity of infant physical status to 4- year IQ and language abilities). One frequently used physical measure is head size and circumference (e.g., Broman, 1989; Hack & Breslau, 1986). For example, Hack & Breslau (1986, 1988) found that head circumference at 8 months of age was one of the best growth parameters for both brain size and later intelligence at 3 years of age. Development of lateralization in hand preference in relation to brain specialization has also been used as a predictor for later IQ. Gottfried and Bathurst (1983), for example, found a relationship between handedness and later IQ for girls but not for boys. However, DiLalla et al. (1990) found no consistent results using two

measures of handedness. In the same vein, some researchers have attempted to use the proportion of sleep time (Beckwith & Parmelee, 1986), or the quality of the cry (e.g., Huntington, Hans, & Zeskind, 1990), or smiling (Crockenberg, 1983) as a predictor of later cognitive measures. Crockenberg reported that 3-month smiling and eye contact predicted MDI at 21 months of age.

In psychological domains, some psychologists have found significant correlations between infant preverbal communication and task orientation with parental IQ and/or infant later mental functioning (e.g., Benson, Cherny, Haith, & Fulker, 1993; DiLalla, et al., 1990). Socioeconomic status, including the level of parental education, has frequently been used as a reliable indicator of later IQ in infants (e.g., Laucht, Esser, & Schmidt, 1994; Rose & Wallace, 1985). Individual differences in sustained attention and manipulative exploration of objects at 8-9 months have also been found to be correlated with later cognitive function at 2 years of age (Kopp & Vaughan, 1982; Ruff, McCarton, Kurtzberg, & Vaughan, 1984).

Other measures frequently used as predictor variables have been home environment and caregiver interaction with the infant (Beckwith, Cohen, Kopp, Parmelee, & Marcy, 1976; Ruddy & Bornstein, 1982; Sigman, Cohen, & Forsythe, 1981) Maternal stimulation during infancy (e.g., feeding and teaching interaction, general environmental quality), for example, predicted cognitive/language competence at 2 and 3 years of age (Bee et al., 1982). Maternal encouragement of attention (attempting to direct infant attention toward an aspect of the environment), and maternal interaction with the infant to elicit responsiveness (smiling, vocalization, looking, touching and so on) predicted language comprehension (r s ranged from .29 to .35 all $ps < .05$) at 13 months (Tamis-LeMonda & Bornstein, 1989) and school performance and cognitive skills (WISC-R Vocabulary scores ($r = .34, p < .01$) at 6 years of age (Coates & Lewis, 1984). Some measures of visual expectation (defined in

terms of an infant's capability of forecasting a future event when they have no control over those events [Haith, Hazan & Goodman, 1988]) at 8 months have also been reported to be reliable predictors of IQ at 3 years of age, $r = .29$ to $r = .47$ (DiLalla et al., 1990) .

2-5 PREDICTABILITY OF INFANT DEVELOPMENTAL SCALES TO LATER IQ IN CHILDHOOD

Infant developmental scales were originally designed to measure different aspects of an infant's development and not as predictors of later mental functioning (Yang, 1979). For example, Griffiths has agreed that "... it is not to be expected that tests applicable to and legitimate as tests in the period of infancy, are necessarily similar in type to, or individually highly correlated with, tests suitable for measuring intelligence in older subjects" (Griffiths, 1986, p. 41). Similarly, Lewis and Sullivan (1985), after a comprehensive review of infant developmental tests, concluded that an infant's performance on blocks and rattles was not related to subsequent intellectual functioning. However, researchers have repeatedly attempted to judge the utility of such scales as predictors for subsequent intellectual status (see review by Brooks-Gunn & Weinraub, 1983). The rationale has been that, to the extent that early abilities constitute fundamental aspects of intelligence at a later age, scores on an infancy sensorimotor test should correlate with scores obtained from intelligence tests later in time. Several reviews based on many longitudinal studies have failed to provide an extensive predictive value in favour of conventional infant tests, at least during the first 18 to 24 months of life. This has been the case for normal babies (Bayley, 1970; Bee et al., 1982; Honzik, 1983; Lewis, 1983; Lewis & McGurk, 1972; McCall, 1979a; McCall et al., 1972; Wilson & Matheny, 1983) as well as for infants at risk (Hunt, 1981; Rose & Wallace, 1985). However, prediction from low score infants (e.g., Mental Development Index

[MDI]; Bayley Scales below 80) is more accurate than for normal babies (Crowe, Deitz, & Bennett, 1987; Honzik, 1983; but see Zelazo [1982] who holds that conventional infant tests are more likely to lead to a depressed estimate of intellectual ability for the developmentally disabled children). For the most part, the predictive validity of infancy tests has been unimpressive. The conclusion reached by Bayley summed up the inadequacy of developmental scales to evaluate later IQ: "It is now well established that we cannot predict later intelligence from scores on tests made in infancy. Scores may be altered... by developmental changes in the nature and composition of behavior tested..." (Bayley, 1955, p. 806).

In an early study, Escalona and Moriarty (1961) assessed the predictability of the Cattell and Gesell scales administered to infants aged between 4 and 32 weeks in relation to the WISC test administered to children at 7 to 8 years of age. They concluded that, for subjects younger than 20 weeks, there was no reliable correlation between infancy and childhood tests, and that infants' scores obtained between 20 and 32 weeks had only a limited predictive value. Similarly, an analysis of longitudinal data from Fels' population led McCall et al. (1972, p. 745) to conclude that "... until the second year of life there is relatively poor prediction from infant tests to IQ assessed in middle or late childhood". However, McCall qualified that statement by suggesting that some specific items (e.g., verbal items) and abilities might provide a better estimate of later IQ than others. There is evidence that this could be so (e.g., Ruddy & Bornstein, 1982; Siegel, 1981). For example, measures of awareness of surrounding and reaction to sounds (i.e., bell, rattle) obtained at 7 months from the Bayley Scales (Bayley, 1969) were predictive of later MDI at 1 and 2 years ($r = .19$ and $.18$ respectively, $ps < .05$) of age (DiLalla et al., 1990). Likewise, Siegel (1979) found that language items obtained from the Bayley Scales at 12 months were predictive of

later intelligence at 36 months of age.

In another study, Wilson (1983) examined data from the Louisville Twin Study and reported that developmental trends from age 1 to 2 years were lower than the correlation between age 2 and 3 years. Similarly, Rose and Wallace (1985) found that the 6 and 12 months Bayley MDI failed to predict cognitive outcomes between 1 to 6 years of age, although the 24-month MDI predicted later IQ at approximately 3, 3½ and 6 years of age. Seemingly, the size of age to age correlation from developmental tests is inversely related to the interval of time between test and retest for normal and at-risk infants (Largo, Graf, Kundu, Hunziker, & Molinari, 1990; McCall et al., 1977; Yang, 1979).

In a review by Fagan and Singer (1983), the correlations obtained for infant tests given during the 3 to 4, 5 to 7 and 8 to 11 month periods of life and later cognitive tests at 3 to 6 years and beyond were only .06, .15 and .21 respectively, for normal infants. For a clinical sample and among at risk infants the predictive validity, although slightly higher, was still low in terms of the magnitude of prediction. Compared with the long-term stability of intelligence tests for older children (Hindley & Owen, 1978; Moffitt, Caspit, Haskness, & Silva, 1993; Sameroff, Seifer, Baldwin, & Baldwin, 1993) the predictive value of infant tests was therefore very low.

There are some indications that the sensitivity of developmental instruments in detecting at risk infants is also low. Fagan et al. (1986) and Fagan, Singer and Montie (1985) have reported that MDI scores from the Bayley Scales given at about 8 months to a sample of 26 delayed infants correctly identified only 45% of delayed children and 53% of the normal children at 3 years of age. It has also been reported that developmental scales cannot identify individual infants who would be considered as gifted later in life. For example, Shapiro et al. (1989) have attempted to determine the ability of the Bayley MDI in

infancy to predict giftedness at 7½ years of age. The relationship between the Bayley Scales and later WISC-R was statistically significant ($r = .35, p < .001$). However, of 36 children who later proved to be gifted (i.e., children with IQ greater than 135 [WISC-R]), only two infants had an MDI greater than 134 at 13 months.

Nevertheless, a close examination of the literature finds clearly conflicting results reported for the predictive validity of traditional infant tests. The number of studies in which a reliable correlation has been found between infant sensorimotor activities measure and later cognitive outcome is not so small as to be disregarded (e.g., Broman, 1989; Roe, McClure, & Roe, 1983; Ross, Lipper, & Auld, 1985, 1986; Siegel, 1979, 1981, 1989). For example, in a longitudinal study conducted by Roe et al. (1983), it was found that scores on the Gesell Scales obtained from infants aged between 5 and 15 months correlated significantly with the WISC-R Performance Scale at 12 years of age (r s ranged from .43 to .66). The infancy scores did not predict the Verbal Scale, however. Roe et al. (1983) concluded that "... the failure of other investigators to find prediction of later intelligence from infant tests may have been due to the common tendency to think of "intelligence" as a unidimensional quality and to think that all "intelligence" tests measure the same underlying mental ability" (p. 145).

Siegel (1983) found correlations ranging from $r = .26$ to $r = .68$ between the MDI and PDI from the Bayley Scales, given at 4, 8, 12, 18 and 24 months of age to premature babies, and later intelligence scores at 5 years of age. She also reported significant correlations between Bayley scores at 4 to 12 months, the WISC-R at 6 years, and achievement scores at 6, 7 and 8 years of age (Siegel, 1985 cited by Siegel, 1989). Recently, Laucht et al. (1994), using a large number ($N = 226$) of 3-month-old infants at risk for subaverage IQ and non risk German infants, reported that the Bayley Scales (both MDI

and PDI) predicted later outcome for their sample at 2 and 4½ years of age (*rs* from .22 to .37 all *ps* < .001). However, there was an inverse relationship between the predictive value of early developmental measures and the interval of time for subsequent measures. This implies that, to some extent, there is continuity from the development of mental and perceptual motor functioning to intelligence at early and middle childhood. Siegel (1989) proposed that the problem of unpredictability in infancy tests may lie in the use of correlational studies, which are sensitive to attenuation of range in the distribution of scores. She concluded that, “In spite of a widely held misconception, infant tests predict later test scores and can detect which children are likely to show problems” (p. 101).

2-5-1 Explanation of the low predictive validity of developmental scales

The poor predictability of later IQ from developmental tests has led researchers to two different interpretations. The first is that the nature of intelligence is not stable or constant during infancy. This interpretation resulted in the hypothesis of discontinuity; that is, the nature of intelligence during infancy is highly canalized (McCall, 1981a) and cognitive skills reflect qualitative changes as infants mature (Piaget, 1954; Wohlwill, 1980). For instance, Wohlwill (1980) stated that; “... the essential and inescapable point is the marked discontinuity between development during infancy and the subsequent emergence of cognitive development (p. 393).

However, a second interpretation has argued that the nature of intelligence is continuous, and general intelligence, or *g*, is equally applicable to infant and subsequent cognition. This view therefore holds that the available measures are not suitable for eliciting cognitive power during infancy because they are so heavily dependent on elementary sensorimotor activities that they cannot tap the processes similar to those contained in adult intelligence tests (Brooks-Gunn & Lewis, 1988; Fagan & McGrath, 1981; Rose, Feldman, &

Wallace, 1988; Thompson, Fagan, & Fulker, 1991; Sternberg, 1981). In other words, what is assessed by the conventional infant test, particularly at very early ages is simple sensorimotor activity (e.g., turning head to sound, smiling, eyes following and blinking, manipulating and so on). These skills are not equivalent to the predominantly verbal and symbolic activities assessed by an intelligence test in older children or adults.

Siegel (1989) challenged these assertions, maintaining that, despite a difference in the “surface structure” of abilities measured in infancy and childhood, the “deep underlying structure” in both types of test remained the same. That is, behaviours in infancy and in childhood appear to be different but they are, according to Siegel, two representations of a single process. She alleged, for example, that, although the ability to grasp and reach at 4 months is different from eye-hand coordination at 6 years of age, these functions are correlated with one another and measures of both reflect a similar deep underlying structure- in this case, perhaps maturation of the nervous system. Siegel also suggested that if the prediction phenomenon is reconceptualized as identifying infants who would be likely to obtain a below average score later in childhood, as opposed to the exact score, the developmental test can be used as a predictor for later mental outcomes.

CHAPTER 3

INFANT INFORMATION PROCESSING TASKS

3-1 INTRODUCTION

Attention in infants has long been regarded as an avenue to cognitive assessment (McCall, 1971) and is thought to be the most important measure of infant cognitive integrity by some researchers (Lewis & Sullivan, 1985). Lewis (Lewis, 1982; Lewis, & Baldini, 1979) maintain that marked differences among children in attention appear to reflect individual differences in their central nervous system (CNS) integrity. New infant cognitive measures are predominantly based on infant attention, particularly differential visual fixation to new and previously experienced stimuli. Measuring the way an infant employs and sustains its attention in the environmental context has been found to have cognitive validity, and these functions are thought to be more analogous to the concept of intelligence in children and adults than are simple sensorimotor activities.

The contemporary paradigm for assessing an infant's attention, particularly measurement of looking behaviours, largely benefits from a technique that was initially developed by Fantz (1958a) to measure visual perception in infants. The technique was based on the assumption that, because animal infants (Fantz, 1956, 1958b) and human infants (Fantz, 1963, 1964, 1965) typically look longer at some stimuli than others, they are able to discriminate between two stimuli.

The Fantz procedure of visual preference was very simple. To estimate differential attention he relied on recording the reflection of a stimulus from an infant's eyes, to measure how long the infant looked at targets. Looking at the target by the infant is inferred to

reflect processing of the information contained in the stimulus. Preferential discrimination can be inferred by contrasting the amount of looking (fixation time) an infant accrues to a new stimulus as opposed, to the previously seen stimulus.

Newborns are congenitally equipped to engage in visual activities (Haith, 1980). From birth, infants show a capacity for visual differential and are more responsive to aspects of the environment with which they are not familiar (Fantz & Nevis, 1967). Although newborns can distinguish between highly discrepant stimuli (differing for hue, number of elements, size, brightness) the level of the infant's maturation (postconceptional age), the complexity of stimuli to be discriminated, and the time an infant needs to be exposed to a target to remember it, are critical in preferential looking behaviour (Cohen, 1988a; Fagan, 1970, 1972; Lasky, 1980; Richards, 1997; Rose, 1980, 1981; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982). Thus, for example, Catherwood and her colleagues found that infants at about 20 weeks from birth can categorize stimuli of uniform shape as well as stimuli that have dissimilar shapes in terms of their hue (Catherwood, Crassini, & Freiberg, 1989), whereas 39-week-old infants are capable of discerning categorical similarity and categorical discrepancy (Catherwood, Green, Skoien, & Holt, 1995). The general finding indicates that the optimal level for complexity of stimuli increases as infants get older (Cohen, DeLoache, & Strauss, 1979).

3-2 MEMORY IN THE INFANT

In the past 30 years, taking advantage of the preference technique, researchers have attempted to evaluate infant competencies in the early perception and cognition domains. This procedure is used to examine how infants perceive objects, as well as for tracing the growth of cognitive abilities in normal and/or retarded infants. Using this method,

researchers have found, for example, that newborns younger than 7 days are able to detect patterns with curved rather than straight contours (Fantz & Miranda, 1975), to prefer black and white patterns to plain discs (Fantz, 1963), and prefer horizontal to vertical square wave gratings (Slater & Sykes, 1977). Infants aged 6 to 16 weeks are capable of discriminating crosses from circles and red from black (Saayman, Ames, & Moffett, 1964). Similarly, it has been found that the 3-month-old infants, following familiarization to an object undergoing one of two types of motion, can significantly discriminate a novel motion from a familiar motion (Bahrick & Pickens, 1995). Fagan (1992) reported that a 5-month-old infant can process the attribute of form from the form-colour compound (preference for a green square as opposed to a green diamond after familiarity with a red diamond).

Taking advantage of the method used in preferential looking, researchers have established the ages at which infants can detect similarities and dissimilarities (categorization skills) across different representations (Younger & Gotlieb, 1988), spatial orientation (Colombo, Laurie, Martelli, & Hartig, 1984; Slater, Mattock, Brown, & Bremner, 1991) and spatial relations above and below a horizontal bar (Quinn, 1994). Researchers have also established the ages at which an infant is proficient at face recognition during the early months of life. Based on a series of experiments, Fagan reported that by 5 to 6 months infants are capable of face recognition with different orientation, and memorizing faces for different delayed time periods (Fagan, 1974, 1977; Fagan & Shepherd, 1979). By 6 to 7 months infants are able to discriminate among individual faces of the same age or same sex (Fagan & Shepherd, 1982). Based on the visual preference methods, there is now a large body of information regarding infant perceptual/cognitive abilities (Bornstein, 1992; Cohen, 1979). These findings were undeniably impressive, if not revolutionary (Bornstein, 1989a).

Concerning the predictive validity of the new measures, McCall, (whose conclusion, "... prediction coefficients from infant behaviour to later IQ are sufficiently low to be conceptually uninteresting and clinically useless" [1981b, p. 141] has repeatedly been cited in the relevant literature), finally acknowledged that, "Whatever the mechanism, it seems clear that the prediction phenomenon has been established..." (McCall & Carriger, 1993 p. 77). However, in terms of practical usefulness, the predictive value of these measures has often been found to be no better than other predictors derived from socioeconomic factors (McCall, 1994, 1989). Fantz's procedure now constitutes the basis of two different methods for assessing infant's intelligence, as indexed by their ability to process information. These are (i) habituation and (ii) visual recognition memory. In the next two sections the constructs underpinning these measures, their psychometric properties and the predictive power obtained from habituation and visual recognition memory tasks, will be evaluated.

3-3 HABITUATION

3-3-1 Construct and procedure

Habituation is defined as a decrement in an infant's attentional responding (e.g., visual fixation), as a result of repeated exposure to a stimulus. In the habituation method an infant is exposed continuously to a single "familiar" stimulus. Trials then proceed by presenting a new single "novel" stimulus to measure the infant's reaction to novelty (dishabituation). Decrement in attention is expected to occur as a result of comparing the matched stimulus across different trials and/or because of waning interest in the familiar target. In contrast, a recovery in attention would be expected when an infant makes a comparison between a newly presented stimulus and the mental representation of the familiar target (Bornstein, 1988, 1992).

The construct of habituation is a controversial one. While some researchers believe that habituation is simply a kind of conditioning response (Malcuit, Pomerleau, & Lamarre, 1988 a, b; see also Millar & Weir, 1992), others maintain that it reflects an infant's competency to acquire information about the environment (Bornstein, 1989b; Cohen, 1988a, b). Cohen suggests that habituation may tap memory and it should therefore be regarded as a process involving cognitive functions. To Bornstein, habituation is a type of learning (Bornstein, 1988; Bornstein, Pecheux, & Lecuyer, 1988), which perhaps tap into the same fundamental aspects of general intelligence during infancy that are necessary for mental functioning later in life. More specifically, habituation is held to consist of two different but related cognitive components. At first the infant develops an engram or mental representation of a familiar stimulus, by looking repeatedly at the target and encoding necessary information; and then compares this internal engram with the characteristics of the new stimulus. Discrimination between familiar and novel stimuli and retention of information from the familiar target are inferred to have occurred if the infant shows behaviour consistent with dishabituation or recovery of attention to a novel stimulus following the habituation phase.

This procedure exemplifies the information processing perspective of habituation. It derives from the work of the Russian physiologist Sokolov (1963), who was the first to offer an information processing model for habituation that was applicable to both animals and to human infants. Sokolov's model was a neurologically based, arguing that the orienting response to a repetitive stimulus results in the construction of a mental representation (neurological function) and any new stimulus is compared with this mental representation. According to the information processing approach, infants are capable of perceiving, encoding and categorizing the information to which they are exposed and they can

memorise this information to use in later recognition tasks. Many researchers have acknowledged that habituation reflects the maturation of the CNS and measures an infant's ability to process information (e.g., Cohen, 1973; Colombo & Mitchell, 1988; Deary, 1988; Rovee-Collier, 1988; Zelazo, 1982).

The developmental trend observed in young and more mature babies in the progression of habituation from simple stimuli to more complex stimuli does suggest an information processing interpretation of habituation. Older babies usually habituate more quickly than young babies, plausibly because of more developed cognitive skills (Bornstein, 1989b; Bornstein et al., 1988; Cohen, 1981, 1988a; Horowitz, Paden, Bhana, & Self, 1972; Slater, Cooper, Rose, & Morison, 1989). A second possibility, suggested by Cohen (1988a, b; Leahy, 1976) is that young infants habituate more slowly than older ones because, when they are presented with identical stimuli, they process each stimulus in terms of separate units or elements of information. They therefore have more items to remember, compared to older infants who process the same stimulus as an overall configuration or gestalt. Moreover, there is evidence of individual differences in speed of encoding information among babies. In general, intellectually disabled children and infants who are premature or at risk for later mental disability perform less well when compared to their normal counterparts (e.g., Cohen, 1981; Millar et al., 1991; Ross & Ross, 1981; Shepherd & Fagan, 1981). Still another reason is the level of complexity of the stimulus to be processed, with more complex stimuli requiring longer processing time in contrast to a simple or unidimensional stimuli (Turpin & Rose, 1988; but see Brown [1974] who did not support the hypothesis that the habituation rate is linearly related to the complexity of the stimulus). The construct of habituation is still in dispute, and as Lecuyer (1988) suggested, "... we still have to learn about what infants of different ages are really doing during a habituation

experiment” (p. 479).

Typically, two different procedures are in use for measuring habituation. The first procedure, termed the “fixed trial” technique, was developed by Fantz (e.g., Fantz, 1963). Habituation is assessed during a fixed number of presentations (determined by the experimenter) of a familiar stimulus for a fixed amount of time in each trial (depending on the infant's age). A new stimulus is presented immediately after the habituating criterion has been met. In the second procedure, an “infant-controlled” regimen (Horowitz et al., 1972), the number and the length of trials is, as suggested by the name, determined by the infant. The habituation phase is continued until the infant reaches a predetermined criterion of decrement in its attention (normally 50% decrement in the length of looking relative to the first fixation). A new stimulus is then presented to measure the recovery of attention to new targets. The inter-observer reliability for habituation measures has been found to be high, around 0.95 (Cornell, 1979; Millar & Weir, 1992).

Habituation is not confined to visual behaviour and some researchers have habituated their subjects when taking advantage of other modalities, such as an auditory response (Benson et al., 1993; Morrongiello, 1990; O’Connor, 1980; O’Connor, Cohen, Parmelee, 1984) and audiovisual stimuli (i.e., habituating infants using the film of a person speaking and automobile with engine running approaching and receding with coherent soundtracks; Schiff, Benasich, & Bornstein, 1989). In addition to behavioural responses, experimenters have used other measures, like skin-potential reaction (e.g., Stechler, Bradford, & Levy, 1966) or cardiac change (e.g., McCall, 1979b; O’Connor et al., 1984) or sucking response (Lewis, 1982; Siqueland, 1981) and head orienting (Swain et al., 1993), to habituate their subjects.

3-3-2 Dependent measures of the habituation procedure

A variety of measures obtained during the habituation procedure are normally used as predictor measures. The most frequently used measures are total length of fixation duration to familiar stimulus across habituation trials, the number of trials to criterion, duration of first look, the average fixation duration of looks across all trials, duration of peak fixations (the longest fixation during the habituation trials), total of peak fixation, number of looking responses, percent of response decrement to familiar stimulus (habituation), as well as the magnitude of response recovery (dishabituation). Predictor measures have commonly differed from one study to another. Two measures have more commonly been used as predictors of later mental outcomes. They are the length of fixation to a familiar stimulus (e.g., Sigman, Cohen, Beckwith, & Parmelee, 1986), and the recovery of response to a novel stimulus (e.g., Lewis & Brooks-Gunn 1981). Moreover, some studies (e.g., D. Rose, Slater, & Perry, 1986) have found that certain habituation measures predicted later outcome; but other measures derived from dishabituation (recovery to novelty) for the same infants did not predict later general or specific aspects of mental functions. In contrast, Lewis & Brooks-Gunn (1981) failed to find predictive validity for habituation measures across three months for one sample, whereas dishabituation measures were predictive for both samples. Seemingly, prediction is not consistent across different measures within a study or for one specific measure across different studies. This point is clearly articulated by Malcuit et al. (1988b) who found the predictive value of these measures to be unconvincing and believed that "... the significant correlations reported in the literature are based on too many different variables and measures to be conceptually consistent" (p. 545).

There is also disagreement concerning the predicted measures. For example, while Bornstein and Sigman's (1986) review suggested that verbal IQ is predicted better than

performance IQ, the results of Laucht et al. (1994) indicated that the correlations between habituation measures and nonverbal outcome were superior to those of verbal IQ. Similarly, the results have not always been consistent across gender. For example, Cohen, Gelber, & Lazar (1971) reported that boys habituated more rapidly than girls.

3-3-3 Test-retest reliability and long-term stability of habituation measures

Data obtained from studies in relation to this issue have been mixed, indicating widely different results that have ranged from low or nonexistent test-retest reliability at short-term, to moderate test-retest stability at a relatively longer term. The considerable number of contradictory results in the literature is an indication of the main concern among researchers about the reliability of these measures. Overall, those who are emphatic concerning the usefulness of the habituation measures generally maintain that the reliability of the measures are moderate (Bornstein, 1989a; Bornstein & Benasich, 1986; Colombo & Mitchell, 1990). On the other hand, others remain sceptical about the theory, pointing out that the main problem with these procedures is that the low reliability of the measures does not meet criteria necessary for an adequate psychometric test (Malcuit et al., 1988a).

In the habituation procedure, while several investigators have reported some degree of reliability for qualitative patterns and measures of fixation duration or recovery of attention to a new stimulus (e.g., Bornstein et al., 1988; Colombo, Mitchell, O'Brien, & Horowitz, 1986; Fenson, Sapper, & Minner, 1974) others have failed to find consistency and stability for these measures (e.g., McCall, 1979b; Pecheux & Lecuyer, 1989, 1983; Pomerleau, Maitre, & Malcuit, 1989). For example, Bornstein et al. (1988, p. 132), who examined the stability of individual differences in habituation in 3 to 5 month old infants, reported that "In overview, habituation appears to be stable in infants, but the size of the stability coefficient is considerably influenced by the duration of the interest interval...".

Other influences included the infant's age, the dependent measures, the stimuli and the habituation procedure. For instance, Pecheux and Lecuyer (1983) found that the measure of total fixation time to criterion in 4-month-olds was stable within a session and between sessions two weeks apart, whereas the measure of the number of trials to criterion was never reliable. Another example of this comes from Colombo's examination of the stability of visual habituation in infants aged from 3 to 9 months (Colombo et al., 1986). These researchers found that duration-based and habituation measures showed moderate week-to-week reliability ($r = .30$ to $r = .50$); but over a longer interval only the duration of peak fixation was consistently stable. In addition, the test-retest reliability showed some degree of age-related variation. That is, habituation indices were more reliable at ages 4 and 9 months than at 3 and 7 months. The most perplexing aspect of the results was that, while no consistent short-term test-retest reliability for the measure of dishabituation emerged, a long-term stability was observed for this measure between 5 and 9 months. Colombo et al. (1986, p. 482) acknowledged that "This isolated relation is difficult to interpret because it is the only significant longitudinal correlation for this variable, because it occurs over a long (5-months) interval in the absence of significant correlations over shorter intervals, and because dishabituation did not exhibit within-age reliability at these two ages". Overall, it seems that the most reliable measure of the habituation method is fixation duration in the short-term, but no other measure shows a consistent stability over a longer period of time.

The generality of habituation measures across different stimuli and different sensory modalities (i.e., between rate of habituation to visual and auditory stimuli) is also very low (Kagan, 1989; McCall, 1988, 1989). McCall (1979b), for example, questioned the use of habituation indices as a measure of the rate of processing information, because he found that neither the pattern of habituation across two modalities within an age, nor within modalities

across ages from 5 months to 10 months, were stable (but see Bornstein & Benasich, 1986).

3-3-4 Concurrent and longitudinal predictive validity of habituation

(i) *Concurrent validity*: Some research indicates that habituation is sensitive to individual differences between normal and intellectually disabled infants (e.g., Cohen, 1981) and provides evidence for its long-term predictive validity. To the extent that habituation is a cognitive process it is expected to be an indication of individual variation in terms of mental efficiency, and to be capable of discriminating normal infants from those who are in the potential risk category for some reason, or who have already developed some form of retardation. In an earlier study Miller et al. (1977) for example, showed a concurrent validity for habituation. Infants who showed a steeper decrement in their fixation duration to the stimulus outperformed those who had shallower decrements. A recent study by Millar et al. (1991) compared some habituation measures in 4-12-month normal babies with infants who were at risk because of CNS related factors. They found that, while normal infants exhibited response decrement and recovery to novelty, babies who were at risk showed neither habituation nor dishabituation responses. In other studies it was found that infants who were born prematurely performed more poorly than full-term infants on the habituation task, either in the rate of habituation and/or in response increment to novel stimuli (Caron & Caron, 1981; Sigman & Parmelee, 1974; but see O'Connor, 1980 who failed to demonstrate such differences). The results also suggested that the habituation procedure is sufficiently sensitive to differentiate between normal and intellectually disabled infants (Brooks-Gunn & Lewis, 1988; Cohen, 1981; Lewis, 1982; Lewis & Brooks-Gunn, 1984; McDonough, 1988; McDonough & Cohen, 1982). For instance, Cohen (1981) has found that Down's Syndrome babies did not habituate at the youngest age (i.e., 19 weeks) examined. They did habituate to a familiar stimulus but did not show recovery to a novel test at an intermediate age (i.e.,

23 weeks) and only the oldest group (i.e., 28 weeks) demonstrated both habituation and recovery of preference for a novelty item.

(ii) **Predictive validity:** Using the habituation paradigm during infancy, several researchers have found moderate predictability for IQ at a later stage of development (Brooks-Gunn & Lewis, 1988; Cohen & Parmelee, 1983; Lewis & Brooks-Gunn, 1981; Miller, Sinnott, Short, & Hanis, 1976; O'Connor, 1980; O'Connor et al., 1984; D. Rose et al., 1986; Ruddy & Bornstein, 1982; Sigman et al., 1986; Tamis-LeMonda & Bornstein, 1989; see also reviews by McCall & Carriger, 1993; Bornstein & Sigman, 1986). Miller et al. (1976), for example, found that infants at different ages (2, 3 and 4 month-olds) who habituated quickly were more advanced in cognitive scores derived from the Uzgiris and Hunt Scale at 14 to 15 months of age, relative to slow habituating infants. Cohen and Parmelee (1983) have also found that an infant visual attention test administered at term predicted developmental outcome at age 5 years. Similarly, language comprehension and play at 13 months was found to be predictable from measures derived from habituation at 5 months (Tamis-LeMonda & Bornstein, 1989).

One of the longest periods for predictive validity reported for habituation with preterm infants at term and at 4 months of age came from the Sigman team. Sigman et al. (1986) followed up their subjects at 8 years of age. They found that a measure of habituation obtained as early as at term was significantly related to childhood intelligence assessments (r s ranged from $-.27$ to $-.36$, all p s $< .05$, $N = 91$). Finally, in recent research conducted in Germany by Laucht et al. (1994), a significant but low correlation (r s from $.12$ to $.21$) was found for at risk infants between response decrement and response recovery measures at 3 months and later cognitive outcome at 54 months.

There is some indication of sex differences in the predictability of habituation

measures. For example, Fenson et al. (1974) reported that fixation time and habituation rate predicted rate of satiation in play with a familiar toy and response recovery for a novel toy for females, but this was not the case for males. O'Connor (1980) has also found that responsiveness to auditory novelty was a strong predictor of later Bayley scores at 18 months, but only for girls.

To summarize, Bornstein's (1989b) comprehensive review, based on 7 longitudinal studies, estimated a correlation of .49 between habituation measures during infancy and later intelligence at ages between 2 and 8 years. Bornstein and Sigman (1986) have also found that habituation measures and cognition in later childhood share 15% of the common variance.

3-3-5 Criticism of habituation (theoretical concept and practical issues)

There is considerable dispute concerning the construct of habituation and its utility as an instrument of prediction of cognitive performance later in time (Bloch, 1988; Kagan, 1989; Lecuyer, 1988, 1989; Malcuit et al., 1988a). It is claimed, for example, that some infants are able to distinguish between familiar and novel stimuli before becoming habituated to the stimulus (Kagan, 1989). In this case, then both the concept of habituation measures as predictors and the practice of categorizing infants as fast or slow processors are questionable. In addition, the use of visual fixation measures raises some practical problems. For example, it is very difficult, if not impossible, to differentiate between blank stares and fixation involving processing. This point makes acceptance of rate of habituation as a measure of speed of information processing (Lecuyer, 1989; McCall, 1979b), or as a useful construct (Kagan, 1989) debatable.

While acknowledging the usefulness of habituation as a methodological procedure and as an essential cognitive activity, Malcuit et al. (1988a) have argued against a complex

cognitive interpretation. Instead, they insist that habituation is a form of operant conditioning - that is, a simple activity (e.g., visual fixation) maintained by a “synchronous reinforcement schedule”. In this case a new stimulus or a more complex stimulus serves as a reinforcing target. Others believe that, although habituation is a form of conditioning, it reflects individual differences in attention between infants (e.g., Fagen & Ohr, 1990). However, many researchers maintain that the reinforcement explanation for habituation is misleading because the type of learning in habituation is different from that formed in operant conditioning (e.g., Slater, 1988; Turpin & Rose, 1988). Comparing habituation and contingency learning in normal and high risk and low risk infants, Millar & Weir (1992) suggested that there is a complex relation between measures of habituation and conditioning learning. Both procedures involve distinct aspects of cognitive processing.

For researchers who have questioned the possibility of prediction of later IQ from habituation parameters, there is still an enigmatic relationship between rate of habituation and later mental performance (Kagan, 1989). These researchers have attributed the observed relation between habituation and later outcome to sample characteristics or other cognitive and noncognitive characteristics that are shared in infancy and childhood. For example, Kagan, Lapidus and Moore (1978) observed that parental social class was more successful in predicting cognitive variables than infancy attentiveness measures and suggested that parental stimulation in the home mediated prediction. However, Tamis-LeMonda & Bornstein (1989) noted that habituation measures at 5 months predicted language comprehension and play at 13 months of age and appeared to be not dependent on the mother’s encouragement to attend. They concluded that “... links between early habituation and later cognitive development are direct and not solely mediated by maternal stimulation...” (p. 738).

McCall (1989, p. 184) speculated that infant characteristics play an important role in the prediction of later IQ. He suggested that an infant with a “disciplined, goal-directed, controlled attentional disposition” is the infant who attends to an experimental stimulus and the child most likely to habituate rapidly. A similar interpretation was given by Lecuyer (1988, 1989) who suggested that fast habituating infants are those who have greater ability to remain oriented on task, rather than those who have greater encoding speed for processing. If that is the case, task orientation may mediate the correlation between infancy performance and later IQ, as suggested by McCall (1989) and empirically observed by DiLalla et al. (1990). Therefore, it is possible that one factor that influences the length of visual fixation is the infant’s state of arousal, which subsequently determines the rate of habituation and possibly generates an underlying mechanism for the continuity of intellectual abilities (Sigman, 1988; Sigman et al., 1986). It would appear that the complexity of the habituation paradigm and its measures is bound to cause dispute. While Malcuit et al. (1988a) insist that habituation should be studied in its own right and not as an instrument, Sigman (1988) holds that the study of habituation is important not only in its own right but also because it provides a new measuring instrument.

Despite the popularity of the habituation procedure, from a methodological point of view the use of this procedure has been criticized, and it has been argued that the procedure underestimates an infant's discriminative competence in processing information, because infants can probably process stimuli long before they reach criterion (Fagan & Singer, 1983; but see Caron, Caron, Minichiello, Weiss, & Friedman, [1977] for constraints applying on the familiarisation-novelty method). McCall (1994) has pointed out that data from the habituation procedure suggested that encoding may take place very rapidly, even in a single presentation of a familiar stimulus, and not over several trials of familiarization. Similarly,

Fagan (1974, p. 356) noted that in the habituation procedure "... recognition may be accomplished long before any reliable decline over trials can be demonstrated...". It is claimed that recognition memory as administered within the paired comparison procedure provides a more efficient technique for assessing an infant's recognition memory capacity (Fagan, 1990).

3-4 VISUAL RECOGNITION MEMORY

3-4-1 Theory and procedure

Recognition memory, as opposed to familiarity preference, is another procedure employed to determine infant preference for novelty. The following account is in terms of the work of Colombo, Mitchell, Dodd, Coldren and Horowitz (1988), Fagan (1984b), Fagan and McGrath (1981), Rose, Feldman, Wallace and Cohen (1991), Rose and Wallace (1985), Rose et al. (1988a, 1992) and other researchers in the field.

The general procedure for the paired comparison or recognition memory method consists of a familiarization phase and a test phase. First, an infant is exposed to two identical stimuli or one stimulus, called the "familiar" or "standard" stimulus, for a preset duration, called "familiarization time". The familiar target is then paired for a predetermined time with a new "novel" target in two successive trials, in which the position of the stimuli is reversed to prevent position bias. The two successive exposures of familiar and novel targets together are called the "test phase" or "novelty trial". Familiarization time is determined as a function of infant age (Rose, 1981, 1983; Rose et al., 1982) and the complexity of stimuli to be processed. For the normal infant, the expectation is that the child will cease to look at the familiar target and will instead attend to the novel stimulus for a longer period of time. The assessment of recognition memory is not confined to the

measures of visual fixation only and some researchers take advantage from other measures such as heart rate deceleration (e.g., Richards, 1997) or high amplitude sucking (Siqueland, 1981) to evaluate the the percentage of novelty preference.

Although different measures can be extracted from familiarization and the test phase, the novelty preference is customarily regarded as an index that reflects the processing of information and that correlates with later mental outcome. Novelty preference is operationally defined as the amount of fixation time to a novel stimulus, expressed as a ratio to total looking time. It is computed simply by the total duration of time for which the infant attended to the novel stimulus, divided by the sum of looking time at familiar and novel stimuli, and multiplied by 100. If this percentage of looking time - the time spent looking at the novel stimulus - significantly exceeds the chance level of 50%, it is regarded as an indication of preference for novelty. It is quite possible, as is expected in the case of some aberrant or at risk infants, that when the familiarization time is not sufficient to make discrimination between stimuli possible for an infant, he or she will score around chance level or significantly less than 50% of preference for novelty (i.e., familiarity preference). Significant preference for familiarity could also be regarded as an indication of processing information, but it reflects a more immature or earlier phase of processing compared with novelty scores that represent a more advanced phase of processing (Rose et al., 1982).

Similar to the habituation method, it is assumed that during familiarisation the infant encodes available information and forms a mental engram from the familiar target. When a new stimulus and a familiar stimulus are presented simultaneously, the infant prefers to look at the novel target because it is different from the already formed engram. Thus, a visual recognition memory procedure reflects some cognitive skills involving processing of information.

3-4-2 Psychometric properties of Visual Recognition Memory tasks

The psychometric characteristics of infant information processing measures obtained from paired comparison method are not consistent across different studies. Internal consistency and test-retest reliability at short term and cross-age stability at a longer interval time, are generally low, and moderate at best. Similarly, generality of information processing measures across different modalities is low. For example, none of the measures of cross-modal transfer, tactual recognition and visual recognition memory at one year of age was found to be related to another (e.g., Rose et al., 1991a).

Nevertheless, several researchers have concluded that the measures obtained from this procedure show a moderate degree of reliability (Colombo, Mitchell, & Horowitz, 1988; Rose & Feldman, 1987). For example, Colombo et al., (1988, p. 1207) who assessed the stability of different measures from a novelty task, argued that "... nearly all measures of infants' visual attention showed good reliability from one week to the next and moderate stability across the period from 4 to 7 months of age".

Low reliability is one of the main concerns regarding information processing tasks. It limits the predictive validity of the task and makes the test ineffective. Theoretically, low internal consistency and low test-retest reliability make the construction of a test debatable, and lead to the conclusion that what is assessed by this type of task, at least in terms of the passage of time, is not a stable structure. Although many suggestions have been offered to explain the low reliability of such novelty tasks, the main reason appears to be the uncompromising variability of infants' behaviour at early ages. The rapidity of developmental change, in terms of both physical and mental growth, make the stability of the test at different ages unlikely. Similarly, temperament, restlessness, excitement, fatigue and many other factors can result in differing results from one session to the next. In

addition, the methodological problems in measuring infants' reaction to novelty may also produce a low internal consistency and test-retest reliability for the task within a session and between sessions.

3-4-3 Concurrent validity for measures of visual recognition memory

Evidence for concurrent validity of visual recognition memory tests comes mainly from studies in which infants who are at risk for cognitive delay later in life perform less well on discriminatory tasks in comparison to their normal counterparts. Using a paired comparison procedure, several studies have found that, compared to full-term infants, low birthweight and premature infants display a different pattern of looking (Rose, Feldman, McCarton, & Wolfson, 1988) and less differential attentiveness to novelty, at least in the first year of life (Caron & Caron, 1981; Rose, 1980; Rose et al., 1992; Rose & Feldman, 1990; Sigman, 1983; Sigman & Parmelee, 1974). These at risk infants have also required longer exposure to a target than full-term infants in order to show a significant preference for novelty (Rose, 1980, 1983). Preference for novelty in 7-month-olds intrauterine growth retarded infants who were at risk for later retardation was also found to be lower than for normal infants (Gotlieb, Biasini, & Bray, 1988). Ruff (1988) maintains that differences between full-term and preterm infants are due to the integrity of CNS and, because this integrity (as reflected by visual fixation measures) varies for full-term and preterm, the two groups show different patterns of looking behaviour.

Likewise, infants from non-normal population have been found to score lower than infants from the normal population on preference for novelty in recognition memory tasks. Down's Syndrome infants, for instance, required more familiarization time to process information (Miranda, 1970), and/or longer period of maturation to process stimuli (i.e., several weeks later) than normal infants (Miranda & Fantz, 1974). In addition, when

familiarization time was equal, they scored lower on novelty preference than normal infants (Miranda & Fantz, 1973).

Finally, an adapted form of the FTII was used with infant animals to detect later cognitive retardation. In several experiments conducted in Gunderson's laboratory (Gunderson, Grant-Webster, & Sackett, 1989; Gunderson, Gran, Burbacher, Fagan, & Mottet, 1986), it was found that whereas normal infant monkeys showed significant novelty preference, a low birthweight infant and an infant who was at high risk for later cognitive delay due to developmental problems or exposure to methylmercury, demonstrated only chance performance, or significantly less visual preference for novelty.

3-4-4 Prediction of later IQ from visual recognition memory tasks

Several investigators have reported that variation in early preference for novelty is related to later childhood IQ. Although some researchers found that a combined score of visual recognition memory and cross modal transfer tasks explained 35% to 51% of the common variance between novelty preference and later mental outcome (Rose et al., 1991a), most researchers have reported a low predictability for this type of task, even as low as 4% of the shared variance (e.g., Andersson, 1996).

The first systematic study employing an index of preference for novelty and attempting to relate these scores to later IQ was conducted by Yarrow, Klein, Lomonaco, & Morgan (1975). These authors found that 6-month old infants who played more with a new toy, compared with the toy which they had previously explored (a bell), had a higher IQ at 3 years ($r = .35$ $N = 39$). However, the first study that reported on predictive validity, using the paired comparison paradigm to assess visual recognition memory, was Fagan and McGrath's study (1981).

In the 1980s and early 1990s several such studies were published indicating a

relationship between novelty scores and potential intellect (DiLalla et al., 1990; Fagan et al., 1985, 1986; Rose et al., 1992 [see below]; Thompson et al., 1991; see also Fagan, 1988a, b and Rose, 1988). Ostensibly, measures of infants' visual attention are not only related to general intelligence, but also correlated with some specific cognitive skills such as language (Rose et al., 1991b), exploration abilities and memory (Colombo et al., 1989; Thompson et al., 1991) and academic achievement (Rose et al., 1992). Colombo et al. (1989), for example, attempted to assess the usefulness of a visual discrimination task in predicting cognitive skills at a later age. They found that novelty preference at 7 months of age was correlated with a memory task at 16 months of age ($r = .51, p < .05, N = 22$), but that the same novelty preference measure made earlier at 4 months did not predict subsequent memory performance.

The major work and the greatest number of publications in this area have come from Fagan and his colleagues at Case Western Reserve University, and Rose and her associates at Albert Einstein College of Medicine, both in the United State of America. Colombo and his associates also have extensively studied infants' visual attention, but most of their studies have focussed on the habituation paradigm and on testing concurrent and construct validity, rather than the predictive value of attentional measures. Studies from Fagan's laboratory and the studies which used some tests of the Fagan test will be outlined later in this chapter and in Study 2. Following are some studies conducted by Rose and her associates.

Studies from the Rose Laboratory: In the late 1970s, Rose began her studies on cognitive development in infants, with an interest in determining individual differences between full-term and at risk infants from a premature infants population (Rose, 1977; Rose, Gottfried, & Bridger, 1978, 1979). The first study from which the prediction data were published began in 1980 and assessed visual recognition memory in preterm and full-term

infants (Rose, 1980). The results of this study were published by Rose and Wallace (1985), who reported that novelty scores obtained at 6 months of age were significantly related to cognitive measures of MDI and/or IQ between 2 and 6 years ($r = .53$ to $r = .66$, $N = 13$ to 28). In the second longitudinal study of perceptual development in infancy, Rose and her fellow researchers recruited a cohort of 63 preterm and 46 full-term infants from families who were predominantly from lower socioeconomic circumstances. They found that preference for novelty in tasks such as visual recognition memory, tactual recognition and cross modal transfer at different ages ranging from 6 to 12 months, predicts later mental outcome in childhood. In several publications, Rose reported different predictive values for her subjects at different ages during early childhood. For example, Rose and her colleagues found that novelty preference can predict 1, 1½, and 2 years MDI (Rose, Feldman, Wallace, & McCarton, 1989), 3 to 5 years IQ (Rose et al., 1991a, 1988a) and language development at 2, 3 and 4 years (Rose et al., 1991b).

In the latest publication from Rose's longitudinal studies, Rose et al. (1992) have reported predictive validity on information processing measures (visual recognition memory, cross-modal transfer) and object permanence tasks for her subjects at 6 years of age. The scores on 7-month visual memory significantly predicted later IQ (WISC-R and Raven Progressive Matrices), receptive and expressive language, several domains of perceptual organization, and reading and quantitative scores ($r = .24$ to $r = .47$, $N = 74-77$). However, 12-month visual memory failed to correlate with Full-Scale, Verbal IQ and/or Performance IQ (WISC-R) at 6 years. Rose et al. (1992) also reported that cross modal transfer at one year related to general IQ (Full-Scale, Verbal IQ and Performance IQ; WISC-R) and some subtests of language, reading and quantitative proficiency ($r = .23$ to $r = .31$, $N = 86-89$). With one exception, there was no statistically reliable relationship between

tactual memory, another measure of information processing, and later mental outcome at 6 years of age. The exception was a correlation of .26 between tactual memory and Performance IQ.

3-4-5 Review of meta-analyses of infant visual recognition memory

There are currently three publications in which authors have reviewed studies that reported a correlation between early visual recognition memory and later cognitive efficiency (Bornstein & Sigman, 1986; Fagan & Singer, 1983; McCall & Carriger, 1993). Recently, Slater (1995) also briefly reviewed the literature of the field and summarised some results from visual recognition memory. The reviews reported median correlation (ranging from $r = .37$ to $r = .47$) between novelty preference and later mental competencies. The studies were based on different ages at the initial test (mainly 4 to 9 months) and a subsequent testing session (1 to 8 years).

Bornstein and Sigman (1986) reported that average common variance shared by infancy measures (i.e., recovery of attention) and childhood test scores was 22%. Bornstein and Sigman (1986, p. 268), who were confident about the power of attentional measures in their review to predict future intelligence, finally commented that, "We find it telling that these particular patterns of attention [decrement and recovery of attention] in infants have become established and manifestations of cognition in early life, namely, how rapidly and completely infants construct schemas of their visual and auditory worlds and how resistant to decay or interference those schemas may be."

McCall and Carriger (1993) have also analysed data from information processing tasks. The authors concluded that, "... individual differences in habituation and recognition memory do seem to predict later intelligence test scores better than those associated with standardized tests of general developmental level." (p. 58). This was found to be true for

risk and nonrisk infants, seemed unlikely to be produced by extreme scores, and the level of prediction was substantial when the low reliability of infancy tests was taken into account. However, the reviewers acknowledged that the practical implications of these measures were limited because of “sheer prediction accuracy”, and that they were more important in terms of understanding the process of mental development than prediction itself. (McCall & Carriger, 1993).

3-4-6 Critique of the literature and meta-analyses of the predictive validity of novelty tasks

The field seems to provide ample evidence that attentional measures taken in the first year of life can, to some extent, predict later mental functioning. However, a close examination of these reviews, particularly in the field of recognition memory, raises some concerns about the predictive value of the procedure and its measure, novelty preference. When taken overall, the concluding comments of the reviewers regarding the prediction phenomenon have been generally accepted, but breaking the data down according to the different procedures and the different measures used in the studies reviewed, does reveal that existing evidence may be less convincing than is commonly claimed. In addition, it does appear that the number of independent studies with independent subjects are much lower than the three mentioned literature reviews suggest.

Specifically, five points should be emphasized regarding all data reviewed in these meta-analyses. First, examination of the review by McCall & Carriger (1993) involving 31 samples, reveals that the results of 12 samples are from habituation studies and 19 from recognition memory studies. Ignoring habituation studies that are not of concern in the present project, three types of predictors in recognition memory studies were analyzed. Although different measures may be conceptually similar they are obtained by different

procedures. The most common procedure used by the studies reviewed is visual recognition memory. When the measure of recovery of attention, cardiac deceleration to a new stimulus (the procedure used for these studies is in fact the habituation paradigm and not recognition memory) and visual anticipation (again a different paradigm with a different predictor measure) are excluded, only 15 samples remained in the analysis. While the median raw correlations for Rose's and Fagan's samples are $r = .53$ and $r = .47$ respectively, the median raw correlation for the other five samples is only $r = .29$.

Slater (1995) summarised seven studies that have correlated visual recognition memory scores with later language proficiency, academic achievement, Stanford-Binet and Wechsler intelligence tests outcomes. The mean raw correlation of DiLalla's (1990) study was only $r = .10$; the mean raw correlation for the samples from the Fagan and Rose laboratories were $r = .38$ and $r = .47$ respectively.

Similarly, Bornstein and Sigman's (1986) review reported correlations for the Fagan study and Rose's samples as $r = .46$ and $r = .56$ respectively. In this review, the early study conducted by Yarrow et al. (1975) is reported as finding a correlation of $r = .35$ between preference for novelty and later Stanford-Binet at 3 years of age. Finally, the review by Fagan and Singer involved seven studies, five of which were from the Fagan laboratory, again with correlations ranging between $r = .36$ to $r = .57$.

Table 3-1 presents the number of samples in these four mentioned reviews. As can be seen from the table, the majority of studies reported came from the Rose and the Fagan laboratories and many of studies overlap (see Table's note). It is possible that the relatively identical predictive value found in all the mentioned reviews is a consequence of overlapping between the Fagan and Rose samples included in the reviews. While acknowledging the concluding comments by McCall and Carriger (1993), which stress that

data from *one* laboratory are not responsible for the prediction phenomenon, it is possible that the predictive value (at least in the paired comparison paradigm) is heavily dependent on the source of data, the majority of which came from *two* laboratories only.

Table 3-1

The Number of Samples in Four Meta-Analyses Reviewed Involving Visual Recognition Memory Tasks

Reviews	Fagan's studies	Rose's studies	Other studies	Total
Fagan & Singer (1983)	5	-	1	6
Bornstein & Sigman (1986)	4	1	1	6
Mc Call & Carriger (1993)	4	6	5	15
Slater (1995)	2	30	6	38
Total samples	15	37	13	65

Note: The studies conducted by Rose and her colleagues are Rose et al. (1988, 1989 a b, 1991 a b, 1992) Rose and Wallace (1985). The studies conducted by Fagan and his fellows are Fagan et al. (Fagan, 1984, Fagan & McGrath, 1981, Fagan & Singer, 1983). From 13 samples entitled "Other studies" eight samples used the Fagan test and other used experimental tasks other than the Fagan test. Other studies conducted by Colombo et al. (1989), DiLalla et al. (1990), Fulker et al. (1988), Gottfried, Guerin, & Bathurst (1989), Gottfried et al. (1989) and Yarrow et al. (1975). Those studies that used measures from habituation paradigm or visual anticipation are not included in this table.

A second point is that data in visual recognition memory are usually influenced by individual subjects. That is, excluding data for some subjects and/or partially overlapping subjects, influences the results and subsequently affects the predictive coefficients found. A close examination of studies in the review by McCall & Carriger (1993) revealed that five samples (out of six) from Rose et al. studies came from one original cohort recruited from a public hospital in New York City. Only one sample (Rose & Wallace, 1985) is, in fact, independent from the other samples. At times the samples partially overlap and sometimes

the subjects, from whom different types of recognition tasks, or different ages at the initial testing and/or different ages at subsequent testing sessions have been reported, are almost identical. The same criticism can be applied to seven studies reviewed by Slater (1995) because 30 correlation coefficients out of 38 were drawn from one original cohort. Consequently, data across the majority of samples overlap.

This second point leads to a third comment. The number of independent studies recruiting independent samples is not as many as the reviews have suggested. That is, the samples in most studies for which predictive values are available have been the same, or partially overlapping samples, drawn from one original cohort for whom different reports were provided. For example, data for all publications from Rose's laboratory in which prediction was assessed were drawn mainly from two cohorts (one cohort described in Rose, 1980 and the other one partially described in Rose et al., 1988a). Similarly, some publications of the Fagan studies reported data for the same, or at times overlapping, subjects and/or combined data for some groups of samples. (For example, there is partial overlapping or different combinations of subjects recruited for the Fagan et al., 1985, 1986 studies).

The behaviour of subjects during test taking is another point worthy of discussion. Generally, in studies employing habituation and visual recognition memory procedures, data for some subjects were missing, mainly due to uncooperative behaviour and fussiness and drowsiness. For instance, 136 out of 362 subjects were excluded from the study by Laucht et al. (1994) due to crying, restlessness and falling sleep. In a survey conducted by Wachs and Smitherman (1985) it was found that, on average, the attrition rate for subjects in infancy research was 29%. The point to be made here is that if a large amount of data is lost, the predictability of the task for the remainder of the subjects becomes questionable.

There is some evidence that subject loss in habituation studies is the results of nonrandom individual differences factors and may make the results of studies biased (Wachs & Smitherman, 1985). It is therefore possible that the predictability of the novelty task is contaminated, to some extent, by a relatively select group of subjects who were testable.

Finally, the predictability of future IQ from novelty tests is generally higher for at-risk infants and infants having a known syndrome associated with intellectual disability than is the case for nonrisk babies (e.g., McCall & Carriger, 1993). The point to be made here is that many subjects in the reviews of recognition memory studies were at-risk infants. For example, all subjects in Rose's studies were at-risk preterm infants or full-term infants from predominantly lower class families, where the potential risk is documented. This raises the possibility of obtaining a higher correlations due to the inclusion of this population in samples. An examination of previous literature therefore clearly suggests the necessity of attempting to replicate the reported results with an independent sample from the normal infant population.

3-5 UNDERLYING MECHANISMS

Determination of the degree of continuity and discontinuity for intelligence from early ages to later childhood has long been a controversial topic in developmental psychology. A solution to this problem, if not impossible (Sternberg & Okagaki, 1989), is at least very difficult. Starting with the first longitudinal study by Bayley in the early 1930s, over six decades of research in relation to the continuity of development has shown that there is no convincing answer, at least in relation to cognitive development. For instance, it is held by Sternberg and Okagaki (1989) that the continuity versus discontinuity debate is largely misconceived because both aspects are involved in human development. Similarly

Pomerleau (1989) and Lewis (1989) argue that the nature of development in humans probably involves both continuity and discontinuity.

Theoretically, the correlational continuity between novelty tasks and later intelligence tests, if substantiated, raises questions concerning the potential mechanisms that underpin this continuity. Some theoretical explanations and several models of continuity have been proposed to clarify the nature of the process or processes that might mediate the relationship between visual fixation in infancy and later cognitive skills in childhood. Although some empirical evidence has been advanced to support this theory, the actual processes involved are still unclear (Rose et al., 1989).

Bornstein and Sigman (1986) have proposed three models for continuity of intelligence. In one model of continuity it is assumed that continuity between recognition memory in infancy and intelligence in childhood has been mediated by the same specific processes that underlie both measures. These authors argued that information processing required for habituation and novelty preference involves processes of encoding, abstraction, categorization, and retention and retrieval of information, and thus mental representation is a common phenomenon in both procedures. In the same way, psychometric intelligence tests, particularly verbal tests based on language acquisition, also rely on mental representation. Accordingly, mental representation should underpin the longitudinal correspondence observed between infancy and childhood tests.

Confirming with Bornstein and Sigman's (1986) view, several investigators have suggested that the speed with which an infant processes information and forms an engram from a familiar stimulus mediates the continuity of mental development across the life span (e.g., Colombo & Mitchell, 1988; Fagan, 1984a; Rose et al., 1991a). Jacobson et al. (1992) have found a significant correlation between fixation duration and reaction time in infants as

young as 6 months old, suggesting that infants who look for a shorter time are faster and are more efficient at processing information than those who show a lengthier fixation duration to a stimulus. In a similar vein, Colombo et al. (1988) reported a relationship between short looking and faster processing in infants aged from 4 to 7 months (see review by Colombo & Mitchell, 1990).

There is also considerable evidence in the literature indicating a relationship between different measures of speediness and intelligence in adults and children (e.g., Eysenck, 1987; Vernon, 1983). For example, Anderson (1992) has hypothesized that individual differences in processing speed underlie the continuity of intelligence. Deary (1988) has also claimed that encoding speed is basic to *g*, accounting for somewhere between 25% and 50% of its variance. The significant negative correlation between measures of inspection time and intelligence reported for older children and adults (Kirby & Nettelbeck, 1991; Nettelbeck & Wilson, 1985; see also review by Nettelbeck, 1987) and for 6-year-old children (Anderson, 1986; Nettelbeck & Young, 1989) endorses some degree of relationship between speediness and IQ.

McCall accepts that rate of information processing may mediate the observed correlation, but has suggested that it could not be the only mechanism underlying mental development. On the contrary, he has proposed inhibition as a theoretical construct underlying prediction (McCall, 1994). McCall (p. 113 and 114) suggested that an infant's ability to "... inhibit responding to less salient stimuli, including ones made lower in salience by familiarization, is a *single* mechanism that can explain the prediction phenomenon from *both* paradigms [habituation and visual recognition memory]". According to McCall, both explanatory themes, inhibition and the rate of information processing, are stable threads in the prediction phenomenon.

In another model proposed by Bornstein and Sigman (1986), motivation or self-regulation has been suggested as a construct mediating prediction. Berg and Sternberg (1985) and Bornstein and Sigman (1986) hypothesized that motivation and state of arousal is an integral component of intelligence. Infants who have strong motivation or who are highly self-regulated have more ability to sustain their attention on a task, to persist in doing the task and can direct or inhibit their looking towards or away from the stimulus. They may also turn away quickly from a constant stimulus. These infants may, in childhood, have greater motivation to explore the new environmental context in the test situation, and a greater capacity to sustain their attention on intelligence tests, consequently scoring higher.

Similarly, Cohen (1988a) suggested that habituation and visual recognition memory tasks may tap individual differences in temperament, with more motivated and more curious infants scoring higher than less motivated infants. There is some indication that motivation mediates the correlational continuity in intelligence from infancy to childhood. For example, Sigman et al (1986) found that fixation measures in term (newborns) and 4-month-old infants related to WISC-R scores at 8 years of age. In analysing their data, they concluded that individual differences in an infant's state of arousal constituted one of the important factors in the prediction of later childhood IQ from fixation measures. It is possible, therefore, that measures of attention during infancy and cognitive skills in childhood are mediated by motivational aspect of intelligence.

3-6 THE FAGAN TEST OF INFANT INTELLIGENCE

The assumption that visual recognition memory during infancy has concurrent validity as an index of intellectual behaviours, led Joseph Fagan to develop a rationale for a new measure of infant intelligence that was grounded more on cognitive tasks than infant

developmental tests. The construction of the test was based on the theory that viewed intelligence as processing of information (see, above, the section concerning Fagan's theory of intelligence). The test was developed in various stages and has been updated over the last 15 years. Consequently several different versions of test have been used in different studies. The first data for construction of the test were collected in the early to mid 1970s (Fagan, 1971, 1973, 1976, 1977). The earlier version of the test comprised 12 novelty problems, from which only two to four items were administered at each age (Fagan & Singer, 1983). Fagan and McGrath (1981), in their first longitudinal study, initially tested their subjects ($N = 93$) between the ages of 4 to 7 months and administered subsequent assessments (i.e., the Peabody Picture Vocabulary Test [PPVT], the vocabulary factor of the Stanford-Binet test and/or the WISC-R) at 3.8 years (younger group) and 7.5 years (older group) of age, to test predictability from the novelty task. They reported that novelty scores were significantly related to IQ and verbal tests ($r = .37$ and $r = .57$, $p < .005$) at later ages. In another study, Fagan (1981, cited by Fagan and Singer, 1983) reported a significant correlation of $r = .36$ between the novelty preference test at 7 months of age and later PPVT and Stanford-Binet tests administered at 3 years, using a sample of 52 white and 16 black infants. Similar results have been reported in a study of infants who were tested initially at 7 months and retested at 3 and 5 years of age (Fagan, 1984b). An interesting finding of this study was that novelty preference scores obtained in infancy were more highly associated with later IQ as measured by PPVT ($r = .42$, $p < .01$, $N = 36$) than with later visual recognition performance ($r = .23$, NS) at 5 years of age. This led the author to suggest that the association between novelty preference and later IQ is mediated by more general processes than a simple recognition factor.

Later studies, designed with the intention of constructing a screening device for the

early detection of intellectual deficits, also examined the efficiency of the novelty score for this purpose (Fagan & Montie, 1986; Fagan, Shepherd, & Montie 1987; Fagan et al., 1985, 1986). The population for normative data for the screening device was a total number of 64 infants, suspected to be at risk for subsequent intellectual disability because of prematurity, maternal diabetic, thyroid disorders, growth failure and evidence in some cases of CNS damage. The sample was tested initially at one or more of four ages (12, 16, 22 and 29 weeks of age). The subsequent abilities tests were administered at 3 years of age using the Stanford-Binet and the PPVT. Novelty preference scores equal to or less than 53% detected 6 out of 8 children who had IQs equal to or less than 70, indicating that the device has 75% sensitivity in predicting delayed cognitive development. Novelty preference also had a value of 91% specificity since it correctly identified 49 out of 54 of the normal children (Fagan et al., 1986).

3-6-1 Standardization and statistical information on the FTII (marketed version)

The latest version of the Fagan Test of Infant Intelligence (FTII; Fagan & Shepherd, 1991) is now commercially available. The test has been standardized on 149 middle-class infants, the majority of whom were full-term and at low risk for later retardation (five infants were at risk for intellectual deficit). The test was administered to the infants between two and four times - at 27, 29, 39 and 52 weeks of age - and corrected for prematurity if they were born before 40 weeks of gestation age. The mean percentage of novelty scores is reported as 59% ($SD = 4.5$). At 3 years of age, the subjects were retested and were given the Stanford-Binet Test and the Peabody Picture Vocabulary Test (PPVT, Dunn & Dunn, 1981). The mean IQ score combining the Stanford-Binet and the PPVT for the sample was 108 ($SD = 12.1$). Mean novelty scores and mean IQ scores correlated .36 ($df = 148, p < 0.0001$). As a screening tool, the FTII is reported to be 100% sensitive to the detection of later

retardation, and highly specific in identifying normal infants (91.8%), giving a false positive rate of approximately 8 infants in 100. That is, 8 normal infants in 100 were labelled retarded (Fagan & Detterman, 1992). Internal consistency of the task was as low as .28 for a single test and as high as .48 for multiple administrations of the task (a composite of four ages). The inter-observer reliability of test is .94. No other information regarding the standardization sample and test-retest reliability is available for the marketed version.

Having a reliable screening tool would make it possible for clinicians to detect potentially retarded infants who do not show any obvious sign of abnormality at an earlier age. This would therefore constitute an important breakthrough because advances in medical care have made it possible for a greater number of at-risk infants to survive. As a research tool, a reliable novelty task could help scientists to examine the aetiology of retardation in very young infants.

3-7 DEVELOPMENT OF RESEARCH PROGRAM

Fagan argued that early novelty tasks “are valid constructively, concurrently, and predictively as an index of intelligence” (Fagan, 1988b, p. 261). However, the findings indicating reliability and predictive validity for the FTII are not consistent across different studies and different ages. One source of confusion may be related to the fact that most studies in which the reliability and predictive validity of the test was assessed were based on versions of the Fagan test that were different from the marketed version. The number of novelty items, the nature of stimuli used, the age and demographic characteristics of subjects were different from one study to another.

Although it is generally assumed that infancy tests - particularly visual recognition memory tasks - are not biased by culture, birth order, sex, or parental socioeconomic status

(Bornstein, 1988, 1989a; Fagan 1990), it appears likely that they are affected by an infant's temperament (Ruff, 1990). It has now been accepted that each baby has a different personality and temperament (Peterson 1996; Sanson, Prior, & Oberklaid, 1985). For instance, in one study conducted by Sanson et al. (1985) it was found that Australian babies, compared with American counterparts, were less active and could be categorized as easy babies who are normally cheerful and show happiness and positive responsiveness. However, Australian children are less adaptable to change in the environment and more arrhythmic in daily life. Even within a supposedly relatively uniform culture like America, it has been found that the Chinese-American newborns have a tendency to habituate more readily than the European-American infants (Freedman & Freedman, 1969). Therefore the validation of the Fagan test in an Australian sample would be useful. Furthermore, as reviews of the literature suggest, the number of independent studies providing some evidence for predictability of the Fagan test is not substantial. In fact, little research has been done on the marketed version of the Fagan test.

3-7-1 Research questions

This thesis has included four studies and two secondary investigations. What follows here is a brief account of the rationale on which the four studies were completed.

(i) *Study 1*: Findings concerning the test-retest reliability and long-term stability of the measures are mixed. For the Fagan test there is no study by other than the test's author in which internal consistencies for all four age-related tests have been estimated. Since the size of the predictive coefficient between infancy measures and childhood intelligence tests depends on the size of the reliability coefficient of both measures, the reliability of the infant test becomes a critical factor in studies of individual differences. Possession of 78 novelty reports made it possible to analyse some aspects of novelty problems in the Fagan test (item

analysis). The assumption of this study was that, if the basic concepts underlying the Fagan test remained the same at different points in time, then the test should show a reasonable internal consistency among items constituting the test and also reasonable test-retest reliability.

(ii) *Study 2*: The issue of predictive validity for infant perceptual/cognitive measures has eclipsed the concurrent validity of such tests. Some studies have attempted to assess the concurrent validity of visual recognition memory in infants who were at risk for intellectual disability or had Down's Syndrome or genetic form of retardation (e.g., Cohen 1981; McDonough, 1988). However, few studies have examined the relationship between novelty scores and current infant competencies. On the basis of the forgoing, it is possible that preference for a novel object as an index of infant intelligence should show a reasonable relationship with the MDI score from the Bayley Scales and, to a less extent, with the PDI.

The predictive validity of the information processing measures has also been reported as being low to moderate. Although it is claimed that, operationally, the value of predictability approaches the maximum possible validity (McCall & Carriger 1993), it is evident that the predictive validity is affected by many factors such as the type of stimuli, the number of items, the procedure on which testing is based, the sample size, the combination of samples (e.g., involving outlier infants who suffer from potential intellectual disability signs) and the subject's age at the first test (predictor) and at the second test (criterion).

The ages of the infant at the predictor test and the later test in childhood seem to be critical. Most published studies have recruited subjects younger than 8 months of age, starting as early as immediately following birth (e.g. Sigman et al., 1986), and up to 12 months of age (e.g. Rose et al., 1992). Due to the use of different sample sizes, the variety

of stimuli, and the different procedures, it is difficult from existing data to assess the most appropriate age at which the predictor test should be given. The use of a standard test in infancy, for a sample consisting of normal infants, may reduce the discrepancy resulting from the other variables that affect the predictability of tests. Having used the FTII as a predictor at four different ages (27, 29, 39, 52 weeks of age), Study 2 attempted to establish the first age in infancy on which the prediction could be based.

The FTII attempts to measure an infant's ability to differentiate between a new and a previously seen stimulus (novelty preference). However, it is possible that other activities during the course of information processing can also measure an infant's level of cognitive skills and consequently be related to later mental outcomes. One such measure is exposure time, the amount of time for which an infant is exposed to a target in order to accumulate the familiarization time. There is some evidence that exposure time is developmentally stable (e.g., Rose & Feldman, 1987), and appears more often as a significant predictor of later IQ than the duration of focused attention (Ruff, Lawson, Parrinello, & Weissberg, 1990). This evidence indicates that exposure time could be a reliable attentional measure of infant intelligence and may have some degree of predictive validity for later outcome.

(iii) Study 3 and Study 4: One conclusion endorsed by the reviewers of the literature on information processing during early infancy was that the predictability of information processing tasks was higher for infants whose initial test was given between the ages of 4 and 7 months (e.g., Slater, 1995). This suggests that there is an age-related “window of opportunity” for this type of measure to be useful. Lecuyer (1988) suggests that 5 months of age may be the key age for which most correlations between habituation measures and later cognitive development are available.

Study 3 and Study 4 involved a newly developed visual recognition memory task

based on the Fagan test. This was administered to a number of 15-17 month-old subjects, who were retested at 27 months of age. Two different samples were recruited and for each group a different familiarization time was used, the aim being to test the possibility of prediction of later IQ for older infants. The usefulness of an alternative scoring system for measuring the magnitude of preference for novelty was also assessed in Study 3.

(iv) *The predictive validity of the Bayley Scales:* The literature concerning the predictability of developmental scales clearly illustrates the contradictory results in the field. Considering the vast variability of data obtained from subjects, either during infancy or childhood, it appears that the low predictability reported for the developmental scales is a function of the age of the infants at the first testing. Undoubtedly, developmental scales in early infancy have low predictive validity to later IQ. However, it seems that the appropriate point in time at which the prediction of later IQ becomes possible is earlier than the age suggested by the literature of the field (i.e., 18 to 24 months). Therefore, another aim of the present project was to compare the results obtained from measures of attention with data collected simultaneously by means of conventional developmental scales, to establish the first age at which prediction of later IQ becomes possible. Similarly, the potential value of socioeconomic status as a predictor of later mental functioning was assessed, using such measures as parental education, parental occupation and maternal IQ.

(v) *Rising MDI in young children:* The Bayley Scales (Bayley, 1969) is one of the most widely used tests in research and clinical appraisal. However, the normative data from this test are dated and may not be applicable to infants from the present generation. On the other hand, given that the nature of the developmental scales, particularly in the first year of age, is heavily based on gross and fine motor activities, it is not immediately obvious that normative data need to be revised. Nevertheless, during the last few decades there has

been rapid economic growth and considerable changes in health care and life style in most cultures which have had a great influence on babies, and perhaps more advanced developmental milestones could be expected. In fact, Campbell, Siegel, Parr and Ramey (1986) have reported that the Bayley Scales (Bayley, 1969) overvalued their sample's performance and suggest that the normative data could be improved. If this is true, investigators will be confronted with problems resulting from using an outdated test.

Recently, a new version of the Bayley Scales has been released (Bayley, 1993) and is available to researchers. This test has not yet been validated on an Australian sample. It would be a valuable contribution to the Australian data to provide validation for the new Bayley Scales and to verify if the old version of the Bayley Scales is still accurate. This is of particular importance for studies using the unrevised version of the Bayley Scales in longitudinal designs.

3-7-2 Pilot study

In order for the examiner to become familiar with all aspects of testing and the tests, all measures taken during infancy and subsequent early childhood tests were piloted on a nonstudy sample. First, the Fagan test was administered to 16 infants aged between 6 and 12 months. The administration of the Fagan test was also practiced using the video tape provided as a part of the manual for the test (Fagan & Shepherd, 1991). Based on the 18 training tests of the training certification film, the inter-observer reliability for the present study was $r = .94$ ($SD = .06$). The new novelty task developed for Study 3 and 4 was also trialled on five subjects whose ages ranged from 15 to 21 months.

The Bayley Scales were piloted on two different age groups. The first group ($N = 8$) consisted of subjects aged between 5 and 14 months. The second group consisted of 6 infants with an age range of 17 to 30 months. Both mental and motor scales were

administered to the first group, but only the mental scale was administered to the majority of the second (All subjects were normal infants and two of them were twins).

In addition, an abbreviated form of the Stanford-Binet (including Vocabulary, Comprehension, Short-Term Memory, Pattern Analysis, Copying, Absurdities and Bead Memory) was administered to six children aged between 26 and 40 months (three subjects were part of the group to whom the Bayley Scales were administered). The author was already familiar with the administration of the Stanford-Binet Intelligence Scale, Form L-M for school age children.

While there was no difficulty in administering the Bayley Scales, the author had some difficulties with the administration of the FTII. The first testing sessions took a relatively longer time than the normal time necessary for testing and some infants were not cooperative towards the end of the testing session due to crying or restlessness (perhaps because of the longer time needed to administer the test). During the pilot testing, the infants sat on their mothers' laps but, due to using an ordinary chair (instead of a higher or adjustable chair), observation of the reflection of stimuli in the infant's eyes was difficult. In addition, there was some initial difficulty in providing sufficient light. These problems were rectified after the first few testing sessions.

3-8 AIMS OF THE CURRENT RESEARCH PROJECT

On the basis of reviewed literature the major aims of the present project were:

- (i) To examine the psychometric properties of fixation duration measures as indexed by the FTII and a newly developed novelty task designed by the author for children older than 12 months;
- (ii) to validate the predictive value of the FTII on an Australian infant population;

(iii) to examine the possibility of extension of the novelty preference procedure in predicting later intellectual outcome in older infants;

(iv) to establish a new scoring procedure for measuring preference for novelty to enhance the internal consistency of the novelty task;

(v) to investigate the usefulness of applying an expiring familiarization time in testing older infants with a novelty preference task;

(vi) to contrast the predictive relationship of information processing tasks with those of infant developmental tests and socioeconomic factors;

(vii) to determine the first age at which the prediction of later IQ becomes possible;

(viii) to provide Australian data on the reliability and validity of the recently revised version of the Bayley Scales;

(ix) to investigate the possibility of an MDI gain in an Australian infant population.

CHAPTER 4

THE RELIABILITY AND STABILITY OF INFANT VISUAL ATTENTION MEASURES FROM 6 TO 12 MONTHS OF AGE (STUDY 1)

4-1 INTRODUCTION

Individual differences in response to novelty during infancy seems to be associated with intelligence later in life (Slater, 1995). This measure is based primarily on an infant's disposition to attend and process information and, as Fagan (1992) has argued, it provides a convenient measuring system for assessing an infant's mental faculty. A major criticism in this area is related to the psychometric properties of the new measures for internal consistency and test-retest reliability. The reliability of an infant test limits the maximum predictive correlations.

Results from studies examining the test-retest reliability and short term stability of infant information processing tests are mixed. In their review, McCall and Carriger (1993) reported that the short-term test-retest reliability for habituation and visual recognition memory are typically between .30 and .45. Overall, the sizes of the test-retest, internal reliability and long term stability correlation coefficients are low to moderate, depending on the number of stimuli, inter-test interval, the method under which the test is administered and the infant's age (Andersson, 1996; Benson et al., 1993; Colombo et al., 1988; see sections 3-3-3 and 3-4-2). For example, Rits, Woodruff, & Fagenet al. (1984) reported that while the number of trials to criterion in habituation studies in four-month-old babies was

significantly reliable within a session ($r = .48, p < .05$), no correlation was found between the original measure and a second measure one week later. Combined, these findings confirm Slater's conclusion (Slater, 1995, p. 80) that "Across-age reliability, for all measures, is usually lower, and the longer the age gap, the lower the correlations; with age gaps longer than 2-3 months the r s often approach zero".

Although there have been a number of studies examining the habituation procedure, there are few studies which have examined the reliability and stability of visual recognition memory tasks. One study, Mundy, Seibert, Hogan and Fagan (1983), assessed the consistency of the mean novelty score of 14 developmentally-delayed children aged from 8 to 39 months of age, using the paired comparison procedure. They administered four novelty problems to their subjects twice across a six month assessment interval and found that the cross-age correlation of the average of four novelty scores was moderate ($r = .58, p < .05$), but the correlations obtained between each of the four individual novelty problems were low and nonsignificant.

Two studies from Rose's et al. laboratory have also investigated the reliability and stability of infant information processing measures. In the first study, the test-retest reliability of novelty preference and exposure time scores were examined in infants tested at 6, 7, and 8 months of age using three abstract patterns and three achromatic photographs of human faces as novelty problems (Rose & Feldman, 1987). The stability of the percentage of preference for novelty for each individual problem was highly variable, with some negative correlations and a range between $r = -.18$ and $r = .44$ (average $r = .18$). However, when the novelty scores were averaged across all six novelty problems, relatively higher correlations (ranged from $r = .30$ to $r = .51$) between the three different periods (i.e., 6 and 7, 6 and 8 and 7 and 8 months of age) were obtained. Rose and Feldman (1987, p. 495) also

reported a similar outcome for the exposure time scores, and concluded that "... like the novelty scores, this aspect of visual behaviour shows moderate stability, at least over this age period [6 to 8 months]". The second study assessed the internal consistency and the stability of a composite of 11 novelty problems measuring visual recognition memory and cross-modal transfer (tactual-to-visual) memory in 6, 7 and 8 month pre-term and full-term infants (Rose et al., 1988a). Rose et al. found that the mean of the test-retest coefficients of the novelty scores for individual problems across periods of one and two months was low ($r = .18$) but the scores for composites of the problems were moderately stable, ranging from $r = .20$ to $r = .51$. The internal consistency between measures as estimated by Cronbach's alpha was uniformly low, and the median for all scores was negative ($\alpha = -.10$).

Similarly, using the paired comparison procedure, Colombo et al. (1988) investigated the week-to-week test-retest reliability and month-to-month stability of looking behaviour in 4 and 7 month-old infants, who were tested with letters of the alphabet as stimuli. They reported that the internal consistency of novelty preference measure was extremely variable and low. The week to week reliability, the stability correlations over three months for novelty preference scores and some variables, such as the number of fixations during familiarisation and the proportion of time off stimulus, were low to moderate, but statistically significant.

4-1-1 Internal consistency and test-retest reliability of the FTII

All of the above studies were experimental in nature, and the number of novelty problems, the infants' ages and the type of stimuli and procedure varied from one study to the next, thus making a direct comparison difficult. The test manual (Fagan & Shepherd, 1991) and the technical review (Fagan & Detterman, 1992) for the FTII offers little

psychometric information against which test results can be assessed¹. The few studies utilising some tests of the FTII at different ages have resulted in contradictory findings.

An early study by Fagan and McGrath (1981) attempted to estimate the Spearman-Brown split-half reliability of novelty items given to infants aged 4 to 7 months. They reported split-half correlation coefficients of $r = .41$ to $r = .70$ for novelty preference at different ages. Using the same statistical procedure Fagan et al. (1987) reported a reliability of $r = .58$ for another version of the Fagan test. Thompson et al. (1991) have also found a reliability of $r = .40$ for ten novelty problems in the Fagan test, given at 5 and 7 months of age. However, reliability based on splitting the scale into two parts suffers from the allocation of items to the two halves and can result in different correlation coefficients.

Two studies have endeavoured to predict later intelligence in pairs of twins utilizing the FTII as an infant cognitive measure. In the first study (DiLalla et al., 1990), two versions² of the FTII (the Immediate Version and the Re-test Version approximately 30 minutes later) were given at 7 months and again at 9 months of age. The within session test-retest reliability between the Immediate and Retest Versions was relatively low, with correlation coefficients of $r = .15$ at 7 months and $r = .35$ at 9 months. Moreover, after two months, neither Immediate Version nor Retest Version administered at 7 months of age showed test-retest reliability with 9 months FTII. Using the same design, a second study attempted to replicate the findings of DiLalla's study with a sample of 114 pairs of twins (Benson et al., 1993). At 5 months of age³ the subjects were given one appropriate age-related test of the Fagan test. At 7 and 9 months of age two versions of the FTII (Immediate

¹ In what follows the commercial form of the Fagan Test of Infant Intelligence (Fagan & Shepherd, 1986, 1991) is referred to as the FTII; earlier versions are termed "the Fagan test".

² The novelty preference scores in this study were based on two tests at each age. The first test (Immediate Version) was a complete test of the FTII administered at appropriate age (i.e., 27 and/or 39 weeks) and the second test (Retest Version) was the same test administered at the end of the test session with the same procedure except for familiarisation time which was cut in half.

³ The 5-month Fagan test has not yet been marketed.

Version and Retest Version) were administered to the same subjects. Contrary to the first study, the mean of the novelty preference scores obtained at 7 months from a Retest Version of the FTII correlated significantly ($r = .35$) with the novelty preference scores obtained at 9 months. However, consistent with DiLalla et al. (1990), the correlations for the Immediate Version of the FTII across all ages were found to be negative, indicating no long-term stability of the FTII over 5 to 7, 5 to 9 and 7 to 9 months of age. These contradictory findings for the FTII led the investigators to surmise that "... the low test-retest reliability may indicate that the FTII is measuring different cognitive processes or skills at 5, 7, and 9 months" (Benson et al., 1993, p. 444).

Jacobson and her colleagues (Jacobson et al., 1992) have also used the 27 and 52 weeks tests of the FTII, and found that the novelty scores derived from the FTII between 27 and 52 weeks of age were unstable ($r = .04$). The internal consistency of the novelty preference scores was also very low (Cronbach's $\alpha = .11$ at 6½ months and $\alpha = .01$ at 12 months). The latest publication available for the FTII (Andersson, 1996), also presented evidence of a very low internal consistency for this test. He reported that the internal consistency (Cronbach's α) for the FTII ranged from $-.07$ to $.40$ depending on age, with older infants scoring higher than younger ones.

Taken together, the inconsistency of results regarding the reliability of the FTII suggests that, in addition to difficulties with infants' behaviour and instability resulting from developmental changes, the FTII may have procedural problems or some restrictions on the selection of the stimuli. It is unfortunate that no independent study has item analyzed the FTII. If the test is widely used, it is critical that a comprehensive review of the test be done (Benasich & Bejar, 1992).

First, the unreliability of a task may rest in the selection of novelty items; i.e., whether all items that constitute the test result reliably in above chance level performance. If all items are similar in attractiveness and the complexity of stimuli are adjusted with a preset familiarisation time for different age groups, all items should exceed chance level and discriminate at about the same level of preference. Further, since internal consistency is based on the intercorrelation of items within the test, some items may not be homogenous and, in fact, may be measuring different entities rather than a common entity called “novelty preference”. Inspection of the relationship between individual items and the composite score will make it possible to identify such unreliable items, and the identification and elimination of such items may, therefore, result in an increase in the internal consistency of the test. It would be desirable to determine whether such individual items can be identified.

To date, only one study has addressed the internal consistency of novelty scores obtained from all tests of the FTII (Fagan & Detterman, 1992). They reported that internal consistency of the FTII was as low as $\alpha = .28$ for a single test and as high as $\alpha = .48$ for multiple administrations of the task. Although measures derived from this type of test may produce low to moderate reliability and long-term stability, as yet the psychometric properties of the novelty preference scores obtained from the FTII and its developmental trends have not been identified.

4-1-2 The reliability of other attentional measures

Although the novelty preference score is the most researched measure obtained within the paired-comparison paradigm, other measures of attention and inattention derived from the familiarisation phase may also bear some relationship to an infant's concurrent mental capacities and/or prediction of later intelligence. Two measures of infant looking behaviour are of particular interest - exposure time and the number of fixations an infant

makes to familiarize her/himself with “standard” targets. There is some evidence that these measures are more internally consistent than the novelty preference score (e.g., Jacobson et al., 1992) and are stable in both the short and longer terms (e.g., Colombo et al., 1988; Rose & Feldman, 1987).

Exposure time is the time that an infant takes to accumulate the predetermined familiarization time. Exposure time is shorter when infants spend less time looking away and concentrate more on the task. Consequently, exposure time may reflect individual differences in an infant’s ability to sustain attention (Rose & Feldman, 1987), or may reflect the infant’s tendency toward exploration of the visual environment (Colombo et al., 1988). The test-retest reliability and developmental trends in exposure time as indexed by the FTII are still unknown. Theoretically, as older infants are more self-regulated than younger infants, this measure would be expected to decrease in older infants due to a more focused approach to the task.

Finally, if it is accepted that infant intelligence is measured as response to novel stimuli (Fagan, 1984a, 1992), then there should be intercorrelations between all, or at least some, measures derived from tasks of visual recognition memory. In fact, Colombo found such a relationship between some infant measures of attention and inattention at 4 and 7 months of age (Colombo et al., 1988). They found that fixation duration and novelty scores correlated significantly at both ages, an indication that infants with shorter duration of fixation scored a higher percentage of novelty preference.

In the present study all subjects were tested once at an age consistent with the FTII age range, and the same test was administered to approximately half the subjects twice at two consecutive ages. This design made it possible to assess the internal consistency, test-retest reliability and relatively longer term stability of all measures derived from the FTII.

Overall, the main aims of the present study were: (i) to investigate the psychometric characteristics of novelty scores, the number of fixations and exposure time; (ii) to analyse the novelty items in the FTII for different age groups; (iii) to detect any intercorrelation between different attentional measures and (iv) to test the stability and developmental change of infant attention and inattention measures, derived from a paired comparison procedure over the second six months of age.

4-2 METHOD

4-2-1 Subjects

Initially, 86 (42 females, 44 males) participants were recruited. The sample were all white healthy infants and allocated to one of four different age groups corresponding to the appropriate postconceptional age of the four tests of the FTII (i.e., 67, 69, 79 and 92 weeks. A full pregnancy is 40 weeks, so that these correspond to 27, 29, 39 and 52 weeks following full-term delivery). The criteria met by all subjects were age, having no known physical or intellectual disability, and normal vision as indicated by birth records. Eight subjects were excluded from the final analysis because of crying and fussiness during attempted testing ($n = 6$) and scheduling difficulty or illness at the test time ($n = 2$). There were, therefore, 78 infants involved in the study. Those successfully completing the first testing session in each age group were: 18 subjects (6 girls, 12 boys) 27 weeks of age, 15 subjects (6 girls, 9 boys) 29 weeks of age, 22 subjects (12 girls, 10 boys) 39 weeks of age, and 23 subjects (12 girls, 11 boys) 52 weeks of age. Although the number of boys and girls in each group was unequal, the difference was not statistically significant.

A voluntary second testing session was held for a subgroup from the original 78. Forty infants were volunteered by their parents to participate on two different occasions but

three were lost because of fussiness and appointment difficulties. The data collected from this sample were intended to be used for the purpose of a test-retest reliability analysis. The numbers of subjects remeasured at 29, 39 and 52 weeks of age were 12, 12 and 13 respectively. In order to determine if the two samples (infants with one testing and subjects with two testing sessions) were different from each other in terms of demographic characteristics, a series of chi-square and *t* tests were computed. The two groups were similar in terms of gender, birth order and ethnic background (χ^2 ranged from .00 to 1.70, all *ps NS*) and in terms of parental age, parental education, socioeconomic status as well as gestation age and birthweight (*t* ranged from 1.21 to .94, all *ps NS*).

All subjects lived within a radius of 30 kilometres from the city centre and were recruited from different sources, mainly by advertising in a free locally distributed newspaper, several schools' newsletters and the notice boards at various child care centres. Each family was provided with a background information sheet in which the aims of the project, details of the measures that would be used, and the procedures with which the infant would be tested, were clarified (see appendix B). Table 4-1 shows the birth status and demographic variables for the sample.

(i) **Birth status:** All but two infants were born in Adelaide, South Australia (two were born overseas), between August 1992 and August 1994 without any pre or perinatal complications. There was only one baby whose preterm labour was stopped by surgical treatment at 27 weeks of conceptional age. No subject had more than four weeks prematurity or less than a 2680 grams birthweight. Overall, 92.3% (*n* = 72) of the subjects had 37 weeks or longer gestation periods⁴. Within the confines of this definition, six babies with gestation period between 36 and 37 weeks were slightly premature.

⁴ Liveborn infants delivered before 37 weeks from the first day of the last menstrual period are conventionally termed premature (Nelson, 1995); in Nelson, W. E. (Senior Ed.), R.E. Behrman, R.M.

Table 4-1

Birth Status and Demographic Variables of Sample

	N	M	SD	Range
Variables				
Length of gestational age (weeks)	78	39.4	1.5	36 - 42
Birthweight (grams)	74	3600	537	2680 - 5020
Maternal age	76	30.4	4.5	19 - 41
Paternal age	72	33.1	5.5	21 - 49
Maternal education (years)	78	13.6	2.4	9 - 20
Paternal education (years)	78	14.0	3.0	9 - 21
Occupational status (7-point scale)	74	3.9	.96	2.15 - 6.0
	n			%
Gender				
Female	36			46.2
Male	42			53.8
Birth order				
First born	48			61.5
Second born	16			20.5
Others ^a	14			18.0
Ethnicity				
English speaking families	58			74.4
Non-English speaking families ^b	20			25.6

Note: ^a - 10 subjects were third born and four were fourth born babies.

^b - Iranian, Greek, Indian, Malaysian, Yugoslavian and Polish.

(ii) **Socioeconomic status:** Australia is a multi-cultural society in terms of ethnic background. For the majority of subjects parents were Australian (74.4%). The parents of the remaining subjects ($n = 20$) immigrated to Australia. The proportion of ethnic families was therefore similar to that in the current Australian population (i.e., approximately 25%; Prior, Garino, Sanson & Oberklaid, 1987). According to the 7 point Occupational Prestige Scale⁵ (Daniel, 1983), the average score of the parents' occupational status was 3.9

Kliegman, & A.M. Arvin (Eds). *Nelson Textbook of pediatrics* (15th. ed., p. 454). Philadelphia: W.B. Saunders.

⁵ The scale is an Australian rating scale developed by Daniel (1983) that measures the prestige of each occupation as an indicator of power and privilege in society. The assumption underlying this scale is that a specific value is integrated with each occupation. The scale has 7 points and assigns a rank to each job according to its social standing. A rank of 1 indicates a job of high social standing and a rank of 7 indicates a job of low social standing.

($SD = .96$). Social class was also defined in term of the level of parental educational achievement. The mean of education of both parents was 13.8 years of schooling ($SD = 2.2$). Fifteen subjects' mothers had less than 12 (i.e., 9 to 11) years of schooling and 37.3% of mothers ($n = 29$) graduated from universities. All parents were married (two divorced) and subjects were reared at home.

4-2-2 Measures

A battery of two different types of infant tests was given to subjects in each session. One was the FTII (Fagan & Shepherd, 1991), designed to measure an infant's preference for novelty. The second test was the Bayley Scales (Bayley, 1969), designed to assess an infant's current mental functioning and various motor activities. Both tests were performed using standard procedures and materials.

(i) *The Fagan Test of Infant Intelligence*: The FTII utilizes a measure of an infant's visual recognition memory as a test of infant information processing. The operational score of the test, the "novelty score", is the percentage of time the infant spends looking at novel stimuli in preference to familiar pictures. The marketed version of the FTII is composed of the same 10 novelty problems (novelty items) in each of four different ages (i.e., 27, 29, 39 and 52 weeks postnatal age corrected for prematurity⁶). All novelty items are pictures of human faces. The baby sits on the parent's lap facing the apparatus. The person administering the test sits behind a stage and records on a computer's keyboard the direction and the length of time for which the infant spends looking at the stimuli (see below, procedure). The examiner looks through a magnifying optical lens at the corneal reflection of a picture which is centred over the pupils of the infant's eyes. The pictures are secured on a revolving door and changed by the examiner after each trial. After the test is

⁶ Although there is some controversy about age correction for prematurity when scoring an infant development scale (Sattler, 1988), this procedure was followed here so as to confirm to requirements for the Fagan Test of Infant Intelligence (Fagan & Shepherd, 1991).

completed, a computer print out shows the novelty preference score for each trial separately and for all items as a whole. The number and total time of looking are also recorded and printed (Appendix C).

(ii) **Bayley Scales of Infant Development:** The Bayley Scales (Bayley, 1969) is an individual sensorimotor test designed to assess an infant's concurrent developmental process in terms of cognitive ability and sensorimotor activities. The Bayley Scales was selected because it is one of the most popular measures during infancy (Thompson, Wasserman & Matula, 1996). The test consists of three different scales, each measuring a specific domain of the infant's development (see Chapter 2). Converting the number of items passed by infants to normative data with a mean of 100 and a standard deviation of 16, has made it possible to compare an infant's mental and chronological age (Bayley, 1969).

4-2-3 Apparatus

The laboratory was a 3.5m by 3.5m room equipped with a National VHS colour portable video cassette recorder, type NV-180, in conjunction with a National colour video camera, type WVP-A2N with a wide angle lens, type 0 x .7 for recording testing sessions. However, ethical considerations stipulated that recording of the testing session be restricted to subjects ($n = 12$) whose parents gave permission for the session to be recorded. This recording was intended for monitoring the scoring, if necessary, of the Bayley Scales. The overhead lighting was sufficiently bright so that the reflection of the picture used in the FTII was easily seen in the pupils of the subjects' eyes. The apparatus for the FTII is fully described in the manual (Fagan & Shepherd, 1991). Briefly, the major components of the FTII are an IBM personal computer, a portable stage (Figure 4-1), a computerised testing and scoring system and 14 different pictures (one picture is presented in three different orientations) of babies', men's and women's faces. A high adjustable chair made it possible

for the baby's eyes to be level with the peephole centred in the revolving door on which pictures from the FTII were placed.

Figure 4-1

Portable Viewing Stage Used in the FTII



4-2-4 Procedure

The majority (82.2%) of subjects were tested at home, and the remainder were tested in the laboratory. No subject was tested until they were assessed as being alert, calm and cooperative (or at least in a normal mood as considered by caregivers). Each infant was tested within ± 6 days of reaching the target age for their age group. In order to avoid bias, the same testing procedure was used for all subjects. Infants were first tested on the FTII.

The test was administered in accordance with the standard procedure described in the manual (Fagan & Shepherd, 1991). All novelty problems are presented in fixed order. Each novelty problem consists of two stages. In the first stage, "familiarisation phase" or study time, the infant is presented with one or two identical pictures for a predetermined time. During the familiarisation time "standard" pictures remain exposed to the infant until he or she accumulates the preset time necessary for memorising the standard (familiar) picture. The second stage is a "novelty trial" or test phase that follows immediately after the familiarization period. The novelty trial consists of two trials. In the first trial the previously seen target (standard stimulus) is paired with a new picture (novel stimulus) for a specific time. In the second trial the left and right positions of each of the two pictures are reversed to prevent side preference. The test trial begins with the first fixation by the infant and lasts until the preset time expires. Each test consists of 29 familiarization and novelty trials (Figure 4-2). The number of items, the order of exposure to each specific picture and all pictures are the same across all four ages. The only procedural difference between the versions for the four ages at which testing is recommended is the duration of study time which decreases from 20 seconds for young, to 4 seconds for older infants (the familiarization time also varies for different novel problems within each age). The test time (novelty trials) was fixed at 10.2 seconds for 27 and 29 week-olds and 6.2 seconds for 39 and 52 week-olds. The inter-observer reliability for the present study was $r = .94$ (see pilot studies, section 3-7-2).

The dependent variables from this test used for further analyses, were the percentage of time that an infant spent looking at novel stimuli (the percentage of preference for novel stimulus), the time for which an infant was exposed to stimuli in order to accumulate the

Figure 4-2

Representation of 29 Trials of the FTII for Ten Novelty Items.

Novelty Item 1

Left

Right

1 - Familiarization Trial



2 - Novelty Trial No.1



3 - Novelty Trial No.2



Novelty Item 2

4 - Familiarization Trial



5 - Novelty Trial No.1



6 - Novelty Trial No.2



Novelty Item 3

7 - Familiarization Trial



8 - Novelty Trial No.1



9 - Novelty Trial No.2



Figure 4-2

Continued

Novelty Item 4

Left

Right

10 - Novelty Trial No.1



11 - Novelty Trial No.2



Novelty Item 5

12 - Novelty Trial No.1



13 - Novelty Trial No.2



Novelty Item 6

14 - Novelty Trial No.1



15 - Novelty Trial No.2



Novelty Item 7

16 - Familiarization Trial No.1



17 - Familiarization Trial No.2



18 - Novelty Trial No.1



19 - Novelty Trial No.2



Figure 4-2

Continued

Novelty Item 8

20 - Familiarization Trial No.1



21 - Familiarization Trial No.2



22 - Novelty Trial No.1



23 - Novelty Trial No.2



Novelty Item 9

24 - Novelty Trial No.1



25 - Novelty Trial No.2



Novelty Item 10

26 - Familiarization Trial No.1



27 - Familiarization Trial No.2



28 - Novelty Trial No.1



29 - Novelty Trial No.2



These pictures are reprinted with permission from J.F. Fagan (1996)

familiarization time (exposure time) and the number of glances an infant made during the familiarization trial (number of fixations).

For all subjects, the Bayley Scales were administered after the novelty task. The Bayley Scales testing procedure combines features of both the Binet and Wechsler scales, utilising the age ordering of the Binet and the sub-scale ordering of the Wechsler (Bayley, 1969). Test items for the Bayley Scales are arranged in order of difficulty and grouped by "Situation Codes". That is, a number of items in both the Mental and Motor Scales are arranged in sequence and use a similar stimulus situation or method of administration for different age levels. It is usual for an infant's performance to cover items having different age placements. After establishing the basal level, the items are presented in order of difficulty until the infant fails eight to ten consecutive items. The Bayley Scales consists of three scales. For the purpose of the present study only the Mental Developmental Index (MDI) and Psychomotor Developmental Index (PDI) were administered.

All infants who participated in two different age testing sessions were seen again at the age adjacent to the initial test and tested with both tests (FTII and Bayley Scales) appropriate for that specific age.

4-3 RESULTS

4-3-1 Descriptive statistic and preliminary analysis of the measures of interest

(i) *Novelty preference scores*: The percentage of preference for novelty for each individual novel item is averaged across all items (in this case, ten novel items), to provide a mean novelty preference score. The means and standard deviations of the percentage of novelty scores for all age groups in the initial testing session are shown in Table 4-2. The means for the novelty preference scores ranged from 62.4% to 64.7%.

Table 4-2

The Percentage of Preference for Ten Novelty Items and Means of Novelty Score in Each Age Group

Age	27 weeks (n = 18)		29 weeks (n = 15)		39 weeks (n = 22)		52 weeks (n = 23)	
Novelty Items	M	SD	M	SD	M	SD	M	SD
Novelty Item (1)	58.9	19.9	52.2	21.8	50.4	24.1	55.1	15.5
Novelty Item (2)	76.1***	15.7	72.8***	16.4	68.6***	21.5	69.3***	16.4
Novelty Item (3)	74.9***	14.1	72.9***	13.6	66.1***	20.3	69.2***	16.1
Novelty Item (4)	70.6***	15.3	68.2***	14.0	70.2***	20.4	67.6***	14.7
Novelty Item (5)	65.9**	21.7	73.9***	15.2	74.9***	14.6	69.2***	16.9
Novelty Item (6)	64.8**	18.5	60.9*	14.7	62.8***	15.8	62.9***	16.1
Novelty Item (7)	60.9*	19.0	54.5	16.2	65.7***	17.8	68.9***	16.3
Novelty Item (8)	59.3*	14.4	53.1	24.7	49.3	18.1	60.3*	23.7
Novelty Item (9)	45.7	25.0	55.5	20.1	44.0	18.4	52.5	19.3
Novelty Item (10)	69.5***	19.9	62.9	27.5	69.0***	14.2	70.3***	20.4
Mean novelty score	64.6***	6.4	62.7***	5.1	62.4***	6.2	64.5***	6.5

Note: * $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$. (two-tailed t test) indicating that novelty scores are significantly different from the chance performance.

Before performing further analyses on the novelty scores, the differences between means for males and females were examined. Statistical analyses indicated that there was no significant differences between the two genders for each age group separately (t ranged from -1.41 to .49, all ps NS). T tests were also computed to determine whether novelty preference scores exceeded the 50% chance level. All individual novelty scores (i.e. 40 items derived from the 27, 29, 39 and 52 week tests), as well as the means of the novelty preference scores at each age, then were tested against the hypothetical test value of 50. It is evident from the results in Table 4-2 that the means of novelty preference for all four age groups significantly exceeded chance. However, a more detailed examination of novel items shows that the mean novelty score for items 1 and 9 in all four tests and for item 8 in both the 29 and 39 weeks tests, did not significantly exceed chance level, indicating that the infants may have found these novelty problems difficult. The 29-week test had the highest number of items not exceeding chance since items number 7 and 10 were also among the items that did not exceed chance performance. The results of this study were, then, consistent with the results of Jacobson et al. (1992), who used only the 27 and 52 weeks FTII for black infants from lower class parents. They also found that items numbers 1, 8 and 9 did not significantly exceed chance level.

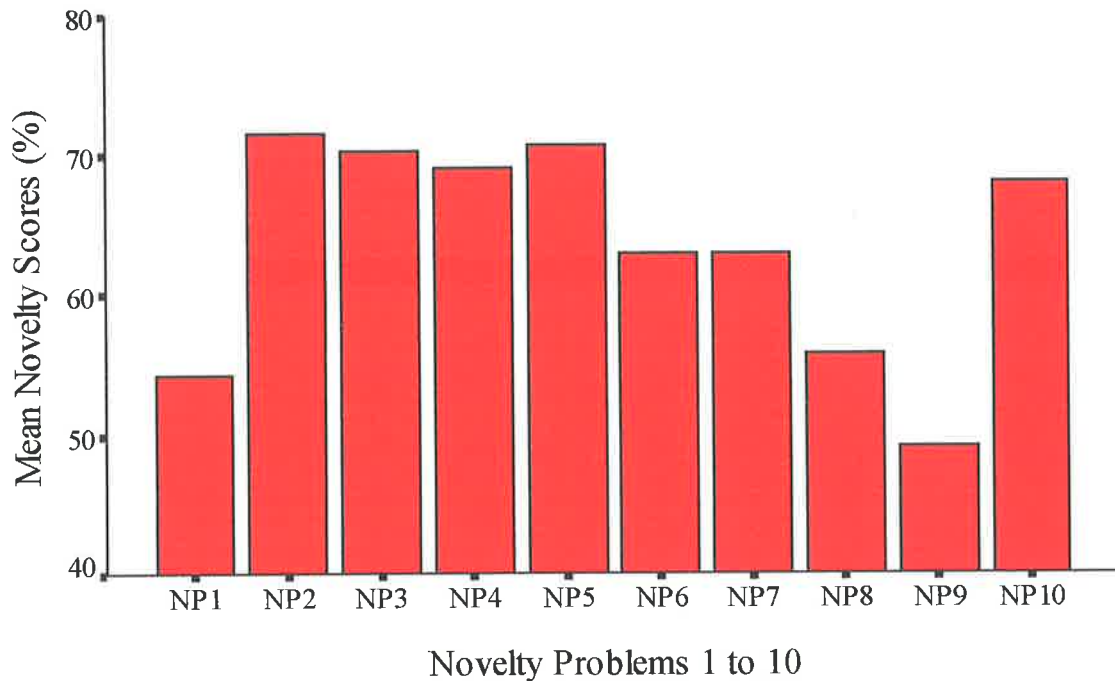
Figure 4-3 illustrates the fluctuation of novelty scores for subjects⁷ (age partialled). As can be seen from the figure, preference for novelty fluctuated across all items. The earlier items (with the exception of the first item) had higher novelty scores than the later items. The implication of these results will be discussed.

Since the normal distribution of measures is important in bivariate correlations and inferential statistics, the normality of the distribution of novelty preference scores was

⁷ The bar graph for each age group separately also indicated that the infants' performance showed fluctuation across all novelty items (see Appendix D).

Figure 4-3

Percentage of Preference for Novelty Across All 10 Novelty Problems (N = 78)



examined before further analyses were undertaken. The normality of the distribution of the novelty scores was assessed using normal probability plots and a Shapiro-Wilks test. The statistical test indicated that the four distributions of the novelty scores were all close to normal. Figure 4-4 shows a boxplot of the means of preference for novelty scores for all four age groups, and Figure 4-5 shows the normal probability plot of the novelty scores for the total sample. It is clear from the probability plot in Figure 4-5 that almost all the scores fall on a straight line, indicating that the observed novelty scores are paired with the expected value from the normal population.

Figure 4-4

Boxplot of Mean Novelty Preference Scores for Each Age Group

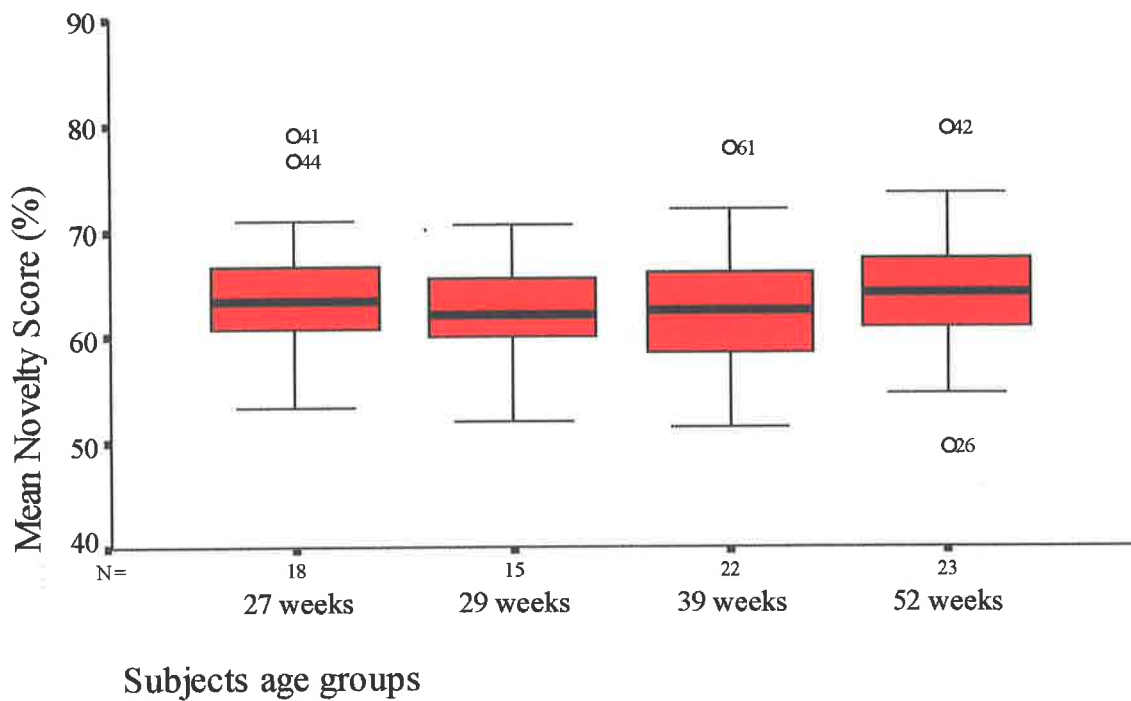
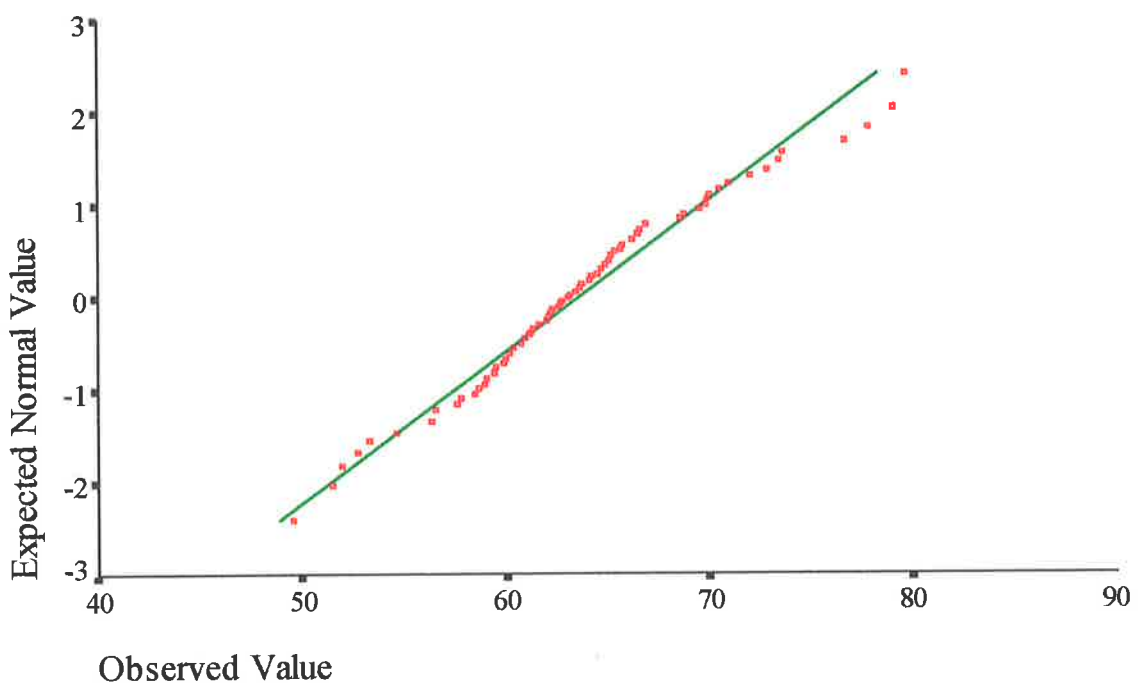


Figure 4-4 shows the box-and-whiskers-plot for the distribution of novelty scores. Although some age groups have one or two outliers, the distribution of novelty scores does not substantially depart from normality.

Figure 4-5

Normal Probability Plot of Mean Novelty Scores for Total Sample (N = 78)



Subsequently, a one way ANOVA was performed to compare differences between the four groups for novelty preference scores (see Appendix E). The resulting analysis showed no age differences between younger and older infants $F(3, 74) = .76, p = .52$, an indication that the predetermined “familiarization time” for all four age groups coincided with changes in developmental age.

(ii) *Number of fixations during familiarization*: The number of fixations is an indication of the length of each glance during which the infant is assumed to process information and make an engram of the exposed stimulus. Familiarization time in the FTII varies as a function of an infant’s age. In order to be comparable for different ages, the number of fixations was counted and transformed to a rate (per 4 seconds). The means and standard deviations of the number of fixations for all age groups are presented in Table 4-3. The data were also analyzed in order to determine any gender and/or cultural background differences. With one exception, no significant differences were found. The exception was for the youngest age group (i.e., 27-week-olds) in which the number of fixation for girls was lower than for boys ($t(16) = 2.1, p < .05$). An examination of the data revealed that items 1 and 3 of the FTII had the highest rate of looking in most age-related tests and for the sample as a whole. The high rate of looking for novelty item 1 suggests that infants may be distracted when the test starts without a warm-up procedure.

Finally, using one way ANOVA, developmental change in the number of fixations was examined (see Appendix F). The results indicated a highly significant linear increase for older groups $F(3, 74) = 11.47, p < .001$. That is, the length of each fixation becomes shorter with age. The results of post hoc comparisons using the Bonferroni test indicated that the number of fixations in the oldest group (52 weeks) was significantly higher than for

the first and second youngest groups and that it was also significantly higher in the 39-week group than the 27-week group (Figure 4-6).

Table 4-3

Means and Standard Deviations of Exposure Time (in Seconds) and the Number of Fixations for Each Age Group

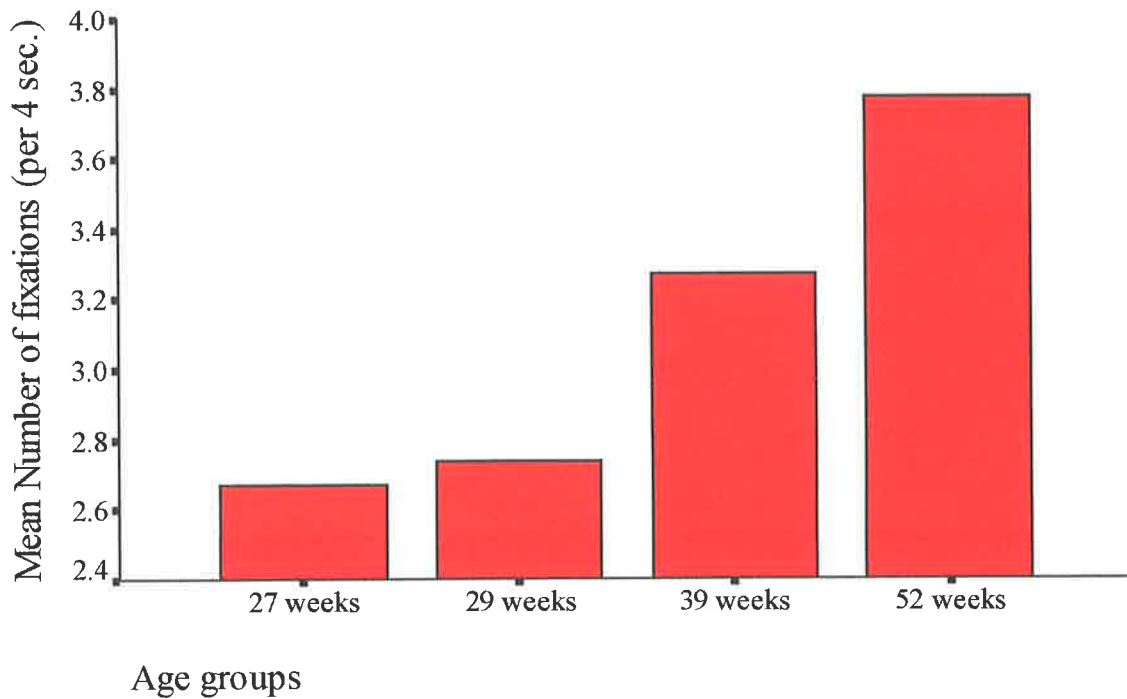
Test age	Exposure time		Number of fixations	
	M	SD	M	SD
27 weeks (n = 18)	7.6	1.5	2.67	.51
29 weeks (n = 15)	7.4	2.2	2.74	.56
39 weeks (n = 22)	7.6	2.6	3.27	.78
52 weeks (n = 23)	8.0	2.4	3.78	.77

Note: Exposure time and number of fixation were adjusted to rate per 4 seconds.

(iii) **Exposure time:** Exposure time is the total time for which an infant is exposed to the stimuli, so as to accumulate the preset familiarisation time. A measure of exposure time for each infant was calculated (due to the age adjusted differences in familiarization time, exposure time was transformed to a rate [per 4 seconds]). Exposure time for most age groups was positively skewed. Table 4-3 presents the average exposure time for each of the four age groups. Overall, exposure time was nearly equal to twice the predetermined familiarization time; that is, an infant needs to be exposed to a stimulus for eight seconds to accumulate four seconds of familiarization time. A further analysis indicated that this measure was not significantly different between the two genders. As is clear from the table exposure time did not show any developmental changes across different ages ($F(3,74) = .29$, $p = .83$; see Appendix G).

Figure 4-6

Simple Bar Chart of Number of Fixations by Age Group



(iv) *Developmental measures (Bayley Scales)*: The means and standard deviations of the MDI and the PDI for all age groups are displayed in Table 4-4. The means of the MDI varied from 110.9 to 115.8. In most age groups, girls scored slightly higher on the mental index than boys and boys scored slightly higher than girls on the psychomotor index, but the differences were not statistically significant (t ranged from 1.10 to 1.30, all ps NS).

Table 4-4

Means and Standard Deviations of MDI and PDI (Bayley, 1969)

Age	MDI (Bayley, 1969)			PDI (Bayley, 1969)		
	M	SD	Range	M	SD	Range
27 weeks (n = 18)	113.1	15.0	91-139	113.8	11.1	92-129
29 weeks (n = 15)	110.9	12.8	86-128	105.2	8.5	87-117
39 weeks (n = 22)	112.6	12.6	89-140	105.6	16.3	72-139
52 weeks (n = 23)	115.8	8.4	93-131	106.2	9.9	86-122

4-3-2 Internal consistency of the FTII

Theoretically, all items that make up a test should be correlated positively with each other if they are measuring a common entity. Using Cronbach's alpha the homogeneity and internal consistency of all four tests of the FTII was examined.

(i) **27 Week Test:** Cronbach's alpha for novelty preference at 27 week test was only .17 (Table 4-5). The inter-item correlations obtained for this test range from -.51 to .66 (mean intercorrelation $r = .02$), with most correlations low and/or negative. A more detailed examination of the item-total scale statistics suggests that there is little relationship between most of the individual items and the sum of the scores on the remaining items. Items numbers 4 and 10 were among those items that affected the low internal reliability of the test. Eliminating these items from the scale increased the alpha to .38 which, however, was still low. As can be seen from Table 4-5 exposure time and number of fixations were more internally consistent than novelty preference scores.

Table 4-5

Internal Consistency (Cronbach's α) for Different Measures Obtained from Visual Recognition Memory Task (FTII)

Age	Preference for novelty	Exposure time	Number of fixations
27 weeks (n = 18)	.17	.53	.38
29 weeks (n = 15)	-.46	.78	.52
39 weeks (n = 22)	.10	.46	.53
52 weeks (n = 23)	.28	.68	.50
Total sample (N = 78)	.10	.62	.64

(ii) **29 Week Test:** Cronbach's alpha for the 29-week test was negative ($\alpha = -.46$), suggesting that most items were not correlated among themselves and the reliability model

was violated. The exclusion of two items (i.e. items 6 and 9) that had the highest negative correlations with the total scale from analysis increased the outcome ($\alpha = .09$). Compared with the novelty preference score, the number of fixations and exposure times were again more reliable, with alphas of .52 and .78 respectively.

(iii) 39 Week Test: The internal consistency of novelty preference scores for the 39 weeks old test was also low ($\alpha = .10$). As was the case for the lower age groups, there was little relationship between individual items and the entire scale and this strongly affected the size of the alpha coefficient. Deleting the two weakest items (i.e., item numbers 4 and 6) increased the alpha to a moderate correlation of .45.

(iv) 52 Week Test: Although at this testing age the internal reliability was higher ($\alpha = .28$) than other ages, it was still very low to be a psychometrically reliable test. Consistent with the analyses at the three younger ages, alpha was calculated with the two least related items (items 3 and 7) partialled out. The obtained alpha increased to .43.

Finally, two other possibilities of increasing alpha were examined. First, the principle of aggregation states that, compared with a single administration, the sum of multiple administrations of a task can minimize the random error of the task (Rushton, Brainerd, & Pressley, 1983). Therefore, the possibility of increasing alpha as a consequence of the inclusion of more novel items was examined. That is, for those infants who had two testing sessions, alpha was computed using the novelty preference scores obtained at two different ages. The resulting alpha for 20 novel items obtained at 27 and 29 weeks was higher ($\alpha = .21$) than the alpha for 10 items, particularly for the 29 week test. However, alpha for 20 novel items obtained either at 29 and 39 weeks or at 39 and 52 weeks did not increase when 20 novel items were aggregated. This suggests that the low internal

consistency is largely due to low inter-item correlations in each age-related procedure of the FTII.

Second, it is feasible that the low internal consistency is a function of the time the infant has to spend in the testing situation. Therefore, the possibility of low internal consistency as a result of fatigue was examined. Two different alphas were calculated. Alphas were calculated for the first five novel items and for the next five novel items. The resulting alpha did not increase and, for the most part, the alphas were lower, perhaps due to the small number (5 in comparison to 10) of novelty items involved in the correlations.

Overall, it would appear that preference for novelty in all four tests of the FTII has a stable low internal reliability, not exceeding .28. In addition, the items which have low or often negative correlations with the total scale were not identical across different tests.

4-3-3 The test-retest reliability of the FTII

(i) *Test-retest reliability of novelty scores:* The cross-age to age stability of infants' looking behaviour was assessed using the Pearson's product-moment correlation coefficient (Table 4-6). The test-retest reliability of novelty preference scores across three different intervals of time (i.e., between 27 and 29 [short-term stability], 29 and 39 and between 39 and 52 weeks of age [longer-term stability]) was variable, ranging between -.17 and .29. Only short-term reliability was positive but still nonsignificant⁸. The scatterplot (Figure 4-7) mimics the information contained in Table 4-6 for correlation between means of novelty scores at different ages.

Rose et al. (1988a) suggested that because "true variance" in each novel problem is predominantly unique, the cross-age correlation would be higher between identical problems

⁸ The significance level of stability has often been assessed by one-tailed probability. Some researchers (e.g., Colombo et al., 1988) have argued that, since an inverse test-retest reliability for these measure is not expected, the significance level should be evaluated only in one direction (positive). However, according to the literature, some researchers have not found any stability for these infants' attentional measures. Thus, two-tailed *t* test have been preferred in the present study.

than between different items. Similarly, Fagan & Detterman (1992, p. 183) hypothesised that “... the Fagan test is composed of items that are measuring different aspects of cognitive functioning”. To test these hypotheses, a series of correlations was computed at the level of each individual novel problem, between all ten items, at three different age groups (i.e., between 27 and 29, between 29 and 39 and between 39 and 52 weeks). Once again, correlation coefficients across individual problems were highly variable, ranging from -.66 to .48. From the 30 correlations calculated, 12 were positive and the remainder were negative, with only three correlation coefficients statistically significant.

Table 4-6

Test-retest Stability of Attentional Measures

Age (weeks)	27 / 29	29 /39	39/52
<u>Measures</u>			
Novelty preference score	.29	-.07	-.17
Number of fixations	.46	.12	-.15
Exposure time	.17	.14	.59*
N	12	12	13

* $p < .05$ (two-tailed).

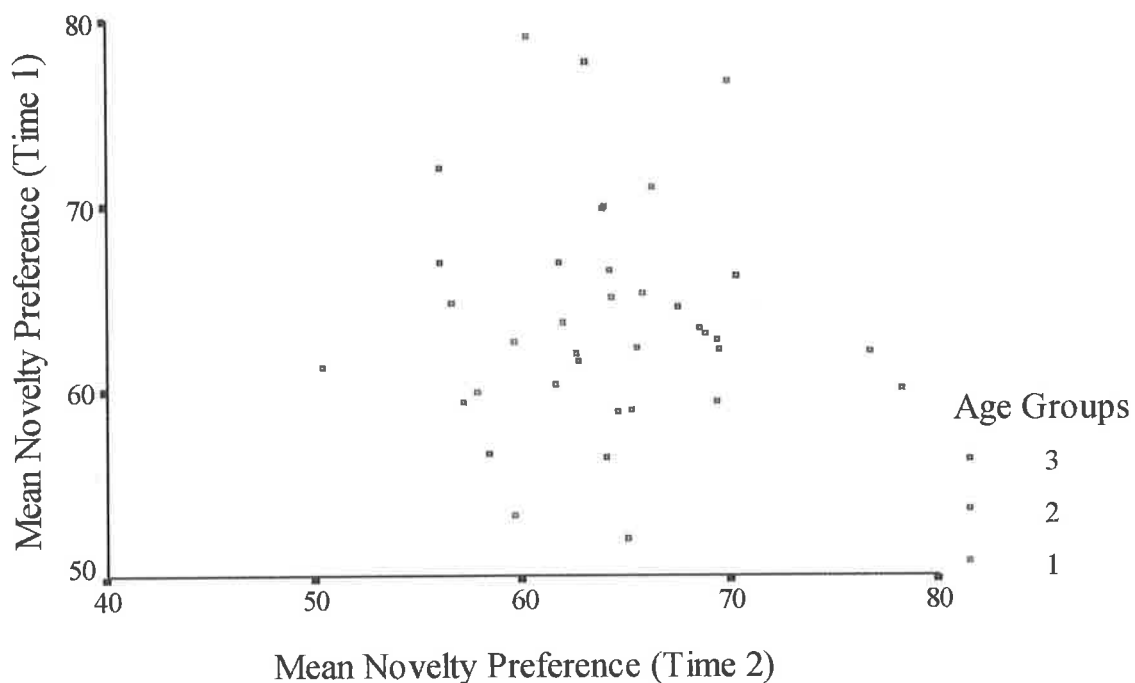
Overall, these correlations confirmed that, in practice, the long-term stability of novelty scores as assessed by the FTII, regardless of how it is computed, is low and inconsistent across different ages. The only indication of stability found here was for the presence of preference for novelty (see Rose & Feldman, 1987). That is, there was some stability in the number of subjects at two successive ages who exhibited preference for novelty above 54.5%⁹. All 11 infants (out of 12), whose novelty scores were higher than 54.5% of preference at 27 weeks, retained a higher score at 29 weeks of age. Of the 11

⁹ In the FTII, a mean novelty score greater than 54.5% is regarded as the presence of visual recognition memory and places the infant in the low risk group for retardation.

infants (out of 12) who achieved a higher score at 29 weeks, 10 subjects scored high novelty preference, again, at 39 weeks of age. Finally, all 13 subjects who were tested at 39 weeks and scored above the cut off point remained at above 54.5% at a later age (i.e., 52 weeks). This result shows stability in "quality" of preference rather than in "quantity" of preference.

Figure 4-7

Scatterplot of Novelty Scores by Age Group



Note: The scatterplot depicts the information contained in Table 4-7 for the test-retest reliability of novelty scores for age group 1 (27-week to 29-week tests), age group 2 (29-week to 39-week tests) and age group 3 (39-week to 52-week tests).

(ii) *Test-retest reliability of number fixations and exposure time:* The cross-age test-retest stability for the number of fixations and exposure time was also computed. The test-retest reliability of the number of fixations over a short interval of time (between 27 and 29 weeks of age) was positive ($r = .46$). However, the correlations between two other age



groups were low or negative (see Table 4-6). Exposure time was the only measure for which the correlation coefficients at all three age groups were positive and significantly stable between 39 and 52 weeks of age. Using a different task, Rose and Feldman (1987) and Colombo et al. (1988) also reported a moderate stability for exposure time at younger ages, from 6 to 8 months and from 4 to 7 months, respectively.

4-3-4 Intercorrelations between attentional measures

A series of bivariate correlations was computed to examine the relationship between novelty preference scores and other attentional measures obtained from the familiarization phase (Table 4-7). Overall, the correlations between the novelty preference scores and exposure time, and between the novelty scores and number of fixations for different age groups, were not consistent. Only the correlation between novelty score and the number of fixations was significant for 39-week-old infants ($r = .48, p < .05$), indicating that short-looking infants (infants who had a greater number of fixations during the familiarization phase) scored a higher percentage of preference for novelty. The results, then, partially replicated those of Colombo et al. (1988), who found a link between the number of fixations and novelty scores for 4 and 7 month old babies. The relationship between infants' exposure time and their number of fixations was also consistently positive at all ages and significant at 29 weeks ($r = .52, p < .05$).

For the sample as a whole, there was a significant relationship ($r = .33, p < .01$) between exposure time and the number of fixations, indicating that infants who looked away frequently (infants with lengthier exposure time) had a higher number of fixations at all ages. Overall, this pattern of correlations suggests that intercorrelations between looking behaviour measures are not straightforward, and depend on several factors, including the infant's age.

Table 4-7

Intercorrelation Between Attentional Measures

Measures	NPS/ET	NPS/NF	ET/NF
27 weeks	.20	-.09	.22
29 weeks	-.43	-.24	.52*
39 weeks	.41	.48*	.24
52 weeks	-.23	.26	.37
Total sample	.03	.18	.33**

Note: NPS is Novelty Preference Score, ET is Exposure Time, NF is Number of Fixations.
 * $p \leq .05$. (two-tailed). ** $p = .01$. (two-tailed).

4-3-5 Correlation between information processing measures and demographic characteristics

In order to examine the relationship between attentional measures and demographic variables, correlations between parental social standing, birth order and ethnicity background were computed. Table 4-8 presents correlations among these variables. As can be seen from the table, with the exception of a correlation between maternal age and

Table 4-8

Correlations Between Attentional Measures and Demographic Variables

	Novelty scores	Exposure time	Number of fixations
Mother age	.23*	.04	-.04
Maternal education	-.00	.34**	.11
Maternal occupation	.01	-.12	.00
Paternal education	.00	.28*	-.06
Paternal occupation	.12	-.26*	-.03
Birth order	-.02	.03	-.23*
Ethnicity	-.11	.19	.02

Note: The numbers of subjects were between 76 and 78 except for maternal occupation and paternal occupation which is 51 and 67 respectively. * $p < .05$. (two-tailed). ** $p < .01$. (two-tailed).

preference for novelty ($r = .23, p < .05$), there was no relationship between this measure and the demographic variables. There was also a significant correlation between number of fixations and birth order, an indication that first born infants have short fixation duration at familiarization period. Exposure time was the only attentional measure that correlated with some demographic variables- parental education and paternal occupation.

4-3-6 Analyses of novelty trials data

Two related but different sets of data can be obtained from the paired comparison procedure - data from “familiarization trials” and data from “novelty trials”. Table 4-9 presents descriptive data obtained from the novelty trials. Two different aspects of the results are of interest; the number of glances in each novelty trial and the time spent fixating on the stimuli in each trial. First, in novelty tasks, at least in terms of theory, it is expected that normal infants, after examining both familiar and novel targets, will favour the novel stimulus over the familiar one. The column headed "Number of fixations at test trials" in Table 4-9 indicates that infants, on average, missed fixating on 5 (out of 40)¹⁰ novelty trials. The frequency table (not reflected here) revealed that most subjects (59%, $n = 46$) ignored at least four, and up to 15 novelty trials (out of 40) during the test trials (i.e., they have only one glance in each trial). That is, their novelty scores (for those trials) were based on fixation on only one target (regardless of the fact that it may be a new or familiar target), without any examination of the next stimulus. This means that this glance has occurred by chance, and not intentionally.

Second, the actual looking time that infants spent studying the novel target was, on average, 4.45 seconds (out of 10 seconds). Although this amount of time is greater than the

¹⁰ The FTII has 10 novelty trials. Each novelty trial consists of two paired-comparison test trials during which infants have an opportunity for at least two fixations (left and right). Given two test trials for each of 10 novelty items and two possible fixations for each test trial, at least 40 fixations are possible if subjects look at least once at both familiar and novel stimuli in each trial.

Table 4-9

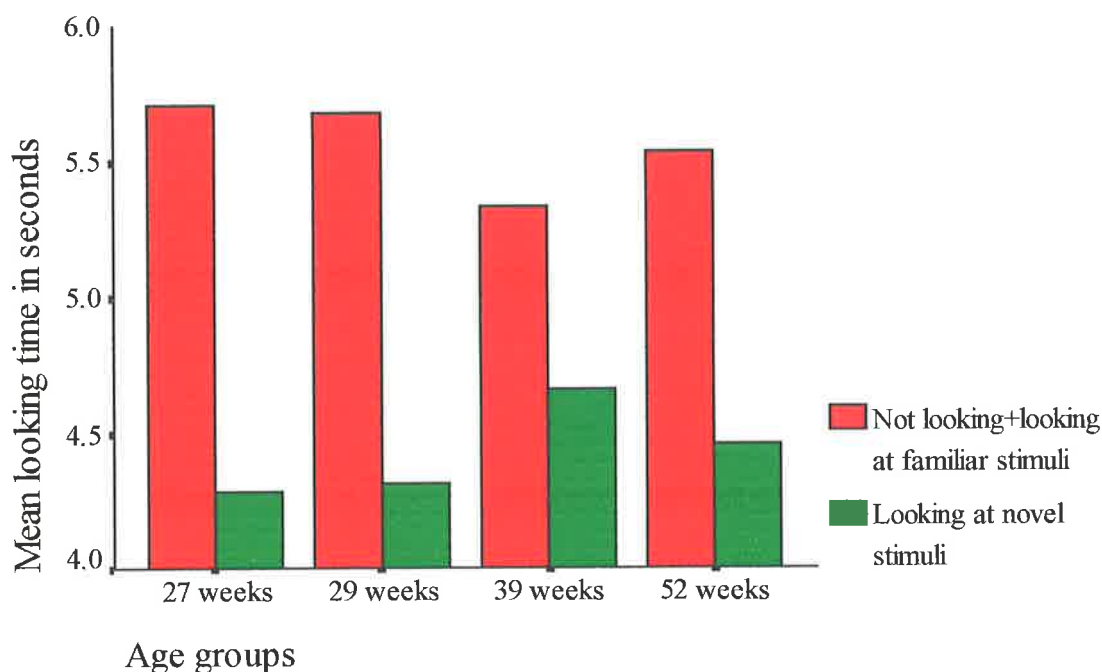
Means and Standard Deviations for Number and Time of Fixations During Novelty Trials

Age groups	Number of fixations at test trials a		Time fixating novelty stimulus b		Time fixating familiar stimulus b	
	M	SD	M	SD	M	SD
27 weeks (N = 18)	34.9	2.9	4.30	0.77	2.40	0.71
29 weeks (N = 15)	34.7	3.6	4.32	1.14	2.51	0.55
39 weeks (N = 22)	35.0	4.1	4.67	0.58	2.81	0.59
42 weeks (N = 23)	35.6	3.6	4.46	0.84	2.42	0.55
Total sample	35.1	3.5	4.45	0.83	2.54	0.61

Note: a- The minimum possible number of fixations is 40 if subjects have at least one look at familiar and one look at novel stimuli. b- Time fixating to novelty and to familiar stimuli during novelty trials are adjusted to rate per 10 seconds for all age groups.

Figure 4-8

Comparison of Time Spent Looking at Novel Stimuli and Remaining Time (i.e., Time Looking at Familiar Stimuli Plus Time Not Looking at any Stimuli)



time spent looking at a familiar stimulus (i.e., 2.54 seconds), it is less than 50% of the total time allocated for looking at both familiar and novel stimuli (see Figure 4-8). Thus, using the Fagan scoring system for measuring preference for novelty, an average of 3.01 seconds of the preset novelty trial time (i.e., $10 - [4.45 + 2.54] = 3.01$ seconds) was not utilized by the infants.

4-4 DISCUSSION

The present study offers clear evidence that the internal consistency of novelty preference scores, as indexed by Cronbach's alpha, for the FTII is low or negative, an indication that test items are not correlated and the internal reliability model is violated. Moreover, some different novelty items in each age-related test did not correlate with the total scale, suggesting little relationship between individual novelty items and the scores on the remaining items. The test-retest reliability for the FTII was also extremely variable and nonsignificant.

In contrast, the numbers of fixations during familiarisation and exposure times, were found to be more internally consistent and exposure times between 39 and 52 weeks of age were significantly stable. While no developmental change was found for novelty scores and exposure time, the results imply a significant developmental trend for fixation duration- that is, the number of fixations increased linearly with age. Of more importance was the inconsistency of bivariate correlations involving different attentional measures (i.e., novelty preference, exposure time and number of fixations) at different ages. When the effect of age was held constant, a significant correlation was found between the number of fixations and exposure time for the sample ($r = .33, p < .01$). No significant correlation was found between other attentional measures, however. The results, then, were inconsistent with the

hypothesis that all three different visual attention measures are dependent on each other in each of the different age groups.

Before further discussion of the results there are several points that should be considered. First, the novelty scores for the Australian sample were between 62.4% and 64.6% at different age groups. In all ages, it was slightly higher than the normative data scores (59%). The novelty scores of the normative data are based on multiple administrations of the test which may result in a more accurate preference score, whereas the novelty score in the current study was based on one occasion only (10 items). However, the scores were close to the novelty preference scores reported by Struthers and Hansen, (1992), Fagan and McGrath (1981) and Fagan and Singer (1983) for normal infants and for a random sample of Norwegian infants (Andersson, 1996). The relatively higher novelty scores in this sample were also consistent with higher scores on other measures of cognitive performance (i.e., the MDI and the PDI).

Second, the novelty scores did not show any developmental change across the second half of the first year of age. The apparent lack of developmental change, particularly in older infants, could be attributed to the nature of the stimuli and a ceiling effect. That is, infants may fail to show a higher percentage of novelty preference because they are already performing at the peak of their capacities (McCall & Carriger, 1993). Similarly, there was no indication of developmental trend for exposure time confirming Rose and Feldman (1987) who also did not find developmental change for human face stimuli in infants aged 6 to 8 months. However, the number of fixations was the only age-related measure of infant attention during the second six months of age. That is, data supported the research hypothesis, because there was a highly significant increase in the number of fixations with age.

Third, in general, individual differences in novelty scores, tended to be independent from socioeconomical factors such as parental education, parental occupation, birth order and ethnicity background. The only relationship found was for maternal age, with babies born to older mothers having a higher percentage of novelty scores, confirming Rose et al. (1989) who reported a significant correlation ($r = .30, p < .05$, two-tailed) between these two measures. In the present study number of fixations was found to be related negatively to birth order, suggesting that first born babies have more glances (i.e., shorter looking time in each glance). If short looking is regarded as a better performance on visual recognition memory tasks (Colombo, Mitchell, Coldren, & Freesman, 1991; Colombo & Mitchell, 1990) then the relation between birth order and number of fixations could be an indication of better performance in favor of first born babies. There were also significant correlations between exposure time and parental education and paternal job status. These relationships are difficult to interpret because exposure time is a measure of inattention and it is expected to be lower in babies from middle class compared to lower class parentage.

(i) Internal consistency of the FTII: Fagan and Detterman (1992) argue that items on the FTII may be sampling different cognitive skills that are either not correlated, or are only weakly correlated with each other. The low internal consistency and relatively high number of novel items that are not correlated to the total scales, supports the notion that the FTII probably is not measuring a single common entity operationalized as novelty preference. These unrelated items are somewhat different from one test to the next. This implies that utilising identical novelty items across six months of developmental change may, in turn, contribute to violation of the reliability of the test.

Moreover, further analysis revealed that the low internal consistency is not simply a function of small number of components (i.e., number of novelty items); alphas for 20 items

rather than 10 items were also low. Furthermore, an item analysis of the test indicated that the number of novel items that did not exceed chance performance is relatively high, reaching a maximum of five items at 29-weeks. Chance performance not only produces an unreliable outcome for the test within a session but also invalidates test-retest reliability between sessions. On the basis of theory, it is expected that all items will significantly exceed the chance value in the normal population. However, the infants showed no significant preference for novelty items 1 and 9 in all four tests of the FTII (and items number 7, 8 and 10 in some tests). The infants' inability to discriminate the first item of the test may be due to a procedural problem with the administration of the test. That is, because the test starts without any "warm up", infants are not oriented to the task and spend time exploring the visual environment and manipulating the apparatus rather than engaging in the task. However, the low novelty score for this item and other items that did not exceed chance performance could also be a function of the degree of difficulty associated with a task or the amount of time allotted for the familiarisation period for these items.

Nevertheless, the low internal consistency for novelty preference scores found in this study supports the results reported by Andersson (1996), Jacobson et al. (1992), Fagan and Detterman (1992) using some tests of the FTII. Using four different types of novelty problems, Rose et al. (1988a) have found a very low internal consistency for novelty scores. They reported that "the interitem correlations ... are low not only because the error variance of problems is high relative to their true variance but also because the true variance itself is predominantly unique rather than shared" (p. 1188). This supports the hypothesis that attentional behaviour during infancy is contaminated by a variety of uncontrollable variables, rendering it impossible to obtain a precise measurement.

(ii) *Test-retest reliability*: The results of the study were even more discouraging with respect to the test-retest reliability of novelty preference scores. Only with the operational definition of stability as the "presence or absence of preference for novelty", the current study indicated that novelty scores are stable. This notion of stability is not convincing for researchers who define reliability in a purely psychometric sense, however. The present study replicated the results of other studies that have used the FTII and did not find test-retest reliability for this test (e.g., Benson et al., 1993; DiLalla et al., 1990; Jacobson et al., 1992).

Exposure time was the only measure of attentiveness that showed positive correlations across all ages and a significant stability ($r = .59, p < .05$) in the oldest age groups (i.e., between 39 to 52 weeks of age). Although the present study extended Colombo et al.'s (1988) findings to 12 months of age, it did not replicate the stability in exposure time for younger infants. Rose and Feldman (1987) maintained that exposure time is an indication of an infant's ability to sustain attention on-task. Infants who direct fixation to the stimulus as soon as it becomes available, and spend more time looking at the target than other aspects of the environment, ultimately have a lower exposure time.

With respect to the reasonable predictive validity of some visual recognition memory tasks, the low reliability of novelty scores is difficult to understand. On this point, several possible explanations should be considered. While some explanations may rest on the structure of the measurement itself, there are also grounds for suggesting that the instability of the measures may be related to factors associated with the infants, rather than the test.

In relation to the structure of the test, it is possible that the nature of the stimuli used in the FTII has some influence on the low test-retest reliability of the task. All items in the FTII are pictures of a human face of different age, sex, affective expression and orientation.

It is reasonable to suggest that the instability of the attentional measures may be due to a change in an infant's ability to recognize social stimuli at different points in time. This suggestion supports Cohen's (1988a) proposition that the observed instability may be attributed to infants at different ages processing different aspects of a single stimulus. According to Cohen, it is possible to obtain a higher reliability if different stimuli (appropriate for different ages) are used at different points in time.

It is also feasible that infants show less variability in their preference for some nonsocial stimuli, such as alphabet letters and geometric forms or abstract patterns, rather than pictures of human faces. In fact, the novel items in studies conducted by Colombo et al (1988) who found some degree of stability consisted of alphabet letters. While Rose and Feldman (1987) found a significant test-retest reliability for abstract patterns they did not find a significant cross-age correlation for face stimuli. In any event, if an infant is viewed as a constructive thinking creature (Flavell, 1992) with a specific personality (Sanson, Prior & Oberklaid, 1985), it is not inconceivable that different social stimuli hold different appeal for them at different ages.

The low reliability might also be attributed to the scoring system of the Fagan test. As results of item analysis suggest, most subjects (59%) missed at least four novelty trials, indicating that at least two novelty items (out of 10) were looked at only by chance. In other words, the novelty score for these items is based on one looking at only one stimulus without involving examination of the second stimulus. This may contribute to the fluctuation of novelty scores within a session, and low test-retest reliability between sessions.

Infant characteristics provide a further possible explanation for the unstable results. For example, the rapid developmental change in infancy and/or different temperament

behaviours may result in the instability of measurement. As Sigman (1988) suggests, the low stability in test performance for young infants may be expected when the rapidity of developmental changes (maturation) and experience (learning) affect the actual condition of the infant from time to time. Similarly, because an infant's behaviour is variable and his or her attention span, temperament and enthusiasm for the test situation changes from time to time, the test-retest instability may not necessarily reflect variation in ability of visual recognition.

(iii) *The relationship between attentional measures:* The existence of intercorrelations between the number of fixations and novelty scores led Colombo to conclude that short-looking infants are more intelligent than infants who have longer fixation durations (Colombo et al., 1991, 1988, 1986; Colombo & Mitchell, 1988). However, the present study failed to find a significant relationship between length of fixations and novelty scores (except for 39-week old subjects). This implies that short-looking infants may not be necessarily brighter than infants with lengthier fixation durations.

Finally, the data illustrated the relationship between the number of fixations and exposure time for subjects as a whole. Infants with high exposure times also have higher numbers of fixations, probably due to a higher shift rate between stimuli and the environment beyond the task. Apparently, these infants are less concentrated on task and make themselves engage with environment factors.

In summary, the current study found no psychometric consistency for the main cognitive measure of the FTII - the novelty preference scores. Ostensibly, in terms of classical test theory, the FTII is not a reliable test. However, further research is needed to provide an empirical basis for this claim. Item analysis may provide some basis for revision to the FTII. For example, identifying and deleting the novelty items that are unreliable for

each age group may improve the test. However, the small number of subjects in each group of the present study indicates that caution should be exercised in generalizing from the results. Further research is warranted.

CHAPTER 5

THE PREDICTIVE VALIDATION OF THE FAGAN TEST OF INFANT INTELLIGENCE

5-1 INTRODUCTION

The theory on which Fagan's new procedure is based argues that the neural integrity required for later cognitive functioning exists from birth, and infants are equipped with elementary functions necessary for later IQ. One such function is an infant's memory. The ability to discriminate between the familiar and new stimuli is considered to be an indication of the infant's memory. This procedure is assumed to provide a reliable marker of the subsequent development of cognitive abilities. Several investigators using the paired-comparison procedure with a variety of experimental stimuli (abstract patterns, human face photos, geometric forms, sounds) in different modalities (visual, auditory and cross modal transfer), and at different ages, have examined the predictive validity of the new measures (see section 3-4-4).

During the 1970s and early 1980s, Fagan conducted several studies in which he investigated infants' recognition memory (e.g., Fagan, 1970, 1972, 1974) and the validity of the procedure for evaluating their current level of mental activities (Fagan & Singer, 1983), as well as the predictability of such visual recognition measures to later mental outcomes (Fagan & McGrath, 1981). Based on these studies, Fagan and McGrath (1981) concluded that "variations in early visual recognition memory both reflect and predict variations in intelligent functioning" (p.129).

In the following years, Fagan and his colleagues published the results of several studies in which the utility of novelty tasks in detecting later mental retardation were shown to be replicable. However, none of the studies was based entirely on the version of the Fagan Test of Infant Intelligence that is now commercially available (Fagan & Shepherd, 1986, 1991). In fact, the studies were based on different experimental stimuli targeted at different age groups from which some of the novelty problems for the marketed version were drawn (Benasich & Bejar, 1992). As yet, only a few independent studies have examined the concurrent and predictive validity of this marketed version of the FTII and, to date, the results have not been consistent in relation to its power to predict subsequent mental outcome.

5-1-1 Concurrent validity of the Fagan test

Concurrent validity for measures of habituation and visual recognition memory is generally based on data from a population which is expected (e.g., Down Syndrome babies), or suspected (e.g., preterm infants), to be different from normal infants in cognitive performance later in life. In comparison with intellectually disabled or at-risk infants, several such studies reported a superiority in preference for novelty in normal infants, either in terms of the magnitude of the novelty score or in terms of exhibiting preference for novelty at an earlier age (e.g., Cohen, 1981; Miranda & Fantz, 1974, see also section 3-4-3).

In relation to the Fagan test, one study (Jacobson et al., 1985) assessed the concurrent validity of three novel items of the Fagan test in 123 infants at 7 months of age, some of whom had been exposed to polychlorinated biphenyls (PCB). The novelty preference score was found to be decreased systematically with higher levels of PCB exposure, supporting the concurrent validity for the Fagan test. A further study reported by Struthers and Hansen (1992) also examined the performance on the FTII for 36 drug

exposed (cocaine and/or amphetamines) infants aged between 37 and 41 weeks, and compared this with 26 infants matched for race and socioeconomic status, who had not been exposed to drugs. The results indicated that infants exposed to drugs had significantly lower scores in the percentage of preference for novelty than infants in the control group.

Another study has attempted to compare the results of a traditional developmental scale with those of the Fagan test. Greenfield (1989, cited by Benasich & Bejar, 1992) compared the outcome of the Fagan test with the Bayley Scales in a sample of 3-month old infants with prenatal complications. Unlike the Bayley Scales, low scores on the Fagan test were associated with birth variables, such as low birthweight, small head size and lower gestation age, suggesting that the novelty scores were a more reliable indicator of developmental status than the MDI.

The only study to use two different age tests of the latest version of the FTII was conducted by Jacobson et al. (1992). They followed the procedure recommended for testing at 27 and 52 weeks in their study of 103 black infants (on two different occasions - 27 and 52 weeks of age) in order to investigate the relation of measures derived from visual expectation (i.e., percent anticipation and reaction time), to other infant tests (e.g., a standard developmental assessment [Bayley Scales], information processing test [FTII], cross-modal transfer task, complexity of play and object permanence [Uzgiris & Hunt, 1975]). A principle components analysis with varimax rotation revealed that the FTII and the Bayley Scales loaded on two separate factors, indicating that they reflect two distinct domains of infant cognitive functioning.

Similarly, Rose and Feldman (1987) reported that, for 7-month-olds, both the novelty preference score and exposure time correlated poorly and nonsignificantly with the MDI and PDI from the Bayley Scale. In another study (Rose et al, 1989), however, they

found a significant correlation ($r = .47, p < .01$) between concurrent MDI and visual recognition memory for 7-month preterm infants.

To the extent that novelty scores reflect a general intellectual capacity (Fagan, 1988b), in principle, it would be expected to find some shared variance with other tests of perceptual and sensorimotor development. However, if information processing measures and developmental indices are independent, such correlations would not be found. Little research has been done to investigate the relationship between the Bayley Scales and the FTII, particularly in the normal infant population.

5-1-2 The predictive validity of the FTII

Few studies have concerned themselves with the predictability of the Fagan procedure. One independent study aimed at the predictability of the infancy measures to later IQ was conducted by DiLalla and her colleagues (DiLalla et al., 1990). Seven domains of cognitive development were evaluated between 7 and 9 months of age. The FTII was also administered to the infants at 7 and 9 months of age. The results reported for the predictability of the Fagan test were mixed. A composite score from the FTII (a combination of four novelty scores obtained at 7 and 9 months by the Immediate and Retest Versions), predicted later outcomes in terms of the Bayley MDI at one year ($r = .15, p < .05, N = 143$) and the Stanford-Binet test at 3 years of age ($r = .32, p < .05, N = 51$). The Immediate FTII at 7 months also predicted 3-year Stanford-Binet IQ, and both Immediate and Retest Versions of the FTII at 9 months predicted MDI at one year of age. However, no novelty preference score at 7 or 9 months predicted later MDI at 2 years of age, no predictive validity was found between 7 months novelty and one year mental outcomes, and no prediction could be made from the 9 month FTII to the 3 year Stanford-Binet IQ. The smallest correlation was $r = .15$, but no correlation exceeded $r = .32$ (all $ps < .05$).

Using the 5 and 7-month test, Thompson et al. (1991) also attempted to estimate the predictive validity of the FTII in a sample of adopted and nonadopted infants. The corrected for unreliability (but not the uncorrected) correlation coefficient predicted the MDI at 2 years of age ($r = .21, p < .05$ [one-tailed], $N = 107$) and both uncorrected and corrected correlations predicted the Stanford-Binet score, and a battery of specific cognitive abilities, at 3 years of age. The composite novelty score did not predict 12-month MDI, however.

More recently, Andersson (1996) examined the predictive validity of the FTII in a random sample of 196 Norwegian infants between the ages of 7 to 9 months residing in two different cities. A follow-up study at 5 years of age indicated that the FTII scores predicted Verbal IQ, Full Scale IQ (WPPSI-R) and two tests of psycholinguistic abilities for the sample from one city. However, in the second city the FTII predicted only Verbal IQ and one test of the psycholinguistic test. The mean of the correlations was as low as .22.

Despite the claim that the Fagan test is now in use in over 100 centres in the United States and nine other countries (Fagan & Detterman, 1992), no independent studies are yet available to replicate the predictive validity of all four age-related tests (i.e., 27, 29, 39 and 52 weeks) of the marketed version of the FTII. The only research to date using all four tests of the FTII is the study published by the test developer and his colleague (Fagan & Detterman, 1992). The reported predictive correlations ($r = .36, p < 0001, N = 149$) are based on multiple administrations of the task, and are also based on inclusion of some at-risk infants. Moreover, no predictive validity was reported for each specific age-related procedure (i.e., 27, 29, 39 and 52 weeks). There is evidence that some tests of the FTII have no predictive validity (e.g., DiLalla et al., 1990). In practice, novelty tasks are not interesting for some infants and multiple administrations of the tasks are generally difficult. In addition, very low internal consistency and test-retest reliability observed in Study 1

place practical limitations on the validation of the test. Finally, it was considered worthwhile to assess the validation for the FTII by using an Australian sample.

Following up all subjects tested in Study 1, the present study attempted to verify several other points. First, an infant's age at the initial test may play a role in the prediction value of the test. Research has indicated that habituation and visual recognition memory tasks would be more reliable for younger infants (Slater, 1995). The FTII has four age groups tests ranging from 27 to 52 weeks. The studies that used different age-related tests did not report consistent results in this regard. For example, Andersson (1996) reported that exclusion or inclusion of 8 month-olds into the analyses did not change the general picture of predictability of the FTII. Using all four tests of the FTII makes it possible to test the hypothesis of higher predictive validity in the earlier testing ages versus testing at later ages.

Second, the prediction of later intellectual functioning for the FTII is based on data obtained from the test phase only (i.e., the percentage of preference for novelty). However, data from the familiarization phase may also have some degree of validity for later IQ. Although the FTII has made no claim for the predictive utility of measures such as exposure time and the number of fixations that infants make to accumulate familiarization time, the validity of these measures for predicting later cognitive outcomes was also evaluated in the same way as novelty preference. These measures were selected because Study 1 indicated that the number of fixations and exposure time were more internally consistent than the novelty preference score.

Moreover, exposure time and duration of fixations showed preterm/fullterm differences, with preterm infants requiring longer exposure time to accumulate the familiarization time (Rose et al., 1988b). Also, longer exposure time at 7 months has been shown to predict low intelligence from 1 to 5 years (Rose et al., 1989). Similarly, the

number of fixations has been found to have some concurrent validity as an indication of infants' abilities for faster processing of information (Colombo et al., 1988, 1991). If fixation duration is considered to be a valid looking behaviour, the assessment of the predictive validity of this measure to later IQ was considered to be an important issue to be investigated.

Finally, a question was raised by Benasich and Bejar (1992) concerning the selection of an arbitrary cut-off score for screening at-risk or retarded infants. There is evidence indicating that novelty scores in general and, more specifically, different versions of the Fagan test, have been shown to misclassify infants at times (e.g., Fagan et al., 1985, 1986, 1987; Fagan & Montie 1986). This is of particular importance, since the cut-off score may limit interpretation of the task. Although the sample in the present study apparently consists of normal infants it is possible to evaluate the sensitivity and specificity of the FTII in terms of detecting low score and high score infants.

5-2 METHOD

5-2-1 Subjects

A full description of the demographic characteristics of the 78 infants and their parents was given in Study 1. To recapitulate, there were 42 males and 36 females in four different age groups corresponding to the four age-related procedures of the FTII (i.e., 27, 29, 39 and 52 weeks corrected for prematurity). On the basis of voluntary participation, 37 infants were nominated by their parents to be involved in a second test. The measures obtained from the second testing session were intended to be used for assessing the stability of the FTII, as well as for an evaluation of the relevance of utilising different numbers of novel problems at different ages for the prediction of later intellectual outcomes.

5-2-2 Measures at follow-up testing

Follow-up cognitive measures were made for the sample through 18 months to 2 years of age using two versions of the the Bayley Scales (Bayley , 1969 and 1993) as a developmental scale, and the fourth edition of the Stanford-Binet test (Thorndike et al., 1986) as a psychometric test of intelligence.

(i) *Bayley Scales of Infant Development (Revised Version)*: The second edition of the Bayley Scales of Infant Development (Bayley, 1993) is an individually administered test, designed to evaluate an infant's current mental and motor development during the first 42 months of life. Although many new items were incorporated, and some items in the first edition were eliminated from the second edition of the Scales, the nature, and the majority of the items in the first version, have remained the same. The new items added to the mental scale are intended to assess an infant's memory and ability to solve problems, as well as numeracy skills, language development and personal-social behaviours (Bayley, 1993). Converting the sum of the raw scores to a standard score results in a "Mental Development Index", a "Psychomotor Development Index" and a "Behaviour Rating Scale". However, for the purpose of the present study, only the mental index was obtained. The mean of the Bayley Scales (Bayley, 1993) is 100 and the standard deviation is 15.

(ii) *The Stanford-Binet Intelligence Scale*: The fourth edition of the Stanford-Binet test (Thorndike et al., 1986) has 15 tests each designed to measure a specific area of cognitive ability. Subtests are combined to provide Standard Age Scores in Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning and Short-Term Memory. These can be used to derive three broad factors of cognitive functioning - Crystallised Abilities, Fluid-Analytic Abilities and Short-Term Memory. Only eight tests apply to children at two years of age, from which five tests were administered to the subjects in this

study. These were Vocabulary and Comprehension tests (i.e., Verbal Reasoning), Pattern Analysis and Copying (i.e., Abstract/Visual Reasoning) and Memory for Sentences (i.e., Short-Term Memory). These five tests were chosen because they are the most suitable to overcome the difficulty of subjects at the lower end of the normative group frequently scoring zero, particularly for Bead Memory and Quantitative tests (Thorndike et al., 1986).

5-2-3 Procedure

Testing for all four age groups was scheduled as follows. Infants who were tested initially at 27 and 29 weeks were tested again at 18 months. Infants who were 39 and 52 weeks old at the first testing were given a follow-up testing one year later, at 21 and 24 months of age respectively. During testing, the subjects were administered two different versions of the Bayley scales (Bayley, 1969, 1993) from which two Mental Development Indexes (MDI) for all subjects were obtained. Since the predictability of FTII in relation to Stanford-Binet test in 2-year-olds is still unknown, another aim of the research was to examine this possibility. Therefore, those subjects who were tested at 18 and 21 months of age were requested to participate again in a final testing session at 24 months during which the Bayley scales (both versions) and the Stanford-Binet were administered together (test scores on the Stanford-Binet scale for 2-year-old were taken during the 24-month-testing session). All subsequent assessments were administered at corrected age for prematurity within 14 days of the subject's expected birthday at 18, 21 and/or 24 months.

In order to keep the effects of internal factors such as fatigue and irritability, as well as the effect of practice and experience constant for all subjects, a fixed sequence of test administration was used. Items common to both mental scales of the Bayley test were administered first. The new items in the second version and all remaining items from the first version were then administered in counterbalanced order to all of the sample. The

Stanford-Binet test was also administered to all children following the Bayley Scale after a short break. The parents were instructed to facilitate their infant's performance but not to interfere or assist the examiner unless they were requested to do so. For some children from ethnic backgrounds, limited parental assistance was provided as required for language items. For some cases ($n = 6$) a single session proved to be too tiring and a second session was therefore required.

5-2-4 Attrition rate

The attrition rate at follow-up was relatively low. Only four subjects did not return for subsequent testing, due to a change of residential address ($n = 2$) and overseas or interstate travel ($n = 2$) at the scheduled time (Table 5-1). Ten subjects for whom two

Table 5-1

Attrition Rate for Sample

Age groups	Total sample (infancy testing)	<u>No returns</u>	<u>Lost one of follow-up testings</u>		
			18 months	21 months	24 months
27 weeks	18	-	2	-	1
29 weeks	15	2	1	-	1
39 weeks	22	-	-	2	3
52 weeks	23	2	-	-	-

follow-up testing sessions were scheduled also missed one of two follow-up testing sessions, mainly because of scheduling difficulties, interstate travel and/or sickness. One more subject was tested at both follow-up sessions, but eliminated from the results.¹ With one exception, no significant differences in any demographic characteristics (t ranged from

¹ One subject from the 27-week-old group who was tested at both follow-up testings was omitted from the analysis after referral to an audiometrist found 65% loss of hearing.

1.21 to .87 all *ps NS*) or cognitive measures (*t* ranged from .94 to .80 all *ps NS*) were found between those who missed the follow-up assessments and the group who participated in childhood testing. The one exception was ethnic background ($\chi^2 (1) = 5.39, p < .05$). Three of the four infants for whom follow-up data were missing were from a non-English speaking background.

5-3 RESULTS

5-3-1 Preliminary analysis

(i) *Childhood measures*: Table 5-2 presents the means and standard deviations of the MDI of the Bayley Scales (Bayley, 1969, 1993) and Stanford-Binet test scores in childhood. A series of *t* tests was computed to determine if there were any gender or ethnic differences between the childhood measures. The MDI scores drawn from the new edition of the Bayley Scales (Bayley, 1993) were significantly higher for girls than boys at 24 months ($t (66) = 2.46, p < .05$)². However, there was no significant difference for gender for the Stanford-Binet test, nor were significant differences found for the MDI and Stanford-Binet scores between children from English and non-English speaking backgrounds in this age range (*t* ranged from 1.66 to .43, all *ps NS*).

It is clear from Table 5-2 that the MDI scores from the two versions of the Bayley Scales are significantly different, indicating that the first version of the Bayley Scales (Bayley, 1969) is no longer a reliable measure of infants' competencies. The number of tests of the Stanford-Binet scale administered to the infants was different (mostly four or five)

² The difference between MDI scores for the first version of the Bayley Scales (Bayley, 1969) was also significantly higher for girls than boys ($t (66) 2.17, p = .034$) at 24 month of age.

Table 5-2

Means and Standard Deviations of Infancy and Childhood Measures

Age	<u>18 months</u>			<u>21 months</u>			<u>24 months</u>		
	N	M	SD	N	M	SD	N	M	SD
<u>Childhood tests</u>									
<u>Bayley Scales</u>									
MDI (1969)	27	117.7	13.6	20	116.7	15.0	68	122.5	19.4
MDI (1993)	27	98.6	10.6	20	100.0	11.1	68	102.7	14.2
<u>Stanford-Binet test</u>									
Vocabulary							56	101.6	10.6
Comprehension							53	103.6	9.0
Pattern Analysis							60	103.2	7.5
Copying							56	98.9	6.7
Short-term Memory							30	104.9	7.7
General IQ							61	102.1	7.9

Note: The MDI is the Mental Development Index, obtained from the Bayley Scales (Editions 1969 and 1993) respectively. The PDI is the Psychomotor Development Index obtained from the Bayley Scales (Bayley, 1969).

because some subjects refused and/or were unable to perform some tests (e.g., Memory For Sentences, Vocabulary). Since the lowest age level of the Stanford-Binet test is 24 months, seven children could not be tested due to inability to perform any item of the test ($n = 5$), or because of uncooperative behaviour ($n = 2$) at testing. The low standard deviation of the test is an indication of the relatively similar performance of most children, particularly on the Abstract/Visual Reasoning factor (i.e., Copying and Pattern Analysis).

5-3-2 Relationship between information processing measures and the Bayley Scales

Both the MDI and the PDI obtained from the Bayley Scales (Bayley, 1969) during infancy were correlated with measures obtained from the FTII at the same age.

Intercorrelations between the MDI and the PDI with information processing measures are presented in Table 5-3. Although the correlations between novelty scores and the MDI and

Table 5-3

Correlations Between MDI and PDI and Measures Obtained From Information Processing Task

Infancy age	27 weeks	29 weeks	39 weeks	52 weeks
Information processing Measures				
Novelty Scores				
MDI	.38	.48	-.02	.38
PDI	.31	.33	.04	.27
Exposure Time				
MDI	.31	-.24	-.08	-.17
PDI	.18	-.15	-.17	.04
Number of Fixations				
MDI	-.11	.08	-.15	.06
PDI	-.02	-.17	-.01	.14

PDI are, for the most part, positive, none of them were significant, suggesting that novelty scores are generally independent of the MDI and the PDI, at least during infancy. Similarly, the relationships between the Bayley Scales (both MDI and PDI) and exposure time and number of fixations were also inconsistent and nonsignificant.

5-3-3 Predictive validity of the FTII

(i) *The relation between infancy novelty scores and later MDI:* The percentage of preference for novelty scores for all four different age groups (i.e., 27, 29, 39 and 52 weeks of age) were correlated with the developmental scores (MDI) obtained at 18, 21 and 24 months. The resulting statistics are displayed in Table 5-4.

As the table indicates, correlations between the 27- and 39-week FTII and later cognitive developmental index at 18, 21 and 24 months are all low and nonsignificant, and did not exceed .24. The predictive validities for 27 and 39 week tests obtained in the present study were similar to those of DiLalla et al. (1990). These authors reported no correlation between the 27 and 39-week FTII and later 24-month MDI. However, the 52 week Fagan

Table 5-4

Bivariate Correlations Between the FTII and Later Bayley Scales at 18 to 24 Months and
(the Number of Subjects) in Each Analysis

Childhood Age Bayley Scales	18 months		21 months		24 months	
	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)
27-week novelty score	.17 (15)	.04 (15)			.19 (16)	.23 (16)
29-week novelty score	-.43 (12)	-.52 (12)			-.53 (12)	-.38 (12)
39-week novelty score			.23 (20)	.24 (20)	.02 (19)	.07 (19)
52-week novelty score					.44* (21)	.49* (21)

Note: MDI (1969) and MDI (1993) are Mental Development Indexes from the first edition of the Bayley Scale (Bayley, 1969) and the revised edition of Bayley Scales (Bayley, 1993). * $p \leq .05$. (two-tailed).

test has a relatively higher correlation and significantly predicted the 24-month MDI for both versions of the Bayley Scales. That is, higher novelty scores at 12 months were predictors of higher MDI at 24 months. The highest correlation coefficients between the 52 week-test and 24 months MDI indicated that it is possible for the FTII to account for up to 24% of the shared variance in later outcomes³

Of considerable interest were the negative correlations (all nonsignificant) between the 29-week novelty task and later MDI at 18 and 24 months. This result for the 29-week FTII was not surprising due to a lack of internal consistency for the 29-week novelty task. In terms of psychometric properties, this test of the FTII produced the poorest results. In

³ On the basis of theory, the correlations between novelty scores and later outcome are expected. Consequently, some researchers (e.g., Andersson, 1996; Thomson et al., 1991) calculated the probability of the significance level with one-tailed t test. In the present study, because the preliminary analysis suggested that for some age groups there was no relationship between the the FTII and later outcomes, it was decided that the t test would be computed with a two-tailed significance level.

fact, novelty scores for five items (out of 10 items) of this test did not exceed the chance level of 50% (see Study 1).

Next, the possibility that an aggregation of novelty scores would result in increased predictability of later mental outcome was examined. Correlations between FTII scores and later MDI were computed for infants for whom two novelty preference scores at two consecutive sessions were available (Table 5-5). As is clear from the table, the magnitude of the correlations at some ages, using the mean of two novelty scores obtained at two consecutive ages (20 novelty items), was higher than the correlations between one novelty score (10 novelty items) and later Bayley Scales. Nevertheless, all correlations between mean novelty scores and the Bayley Scales at 18, 21 and 24 months of age were nonsignificant.

Table 5-5

Correlations Between the Mean of Novelty Scores at Two Ages and Later Bayley Scales at 18, 21 and 24 Months

Childhood age Bayley Scales	18 months		21 months		24 months	
	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)
Novelty scores						
Mean novelty scores (27 & 29 weeks)	.39	.38			.48	.46
Mean novelty scores (29 & 39 weeks)	.30	.14			.02	.37
Mean novelty scores (39 & 52 weeks)			.56	.35	-.13	-.13

Note: The number of subjects varied between 10 and 11 infants in each correlation. The MDI is Mental Development Index from the first edition of the Bayley Scale (Bayley, 1969) and the revised edition of Bayley Scales (Bayley, 1993).

(ii) *Bivariate correlations between novelty scores and IQ*: Correlations were calculated between the percentage of preference for novelty in all four age groups and the 2-year IQ as indexed by the Stanford-Binet (Table 5-6). Since some subjects could not pass the number of items necessary for establishing the basal level of the Stanford-Binet test, the number of items passed (Raw Score) rather than the Standard Age Score⁴ was entered into the correlation equation. As the table illustrates, there was no correlation between the percentage of novelty scores obtained before 12 months of age and the later Stanford-Binet score. The exception was a significant correlation between the 39-week FTII and Memory for Sentences. However, novelty scores at 52 weeks of age significantly predicted Vocabulary and Pattern Analysis factors. The correlation between 52-week novelty scores and General IQ was nonsignificant. The results, then, suggest that the 52 weeks test of the FTII is a better predictor of later intellectual functioning (compared with other age-related tests), both in terms of developmental scales and as a psychometric test.

5-3-4 The sensitivity and specificity of the Fagan test

The lack of evidence found for a reliable relationship between novelty scores and later outcomes at some ages, places in doubt the clinical utility of this measure. Although all infants involved in the sample were apparently normal babies (i.e., non-risk and healthy infants), at least at the first testing during infancy, it was still possible to investigate the clinical utility of the FTII for the sample, because mildly intellectually disabled infants may not necessarily be detected at such an early age. In order to investigate this aspect of the test, all novelty scores for each of the four age groups and later Bayley MDI (Bayley, 1993)

⁴ The correlation between the Raw Score and Standard Age Score in all subtests was high and ranged from 0.93 to 0.1. Very similar correlations were obtained when Standard Age Scores and novelty scores were correlated.

Table 5-6

Correlations Between Novelty Preference Scores and Later Stanford-Binet Subtests at 24 Months of Age and (the Number of Subjects) in Each Correlation

Infancy age	27-week novelty score	29-week novelty score	39-week novelty score	52-week novelty score
Stanford-Binet test				
Vocabulary	.31 (12)	-.53 (10)	.11 (15)	.46* (20)
Comprehension	-.17 (12)	-.19 (9)	.12 (12)	-.16 (20)
Pattern Analysis	.42 (14)	-.40 (11)	-.32 (15)	.49* (20)
Copying	.26 (12)	.08 (9)	.20 (15)	.21 (20)
Memory for Sentences	.18 (7)	<i>a</i>	.68* (9)	-.47 (10)
General IQ	.30 (14)	-.56 (11)	.12 (15)	.33 (21)

Note: *a* The number of subjects involved in correlation for this test was only 4 infants. * $p < .05$ (two-tailed).

at 24 months were rank ordered from the lowest to the highest scores. (This age was selected because generally the correlation between novelty scores and later IQ is higher and stronger for older ages [e.g., Rose et al., 1989]). The novelty scores and MDI were then dichotomized. Next, the mean Fagan test scores $\leq 53\%$ and MDI ≤ 70 were compared. (The mean novelty scores of 53% or less is selected by Fagan as an indication of the existence of intellectual disability; the MDI score of 70 or less is traditionally viewed as grouping infants for intellectually disability).

As defined, there were four subjects with low novelty scores and three subjects with low MDI. None of the four infants who had novelty scores equal to or less than the cutoff point was found to have an MDI equal to or below 70 at a later age. Similarly, none of the

infants who had an MDI ≤ 70 was among the infants who scored less than 53% on the FTII (Table 5-7). However, if an MDI less than 85 (this cutoff point typically is viewed as below average) and novelty score less than 54.5% (this cutoff point is selected by Fagan as identification of suspected infants) are used as cutoffs, three out of eight low scoring (i.e., equal or less than MDI 85) children at 24 months were correctly detected by low novelty scores in infancy (i.e., 49.6%, 52.7% and 53.3%), indicating 37.5% sensitivity for the FTII. In comparison to its sensitivity, the FTII was highly specific in identifying 91.7% (55 out of 60) of children with MDI > 85 (Table 5-8). In fact, using a FTII score of ≤ 54.5 and an MDI score of ≤ 85 , three of five infants predicted to be at risk during infancy had a low MDI score at 24 months, and 58 of 63 children predicted to be normal were found to be so (see also case study in footnote⁵). The results, therefore, did not support 100% sensitivity of the FTII as described in the normative data (Fagan & Detterman, 1992).

5-3-5 Predictive validity of exposure time and number of fixations to later mental outcomes

The predictive validity reported so far is based on data obtained from the test phase. Correlations were also computed between two measures from the familiarization phase -

⁵ Case study of a Down Syndrome infant: During data collection, S., a Down's Syndrome boy, was referred by the Down Syndrome association. The infant was born three weeks premature with a birth weight of 3100 grams. At 9 months of age (corrected for prematurity) the infant was tested with the corresponding age-related procedure for the Fagan test. The infant scored a mean of preference for novelty of 58.3%. The MDI and PDI for the baby were 69 and 66 respectively (first edition of the Bayley scales). Parents were invited to participate for another testing at 12 months. The 52-week FTII also was administered. The percentage of preference for novelty was 68.4% with seven novel items preferred significantly in the direction of novelty, suggesting no mental retardation in terms of preference for novelty. The MDI and PDI for the second testing with the previously mentioned Bayley Scales were 50 and below 50 respectively. S. was quite cooperative during both sessions and he seemed to enjoy testing. Although S. was diagnosed as a Down Syndrome case, he again was tested at 21 and 24 months of age for prediction of later mental outcomes. The results as expected at 24 months was a MDI of 62 with the first version (below 50 with the new version) of the Bayley scales. The case study described here and in the results section, provides a missed outcome for the Fagan test.

Table 5-7

Contingency Table of the Success of the FTII in Predicting Normality and Deficits in Intelligence (Using Cutoff Point $\leq 53\%$ for the FTII and ≤ 70 for the MDI of the Bayley Scale at 24 Months of Age)

		Bayley Scales (Outcome)	
		Low score MDI ≤ 70	High score MDI > 70
Fagan Test (Predictor)	Low score FTII $\leq 53\%$	0	4
	High score FTII $> 53\%$	3	61

Table 5-8

Contingency Table of the Success of the FTII in Predicting Normality and Deficits in Intelligence (Using Cutoff Point $\leq 54.5\%$ for the FTII and ≤ 85 for the MDI of the Bayley Scale at 24 Months of Age)

		Bayley Scales (Outcome)	
		Low score MDI ≤ 85	High score MDI > 85
Fagan Test (Predictor)	Low score FTII $\leq 54.5\%$	3	2
	High score FTII $> 54.5\%$	5	58

exposure time and number of fixations, and later developmental scores and Stanford-Binet scores. All exposure times across all novel problems that have a familiarization phase (three novel problems of the FTII have no familiarization) were standardised. A single measure of

exposure time (the average of exposure time across all novel problems) for each subject in each age group was correlated with mental outcomes (Bayley Scale and Stanford-Binet test) between 18 and 24 months of age (Table 5-9).

Although most correlations between the 39 and 52-week FTII and later outcome were negative, a direction that was expected, only the relationship between exposure time at 39 weeks and MDI (Bayley, 1993) at 21 months was statistically significant ($r = -.48$, $p < .05$). The negative direction of the correlations suggests that children who scored lower on childhood tests had a longer time off task period during the familiarization phase. The number of fixations (not reflected in the table) was not found to be consistently related to outcome at 24 months, either in terms of the Bayley Scales or the Stanford-Binet test.

Table 5-9

Correlations Between Exposure Time and Later MDI and General IQ

Childhood age Infancy age	18 months		21 months		24 months		
	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)	Genera IIQ
27 weeks (N = 15-16)	.16	-.03			.26	.30	-.12
29 weeks (N = 12)	-.22	-.00			.20	.10	-.17
39 weeks (N = 19-20)			-.34	-.48*	-.42	-.32	-.36
52 weeks (N = 21)					-.03	-.00	-.12

Note: The number of subjects involved in correlations between exposure time and general IQ were 14, 11, 15 and 21 for the 27, 29, 39 and 52 week tests respectively. * $p < .05$. (Two-tailed).

5-4 DISCUSSION

The present study was the first independent study to investigate the predictive validity of all four tests of the FTII in a population of Australian infants. Analyses suggest

that individual differences in the measure of preference for novelty obtained from the visual recognition memory test of the FTII do not reliably predict later intellectual functioning. Correlations were generally low, and those statistically significant were found in the oldest age group only. The results were not consistent across all types of intellectual outcomes. Moreover, at the individual level, the rate of misclassification was relatively high. This suggests a conservative interpretation of the results from such measures, at least for infants up to 2 years of age. With respect to exposure time, the measure obtained from the familiarisation period, only the 39-week exposure time predicted later MDI at 21 months of age. No correlations were found between other measures of the familiarisation phase, number of fixations and later mental outcome.

Specifically, with the exception of an isolated significant correlation between the 39-week FTII and Memory for Sentences at 2 years of age, the first three novelty tasks of the FTII (i.e., 27, 29 and 39 week tests) had no predictive validity, in terms of the Stanford-Binet test or MDI scores, at least for the normal sample from a population of middle class parentage. Rose et al. (1989) also found no relationship between novelty scores (using their own experimental stimuli but the same procedure) obtained for full-term infants at 7 months and later MDI at 1, 1½ and two years. However, the results are not consistent with Thompson et al. (1991), who reported a significant correlation between novelty scores and later MDI at 24 months. In reference to Thomson et al., it should be noted that a composite score was used in the bivariate analysis. They used a combination of six novelty items at 7 months and four novelty problems obtained at 5 months.

There was some weak evidence of predictability of later intellectual outcome in terms of both the psychometric IQ test and developmental scales, for the 52-week test. Vocabulary and Pattern Analysis tests from the Stanford-Binet, as well as the 24-month

Bayley MDI, were significantly predicted by the 52-week FTII. Although in the present study the correlation between MDI and general IQ at 2 years of age was high ($r = .83$, $p < .001$), the relationship between general IQ and the FTII did not reach significance, suggesting inconsistency in the predictive validity of the test.

Furthermore, the predictability of novelty scores did not change as a result of using aggregate novelty scores obtained from two consecutive ages. Generally, larger composite novelty preference scores have resulted in higher prediction coefficients (e.g., Rose, et al., 1988a), but this was not the case in this study. This was probably the consequence of the poor internal consistency and the low or nonexistent test-retest reliability of the FTII. The unpredictability of the FTII is therefore at odds with evidence indicating that novelty preference scores taken at this age predict later mental functioning (e.g., Colombo et al., 1989).

5-4-1 Explanations for low predictability of the FTII

While the number of novel problems for all age groups was the same, and all subjects came from a relatively similar social background, it is not clear why the tests administered to younger infants did not predict later outcomes, even at 2 years of age. Several explanations for the unpredictability of the FTII are possible. One possible explanation originates from the theoretical background and/or psychometrical properties of the task. The low predictive validity of the FTII is suggestive of a "... lack of a clearly articulated theory for its design and validation" (Benasich & Bejar, 1992, p. 163). Alternatively, it might be argued that the psychometric properties of the test are not systematically standardised. The present research, consistent with other studies (e.g., Andersson, 1996; Jacobson et al., 1992), failed to achieve a reasonable internal consistency and test-retest reliability for the FTII, precluding substantial predictive validity.

A second possibility is that longer fixation to the stimuli, particularly for younger infants, may not be an indication of the extent of the infant's processing of information contained in the target. As suggested by Malcuit et al. (1988), variation in the length of visual fixation may be determined by temporary fluctuations in motivation, rather than individual differences in the processing of available information.

It is also possible that the rapidity of developmental change in various cognitive functions during infancy might contribute to the low predictability of the novelty task. That is, the novelty preference score may not be a reliable indicator of an infant's cognitive skills at some stages of development. On the other hand, the childhood age at the follow-up testing is also important. Preference for novelty may be a reliable predictor of later IQ, but not in early childhood because general intelligence becomes more stable in later childhood (Humphreys & Davey, 1988). This will be discussed further in Chapter 9.

Characteristics of the current sample may provide another plausible explanation for the poor predictive validity of the FTII found here. Laucht et al. (1994) have found that prediction from novelty tasks is generally higher for at-risk infants than for non-risk infants. The sample in the present study was a homogeneous group with a restricted IQ range, at least during infancy, and was drawn mostly from middle class parents, among whom the incidence of attentional problems is presumably lower than in a sample with a very low socioeconomic status (Rose et al., 1988a). Further support for this explanation is provided by the data from the correlational analysis. For example, excluding one low score infant from the 27 week test reduced the correlations between novelty scores and subsequent tests to zero.

Before further discussion concerning the clinical utility of the FTII, two aspects of the data deserve more consideration. First, the concurrent validity of the visual recognition

memory task (FTII) in terms of a traditional developmental scale was not supported in the present study. Although most correlations between novelty scores and concurrent MDI and PDI were positive, the direction that was expected, none of the correlations were significant. Similarly, the relationships between the number of fixations and exposure time and the MDI or PDI were not consistent across all age groups, suggesting that information processing measures are independent of measures of the infant's general competence, at least in the first year of life.

Second, is the link between novelty scores and subsequent specific cognitive skills that might be predicted by novelty tasks. Data from Jacobson et al.'s study (1992) suggest that the mechanism underlying an infant's ability to process information consists of a number of relatively independent cognitive processes, rather than a single ability. The significant correlations between 52-week novelty preference score and later Vocabulary and Pattern Analysis factors from the Stanford-Binet, suggest that both verbal and nonverbal skills might be predicted in the child's performance on subsequent intelligence tests. However, Fagan and Knevel (1989) have argued that the FTII was more strongly related to the verbal tests of the Stanford-Binet than to the nonverbal items. Thompson et al. (1991), and Rose et al. (1991b) have also reported that novelty scores are related to later language as well as to general IQ. The relationship between the 52-week FTII and MDI at 2 years of age suggests that, to the extent that the MDI is a reliable measure of *g* factor the novelty task may tap some of those cognitive skills underpinning a general factor during infancy.

5-4-2 Critique of the clinical utility of the FTII

Fagan and Detterman (1992), and Fagan et al. (1985, 1986) claimed that the real worth of the Fagan test lies in its clinical utility for the diagnosis of infants who potentially will show disability in their intellectual activities. The present study indicates that the

clinical utility of the test is low, particularly in relation to its sensitivity (ability to identify infants who are at risk for subsequent intellectual disability). Results have confirmed the relatively high rate of misclassification that has occurred in other studies (e.g., Cohen & Parmelee, 1983). It is recognized, however, that the sample in the present study consisted of nonrisk infants and this limits the evaluation of the clinical utility of the test. The number of infants who were misclassified as at-risk for later retardation (the test misclassified four normal infants), was relatively high. In addition, the FTII in infancy did not identify the subjects who had an MDI equal to, or less than, 70 at 2 years of age. Similarly, the test could not identify a Down's Syndrome child who was tested twice during infancy and scored more than 65% of preference for novelty. Overall, the sensitivity found here for the FTII, as applied to a non-risk infant population, was much lower than that reported for normative data. The results of normative data suggest that the FTII is 100% sensitive to the presence of later retardation. However, the low sensitivity found here is in accord with the rate of misclassification reported by other researchers (e.g., Andersson, 1996). Andersson reported a sensitivity as low as 22% in the classification of infants with low verbal IQ, and the sensitivity of the test found here did not exceed 37.5%.

In summary, the results of this study do not provide evidence for predictive validity of the FTII. For the visual recognition memory tasks, at least in terms of the FTII, prediction may not be possible before 52 weeks of age, and at 52 weeks the predictability is still low and may not be useful in a practical sense. The preference for novelty at this specific age provides an estimate of later developmental scales and only some factors, but not general IQ, as indicated by the Stanford-Binet test. The rate of misclassification questions the clinical utility of the test, supporting McCall's suggestion (1988b) that there is still not enough data to justify marketing this type of task. Further, the lack of predictability

for the FTII, at least for some ages, highlights the validity of the issue raised by Lecuyer (1989) and Malcuit et al. (1988a, b). Lecuyer (1989, p. 149), for example, suggested that most negative results did not get published because they did not reach "... this magical level [.05] of significance". Considering these results, it seems that the FTII and the theory on which the test is based require further research. The test should also include the whole age range of the second half of the first year, not just four specific ages. Normalizing the test for only these specific ages makes the scheduling of testing difficult.⁶

⁶ An article based on this chapter has been presented in the 24th Annual Experimental Psychology Conference, Geelong, Australia. The abstract has been published in: Tasbihsazan, R. Nettelbeck, T., & Kirby, N. (1997). EPC 97 Abstracts for oral presentations, p. 52. Geelong: Deakin University (see Appendices).

CHAPTER 6

PART I

PREFERENCE FOR NOVELTY AT 16 MONTHS AS A PREDICTOR OF LATER MDI AND IQ IN EARLY CHILDHOOD

6-1 INTRODUCTION

It has been suggested that information processing measures, in particular the percentage of preference for novelty, have predictive validity if taken during the first eight months of life. Beyond this specific age range, the predictive validation of processing information measures would diminish (Slater, 1993). It appears that this conclusion was reached, not because of the failure of studies to predict later IQ for older infants, but because of the number of studies in which prediction was made for younger babies. That is, the number of studies involving subjects older than one year of age, is rare. For example, in the 31 studies reviewed by McCall and Carriger (1993), there was only one sample consisting of infants as old as 12 months. In their interpretation of the results of the studies, McCall and Carriger concluded that “... the possible decline in predictability after 8 months, if substantiated by further study, would be more unusual and provocative” (p. 76). Using visual recognition memory and cross-model transfer task in one year old infants, Rose et al. (1991a) found that, while performance on cross-model transfer was significantly related to later IQ, visual recognition measures obtained at the same time were not consistently predictive of later outcomes. They concluded that “it is not clear why the problems of visual recognition given at one year were less predictive than those given at 7 months” (p. 734),

and speculated that the stimuli (three-dimensional geometric forms) used in their study were not sufficiently complex to capture more sophisticated cognitive functioning.

Similar to McCall, Slater (1995) also proposed that the temporal window of opportunity for measures of visual attention is somewhere between birth and 8 months of age. Once again there was only one 12-month-old sample in his review of 24 studies involving visual recognition memory tasks. Slater (1995, p. 84) suggested that "From the middle of the first year, or thereabouts, we need other measures of cognitive abilities".

One reason for such a low number of studies involving visual recognition memory in infants older than 1 year of age is the difficulties that researchers confront in using this procedure with older infants. For example, Fagan (1990) recognized that testing infants older than 12 months with a visual preference test is difficult. Most infants in this age range have begun to walk alone (Myers, 1986) and are more active. In practical terms, the feasibility of completing a study which requires an infant's cooperation, with few motor activities, diminishes.

Apart from the practical difficulties associated with the visual recognition memory task in more mature infants, it is also of interest to examine whether a reliable estimate of subsequent mental functioning can be obtained within this age range (i.e., beyond 12 months of age). Theoretically, prediction at some specific age may not be possible due to the rapidity of development and qualitative changes during the first two years of life. For example, the results of Wilson's study (1983) suggest that a transition in cognitive abilities occurs during the first half of the second year. Similarly, Piaget theorised that this approximate age (after 12 months) marks the onset of a qualitative transformation in infant competencies. Infants enter the "Tertiary Circular Reaction" stage in which her/his

conception of causality is developed and the infant tries to examine new properties in objects and events.

6-1-1 Predictive validity of novelty task beyond 12 months of age

Although there are a number of studies in which 12 month (and older) infants were tested with measures of information processing other than habituation and visual recognition memory (e.g., Gopnik & Meltzoff, 1987), the studies that have used these two procedures are rare. In an early study, Faulkender, Wright and Waldron (1974) found that habituation of visual attention occurs in 3-year-old infants. They suggested that while habituation is present "... it may have begun to give way to other processes of learning that predominate in older children" (p. 1009). In another study Johnson and Brody (1977) assessed the pattern of relationships between visual attention, sensorimotor development, play tempo and motor activity in one year old infants. They reported that not only gender, but also the state of the infant (cooperation and/or irritability or drowsiness), mediated the relationship between some measures of visual habituation and measures of sensorimotor development, play tempo and locomotor activity. However, this study was not concerned with the predictive validity of attentional measures.

Study 2 found that the predictive validity of the FTII for younger babies was disappointing, indicating no predictive validation for the first three tests (i.e., 27, 29 and 39 weeks) using the Fagan procedure. There was, however, weak evidence of predictability from the 12-months test to later MDI at 24 months. Older babies, possibly due to a more mature brain, should concentrate more efficiently on a task, and be better able to encode, retrieve and discriminate familiar and novel stimuli. In spite of some predictive validity reported for a novelty task administered prior to 12 months of age, the validity of this procedure beyond 12 months of age is still unknown.

Apart from preference for novelty, other measures obtained from novelty tasks may also bear some predictive validity to later IQ for older infants. Exposure time, for instance, can be viewed as a cognitive aspect of an infant looking behaviour. That is, this measure could be regarded as a measure of a latency period during which an infant attempts to memorise (or rehearse) the familiar targets, an activity similar to those cognitive functions in older children when they try to memorise a specific target. Miller and Vernon (1992) held that individuals with deficits in short-term memory need to spend more time and effort maintaining information in an active state by activities such as rehearsal of the information. If this were the case, a negative correlation between this measure and later mental outcome, particularly memory, would be expected. In fact, Rose et al. (1989) found that exposure at 7 months for full-term and pre-term subjects showed a modest negative correlation with later outcomes, including Full Scale IQ (WPPSI) at 5 years of age. They also found that length of exposure time was associated with medical factors such as low birth weight, short gestation age and smaller head circumference (Rose et al., 1988b).

The present study was an attempt to assess the predictive validity of the measures obtained from a visual recognition memory task at 16 months for the performance on MDI/IQ tests at 27 months of age. In addition to recruiting a number of healthy normal infants, two Down's Syndrome subjects were included in the current study. These inclusions made it possible to investigate the effect of outlier subjects on the predictive value of the novelty task. Generally, Down's Syndrome babies demonstrate extreme deficits in their performance on visual tasks (Cohen, 1981; Miranda & Fantz, 1974; Ross & Ross, 1981). Children with Down's Syndrome have more basic differences in CNS integrity and require more familiarization time to acquire knowledge of a familiar stimuli (MacTurk, Vietze,

McCarthy, McQuiston & Yarrow, 1985). For comparison with normal infants, familiarization time was the same for all infants in the study.

6-1-2 Time-fixed procedure for measuring novelty preference scores

Traditionally, the magnitude of preference is assessed during the test phase in which a familiar stimulus is paired with the novel target. The time for the test phase expires during this period and it is difficult to justify that the novelty score is a result of the tendency to look at a novelty stimulus or simply a chance performance. This is of particular importance because many infants (see Study 1) do not study both stimuli in the test phase (see also McCall, 1994). In these cases, neither a score of 0% nor a novelty score of 100% should be taken as an indication of information processing. If the procedure were changed so that the time for the test period had to be accumulated, an infant would have more opportunity to select the stimulus at which s/he wished to look. Because time is not expiring, a bright infant can deliberately choose the right stimulus (the new one) with this procedure, even if s/he failed in his or her first attempt and looked at the familiar target. In the present study, a time-fixed approach was used to capture the infants' attention to both familiar and novel stimuli (see procedure) in the belief that it might increase the predictability of this type of task. Using this scoring procedure reduced chance performance and therefore one would expect a higher internal consistency across the novel problems as well as a higher preference for novelty score.

While the extension of the novelty task to infants older than 1 year of age is the main concern of the present study, the examination of developmental trends for attentional measures in this age range was also of particular interest. Although no developmental trend has been found for novelty scores and exposure time during the first year of life (see Study 1), it is plausible to suggest that the skills associated with these measures undergo

developmental change in the second year of life. As a consequence, older infants would show more self-regulation and have a better attention span during test taking, resulting in decreasing the number of fixations and requiring less exposure time during the familiarization period.

The main concerns of the current study, then, were threefold. First, an attempt was made to extend the age range for the novelty task to 15 to 18 months of age, and to evaluate if testing beyond 12 months was feasible in practice and theoretically useful for predicting later intelligence in childhood. Second, a time-fixed procedure was used to assess if this scoring system can significantly increase the magnitude of preference for novelty, as well as internal consistency of the novelty task. Third, developmental change for attentional measures during the first 16 months of life was examined. The practical implications of these findings were also of interest. For those infants who become disadvantaged for reasons such as head injury or encephalitis in the later infancy stages, it would be possible to predict if they were at risk of developing a syndrome for later retardation.

6-2 METHOD

6-2-1 Subjects

A total of 16 infants (9 female, 7 male) participated in Study 3. Fourteen healthy children were drawn randomly from an original cohort ($N = 30$) and two Down's Syndrome infants were included. The remainder of the cohort was assigned to Study 4. The subjects' age range was 15 to 17.5 months ($M = 16.1$, $SD = 0.93$). Subjects were recruited from different sources, mainly by advertising in a free, locally distributed paper. A notice was also forwarded to the Down's Syndrome Association in Adelaide, South Australia. All but two subjects were born in Adelaide, South Australia between August 1992 and March 1993.

The other two subjects were born overseas within the same period of time. With one exception (one subject was living with his separated mother), the subjects' parents were married and all subjects were reared at home.

(i) **Birth status:** Two infants were born with a congenital syndrome (i.e., Down's Syndrome). The remainder were all healthy infants with no prenatal or obstetrical complications, and with normal vision as indicated by their personal health records completed at birth, and thereafter at routine examinations. Most subjects (56.3%, $n = 9$) were first born babies. Description of infant birth data and the parental characteristics of the subjects are displayed in Table 6-1.

Table 6-1

Birth Status and Demographic Characteristics of the Sample in Study 3

Demographic Variables	N	Mean	SD	Range
Gestation age (weeks)	16	39.4	1.6	36.5-42
Weight (kilograms)	16	3355	585	2.2-4.2
Maternal age (years)	15	31.1	4.8	25-40
Maternal education (years)	16	13.9	2.5	11-18
Maternal job status (7-point scale)	11	3.3	.92	1.8-4.8
Paternal age (years)	15	35.5	7	25-49
Paternal education (years)	16	15.7	3.6	10-21
Paternal job status (7-point scale)	14	3.1	1.2	1.5-5.8

(ii) **Socioeconomic status:** Developmental research in Australia enjoys the advantage of the richly diverse blend of cultural backgrounds in this country (Peterson, 1996), and the present study was no exception. The infants participating in the study came from different cultural backgrounds. The majority (62.5%), were of Australian parentage but the parents of the other subjects were from non-English speaking families who

immigrated to, or resided in, Adelaide. Parental education and job status were used as indices of socioeconomic status (SES) for subjects. The mean schooling years for parents was 14.8 years ($SD = 2.6$). Nine subjects' mothers graduated from universities or had at least two years of tertiary education. Social standing was also assessed by a 7-point rating scale of parents' occupational status (Daniel, 1983), and this indicated that the majority of subjects came from middle to upper middle class parents.

6-2-2 Measures during infancy

Two different types of tests, each reflecting a specific domain of an infant's current functioning, were administered to all subjects at approximately 16 months (69 weeks) of age. The first was a visual recognition memory task that is generally interpreted as a measure of an infant's ability to encode, discriminate, retain and retrieve accessible information, as well as the speed with which the infant can perform those functions. The second measure was a standard infant developmental test that is known to be representative of infant developmental milestones. The stimuli and design for the experimental visual recognition memory task used in this study are described below. The Bayley Scales of Infant Development (Bayley, 1969) has been fully described in Chapter 4.

All subjects were successfully tested during infancy. However, two infants did not complete all five sets of the novelty preference task. Three or four sets of novelty items only were administered to these subjects due to fussiness, high motor activities and/or boredom after administration of the first few items.

(i) *Visual recognition memory task*: In order to measure an infant's differential preference for novelty, a visual recognition memory task was developed together with a computerized testing and scoring system. The task consisted of five sets of novelty problems (each set had two novel problems) measuring preferential fixation in three

different conceptual categories - animal, abstract patterns and human faces. Based on an item analysis of different stimuli given at 3 to 7 months, Fagan and Montie (1986) argued that human faces are more predictive than abstract patterns. However, for older children it is possible that other categories of stimuli (e.g., a life-like picture of an unfamiliar animal) could account for a larger proportion of variance in later outcomes. Each novelty problems consisted of a set of four conceptually related pictures. Two identical pictures were used for familiarization and two other pictures from the same category were used for measuring the infant's preferential attention. According to discrepancy theory, novel and standard stimuli that are clearly related in their physical or "conceptual" form are more likely to stimulate infant attention (McCall, 1988a). All stimuli, as described below, were two-dimensional drawings sized 18 x 18 cm on laminated matt white cards (see Figure 6-1).

The first set of stimuli was defined as an animal category. Different categories of animals and birds have previously been used as stimuli in infant attentional studies (e.g., Catherwood et al., 1995; Roberts, 1988; D. Rose et al., 1986). Evidence indicates that by 12 months of age an infant is able to process categories of pictures or objects, such as drawings of animals (Cohen, 1988a). The set consisted of three coloured lifelike drawings of three different felines, from species believed to be less familiar to infants (leopard, tiger, ocelot).

The second set of stimuli comprised three black and white computer generated drawings of men's faces on a white background. All stimuli were drawn using Clipart from Micrografx Charisma, Version 2.1.

The next set of stimuli consisted of coloured abstract patterns (Set 4). Abstract patterns are frequently used as stimuli in infant's looking behaviour studies and even young infants are able to discriminate a familiar pattern from a new one (e.g., Thompson et al., 1991). The two patterns used for familiarization were coloured patterns on a white

Figure 6-1

Representation of Novelty Items of the Visual Recognition Memory Task



















	Left	Right
Novelty Item 1		
1 - Familiarization Trial		
2 - Novelty Trial No.1		
3 - Novelty Trial No.2		
Novelty Item 2		
4 - Familiarization Trial		
5 - Novelty Trial No.1		
6 - Novelty Trial No.2		
Novelty Item 3		
7 - Familiarization Trial		
8 - Novelty Trial No.1		
9 - Novelty Trial No.2		

Figure 6-1

Continued

Novelty Item 4

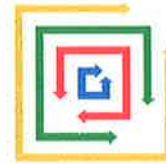
Left

Right

10 - Familiarization Trial



11 - Novelty Trial No.1



12 - Novelty Trial No.2



Novelty Item 5

13 - Familiarization Trial



14 - Novelty Trial No.1



15 - Novelty Trial No.2



background, comprising two groups of five brightly coloured arrows. The two other patterns used for novelty problems were similar in size, colour and the number of elements. Each consisted of four increasingly smaller squares with a corner gap in the outline, nested one inside the other. The only difference between the two novel targets was arrow heads on the incomplete corners of the four squares in one pattern, but not the other. All abstract patterns were made from Chartpak colour film on a white background paper, and were matt laminated.

The last two sets (Set 3 and Set 5) of stimuli were human faces chosen from the FTII pictures. There were three different pictures of a baby's face and a picture of a woman's face with a different spatial orientation. All pictures were monochrome on a white background. It is now well documented that by 6 to 7 months of age infants are easily able to differentiate between pictures of human faces of the same age or sex (Fagan & Shepherd, 1982).

6-2-3 Apparatus and procedure

The test laboratory and equipment were described in Chapter 4. An IBM computer and its components, a computerized testing and scoring system programmed specifically for this study, were also available. Stimuli were presented by means of the portable stimulus presentation stage used in the Fagan test.

The tests of preference for novelty and developmental examination were administered to all subjects aged 16.1 months ($SD = 0.93$), adjusted for prematurity. The majority ($n = 12$) of parents preferred their children to be tested at home. The tests were scheduled for one session. However, three subjects, including two Down's Syndrome children, were tested over two sessions, due to a short attention span and fatigue.

Each infant was tested firstly on the visual recognition memory task and then cognitive abilities were assessed using the Bayley Scales. Attempts were made to test the children at the time when they were judged to be alert and cooperative. Each infant sat on a caregiver's lap at the desk (or sitting by him/herself on a high chair supported by the caregiver) facing the viewing stage. The parents were instructed to facilitate their infant's view of the pictures but were requested not to prompt their child (naming and describing the pictures), not to point to the stimuli or interfere with the infant's looking behaviour. As a warm up procedure, each infant was exposed to two pictures (for about 30 seconds) fixed on the stage and encouraged to look at them. In the test situation, each trial started with a familiarisation period in which the infant was exposed to two identical pictures fixed on the stage for a predetermined time (10 seconds), and followed by two 6-second test trials. In the first test trial, a previously seen target (familiar stimulus) was paired with the first new stimulus of the same set, and both were presented to the infants simultaneously. In the second test trial, the familiar target was paired with the next new stimulus of the set. The stimuli remained exposed until the infant accumulated the preset amount of fixation time, both in the familiarization and test phases. Based on the aim of the research, study time was set equally for both normal and Down's Syndrome infants. This procedure was necessary because it has been found that if the study time is increased, Down's Syndrome and other at-risk babies can also differentiate the stimuli (Miranda, 1970; Mundy et al., 1983).

The order of stimuli presentation (see Figure 6-1) was fixed for all subjects to keep the effects of "fatigue and proactive and retroactive interference" constant (Rose et al., 1988a, b). Pictures of animals were presented first. They were followed in order by computer generated drawings of men's faces, babies' photographs, abstract patterns and a woman's face with different orientations. The novelty scores across all problems were then

averaged to increase the reliability and the predictive validity of the task (Colombo et al., 1988; Fagan et al., 1986).

The novelty problems used in this study were administered slightly differently from the method recommended in the manual for FTII. Typically, after the familiarization period, an infant is exposed to just one novel stimulus, counterbalanced across the right and left position in two tests trials. In addition, the preset time of the test phase starts with the infant's first fixation and runs out, regardless of how long the infant fixates on the stimulus and how many glances occur. In this study, however, the infants were presented with two different novel stimuli, one in each trial and, similar to the familiarization period, the subjects had to accumulate the 6-second preset time for each of two test trials.

The Bayley Scales were also administered after a short break and a mental scale, a psychomotor scale form were filled out for each subject. Overall, five different dependent variables were drawn from the testing session during infancy. They were: a novelty preference score, the number of fixations the infant made to accumulate the familiarisation time, the exposure time, a mental scale and a motor index. All subjects were tested by the researcher.

6-2-4 Follow-up measures

The subjects were visited and retested at 27 months of age (all tests scheduled within two weeks after the 27 months birthday, corrected for prematurity). At follow-up, both the Bayley Scales (Bayley, 1969, 1993) and an abbreviated form of the Stanford-Binet were given, (two Down's Syndrome children were not given the Stanford-Binet test). Both of these tests were scheduled for one session. However, two testing sessions were given if necessary (one normal subject and one of the Down's Syndrome babies had two testing sessions). In order to prevent the effects of fatigue and/or learning because of the similarity

between some items of both tests (e.g., some items of Pattern Analysis), the Bayley Scales was given first and the Stanford-Binet at the end of the session after a short interval. Follow up measures for one subject were missing due to withdrawal from the project.

6-3 RESULTS

6-3-1 Infancy measures

A preliminary analysis was performed to ascertain if there were differences on the infancy measures due to ethnicity and gender. A series of *t* tests revealed no significant differences between the two genders (*t* ranged from .87 to .54, all *ps* *NS*) or between Australian infants and infants from non-English speaking backgrounds (*t* ranged from 1.08 to .94, all *ps* *NS*) across all the dependent variables. The latter group of nonsignificant *t* tests implies that, at least at this age, attention measures and developmental indexes are not affected by cultural background. This result supports the conclusions reached by Cohen and Parmelee (1983) but some researchers have found significant differences in attentional scores for these two groups (e.g., Sigman et al., 1986). However, given the small number of subjects in each group this result should be interpreted cautiously.

(i) **Novelty scores:** Means and standard deviations for each novel item are presented in Table 6-2. A two-tailed *t* test procedure revealed that all but one novelty problem score exceeded the chance level of 50% in the direction of preference for novelty. The exception was novelty item No.7, an abstract pattern stimulus. The mean percentage of preference for novelty was 63.8% (*SD* = 10.4) and differed significantly from chance ($t(15) = 5.31$, $p < 0.001$). In terms of mean preference scores, all subjects but two displayed differential preference in the direction of novelty. The exceptions were the two Down's Syndrome subjects. These two infants performed at chance level for some items and/or preference in

the direction of familiarity for most items, suggesting that the task was difficult for them and/or the study time was not sufficiently long to enable them to process the information contained in the stimuli. Using Cronbach's alpha, the internal consistency of the novelty scores was assessed. Alpha for the novelty scores was higher ($\alpha = .74$) than those reported for the FTII. However, excluding the Down's Syndrome subjects reduced alpha to zero (Table 6-3).

Table 6-2

Means, Standard Deviations and *t* Values of Percentage Novelty Scores for all Ten Novel Items

	Mean	SD	<i>t</i> -value
Novelty item 1	68.9	24.1	3.13**
Novelty item 2	65.9	17.0	3.74***
Novelty item 3	62.6	19.2	2.63*
Novelty item 4	65.8	18.8	3.36***
Novelty item 5	62.6	20.1	2.40*
Novelty item 6	61.0	20.1	2.11*
Novelty item 7	57.0	20.1	1.29
Novelty item 8	62.4	17.8	2.70*
Novelty item 9	66.2	18.2	3.30**
Novelty item 10	65.7	13.7	4.30**
Mean novelty score	63.8	10.4	5.31***

Note: The *t* values refer to the significance of the mean of each novelty score exceeding the chance value of 50% using a two-tailed *t*-test.

* $p < .05$. (Two-tailed). ** $p < .01$. (Two-tailed). *** $p < .001$. (Two-tailed).

Table 6-3

Internal Consistency Coefficients (Alpha) for Attentional Measures at 16 Months

Attentional measures	Down's Syndrome included	Down's Syndrome excluded
Novelty preference score	.74	-.11
Number of fixations	.77	.75
Exposure time	.90	.56

The bar graph obtained for novelty scores indicated that the preference scores fluctuated across different novelty items. Upon close examination of the data, however, it is evident that the fluctuation is mainly due to novelty scores for the two infants with Down's Syndrome (Figure 6-2 a and b). The probability plot departed significantly from normal with the inclusion of these two infants, but when they were excluded from the analysis, the Shapiro-Wilks test indicated that the distribution of novelty scores was normal.

A comparison with normative novelty preference scores was not possible since no previously study for this age is available. However, the mean of 63.8% was consistent with scores obtained in Study 1 for younger infants (i.e., 27 to 52 weeks) using the FTII. There were no developmental changes between the two novelty scores obtained at 12 and 16 months of age ($t(37) = .30, NS$)¹, or across all ages between approximately 6½ and 12 months (Study 1)².

(ii) *Number of fixations and exposure time*: The number of fixations to each stimulus during the familiarization phase and the time the infants spent to accumulate the preset study time (exposure time) are presented in Table 6-4. The number of fixations for older infants in the present study was significantly lower than for younger infants (i.e., 52 weeks) in Study 2 ($t(37) = 5.38, p < 0.001$). That is, older infants look at stimuli more steadily than younger ones. The internal consistency of number of fixations did not change as a results of inclusion or exclusion of Down's Syndrome infants.

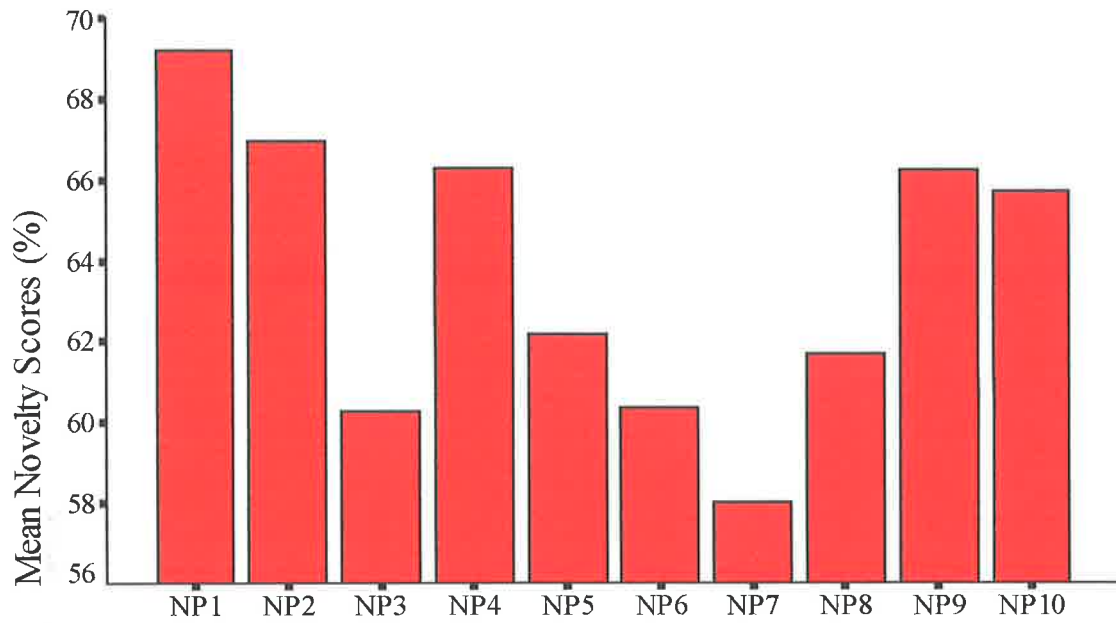
~~Exposure time was also significantly less for 16-month-olds than for 12-month-old infants ($t(37) = 3.26, p < .01$), an indication that older babies are more self-regulated~~

¹ In order to compare the magnitude of novelty scores without the inclusion of the Down's Syndrome subjects, another t test also was performed with the Down's Syndrome infants excluded from the analysis. Although the mean of novelty preference scores was higher (67.1%), there was no significant difference between this measure for 12 month and 16 month subjects.

² Although two different novelty tasks were used at ages 12 and 16 months, the theoretical assumption underlying both tasks and some stimuli were identical. That is, both measured an infant's ability to discriminate between familiar and novel targets.

Figure 6-2 a

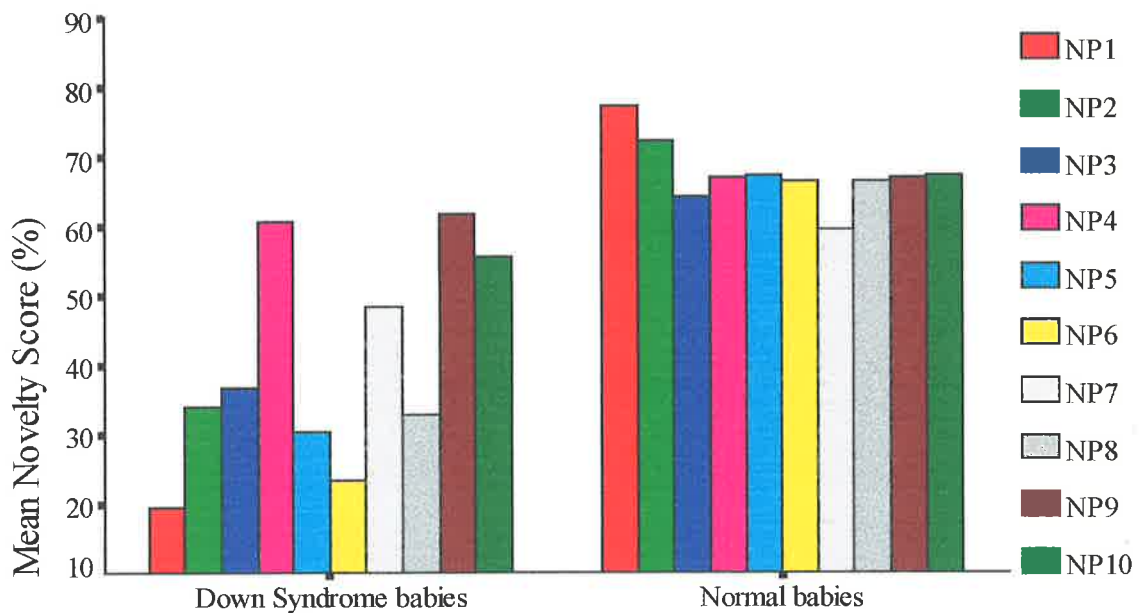
Bar Graph of Percentage of Preference for Novelty (N = 16)



NP 1, 2... = Novelty Items 1 to 10

Figure 6-2 b

Preference for Novelty in Normal and Intellectually Disabled Subjects



The Normality of Subjects

~~and more absorbed in the task. On average, an infant needs less than 4 seconds extra time to accumulate the 10 seconds familiarization time. This implies that, unlike novelty scores, there is a developmental change related to age for these two measures (see Figure 6-3).~~

The means and standard deviations for the developmental scales are also presented in Table 6-4. The MDI and PDI were 108.2 and 103.6 respectively. It was clear that the inclusion of the two Down's Syndrome infants decreased the MDI and PDI scores and increased the standard deviation.

Table 6-4

Means and Standard Deviations for Information Processing Measures and Developmental Indices

	Down Syndrome included (N = 16)		Down Syndrome excluded (N = 14)	
	Mean	SD	Mean	SD
<u>Information Processing Measures</u>				
Mean novelty score (%)	63.8	10.4	67.1	5.3
Number of fixations (per 10 sec.)	6.2	1.7	6.1	1.6
Exposure time (per 10 sec.)	13.8	5.9	12.4	1.8
<u>Bayley Scales (1969)</u>				
Mental Development Index (MDI)	108.2	20.2	114.4	11.9
Psychomotor Development Index (PDI)	103.6	19.7	110.3	7.2

6-3-2 Childhood measures

Table 6-5 presents the means and standard deviations of all predicted tests at the follow-up visit at 27 months of age. The mean IQ score for 27-month-old children was 110.6 ($SD = 12.2$). The Stanford-Binet test was not administered to the two retarded subjects. As is evident from the table (see also table's note), the inclusion of the two mentally retarded children again increased the standard deviation of the Bayley Scales. A further analysis indicated that while there were no gender differences in the childhood measures (t ranged from .79 to 1.5, all ps NS), there was a significant difference between cultural groups in some subtests of the Stanford-Binet (i.e., Vocabulary and Memory For Sentences) and in the Partial Composite SAS (IQ = 117 vs IQ = 103). This indicates that native English speaking children scored significantly higher than their peers from other cultural backgrounds. However, these results should be viewed cautiously since the number of subjects in each group is low.

Table 6-5

Means and Standard Deviations of the Bayley Scales and the Stanford-Binet Test at 27 Month of Age (Down's Syndrome Subjects Excluded, N = 13)

Variables	Mean	SD	Range
<u>Bayley Scales</u>			
MDI(1969)	125.2	15.6	63 - 145
MDI(1993)	108.5	11.5	49 - 125
<u>Stanford-Binet Test</u>			
Vocabulary	110.3	10.4	94 - 126
Comprehension	101.8	12.0	84 - 120
Pattern Analysis	104.3	11.4	88 - 126
Copying	110.0	11.4	96 - 124
Memory For Sentences	112.3	9.9	96 - 128
Partial Composite Score (IQ)	110.6	12.2	95 - 128

Note: The Stanford-Binet test was not administered to the two Down's Syndrome infants. With the inclusion of the two Down's Syndrome subjects, the mean MDIs for the first and the second versions of the Bayley Scales were 117 ($SD = 26.1$) and 100.5 ($SD = 23.5$) respectively.

6-3-3 Intercorrelations between infant information processing measures

It was of particular interest to investigate the degree of the relationship between information processing measures. There was a significant relationship ($r = -.61, p < .01$) between novelty scores and exposure time (see Table 6-6). That is, preference for novelty is predictable from exposure time. The relationship between the number of fixations and exposure time was also significant. However, the present study found no relationship between the number of fixations and novelty scores. This is inconsistent with Colombo et al. (1988), who proposed that short-looking infants are those infants who score higher in preference for novelty.

Table 6-6

Intercorrelations Between Attentional Measures at 16 Months (N = 16)

Measures	<i>r</i>
Novelty score and exposure time	-.61**
Novelty score and number of fixations	-.33
Number of fixations and exposure time	.52*

* $p < .05$. (Two-tailed). ** $p < .01$. (Two-tailed).

6-3-4 Concurrent validity of information processing measures with the mental development scale

To the extent that measures of information processing, particularly novelty scores, are indices of infant intellectual performance, some degree of relationship between these measures and developmental scores is expected. The correlation coefficients obtained here (Table 6-7) were highly significant between novelty scores and the MDI and PDI, as well as between exposure time and the MDI and PDI. However, without scores from the two Down's Syndrome infants, most correlations decreased and become nonsignificant. Statistical results did not indicate that short-looking infants were necessarily more intelligent, at least in this age period and in terms of developmental scores (MDI).

Table 6-7

Intercorrelations Between Attentional Measures and Measures Obtained From the Developmental Scale

Attentional Measures	Down's Syndrome included		Down's Syndrome excluded	
	MDI	PDI	MDI	PDI
Novelty scores	.85***	.91***	.45	.55*
Exposure time	-.61**	-.75***	-.27	-.35
Number of fixations	-.28	-.30	-.36	-.41

Note: N = 14 with the exclusion of the Down's Syndrome infants, and N = 16 with these infants included.

* $p < .05$. (two-tailed). ** $p < .01$. (two-tailed). *** $p < .001$. (two-tailed).

6-3-5 Predictive validity of the 16-month novelty score to the 27-month mental development scale

In order to examine the predictive value of the novelty task, a series of bivariate correlations was computed between preference for novelty and subsequent MDI (Bayley, 1969, 1993) at 27 months of age. Table 6-8 presents correlations between the infancy and childhood measures. Two separate analyses were performed due to the inclusion or elimination of data from the two Down's Syndrome infants. The significance levels and the magnitude of the correlations were changed substantially as a consequence of inclusion or exclusion of these children. When the intellectually disabled children were included, with one exception all category related novelty problems and the mean of the novelty scores across all problems significantly predicted later MDI scores (Bayley, 1993) at 27 months of age. That is, higher novelty scores at 16 months predicted a higher MDI at 27 months. The correlation coefficients for each category of problem ranged between $r = .33$ and $r = .82$. The exception was the abstract pattern stimuli, which was not correlated with mental outcomes. Excluding two Down's Syndrome children resulted in decreased nonsignificant correlations for each category related novelty problem and the mean of novelty scores for all novel items. The results, then, suggested that the prediction phenomenon may be dependent on the inclusion of intellectually disabled infants.

6-3-6 Prediction of later IQ from novelty tasks in infancy

Pearson product-moment correlations were computed in order to examine the relation between infant novelty scores and the Stanford-Binet subtests and General IQ. (Table 6-9). Bivariate correlational analyses were based on the number of items successfully completed by the children on the Stanford-Binet, rather than the Standard Age

Scores (SAS)³. As is evident from the table, there is only one modest but significant correlation of .58 between the mean novelty preference score and later Pattern Analysis of the Stanford-Binet test. However, the correlation between infancy novelty scores and later IQ (Partial Composite SAS) was not significant.

Table 6-8

Longitudinal Correlation Between Novelty Preference at 16 Months and MDI at 27 Months of Age

Outcome (Bayley Scales)	Down Syndromes included (N = 15)		Down Syndromes excluded (N = 13)	
	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)
Category of novel problems				
Animals	.82***	.81***	.50	.44
Faces (computer generated)	.46	.51*	.13	.22
Abstract patterns	.33	.47	-.42	-.18
Faces (pictures)	.70**	.81***	.00	.30
Mean of all novelty problems	.79***	.88***	.25	.48

Notes: The MDI (1969) & (1993) are Mental Development Indices from the Bayley Scales, Original Version (1969) and Revised Version (1993) respectively. * $p < .05$. (two-tailed). ** $p < .01$. (two-tailed).

*** $p < .001$. (two-tailed).

Table 6-9

Correlations Between Novelty Scores and 27 Months Stanford-Binet Test

Stanford-Binet subtests	<i>r</i>
Vocabulary	.15
Comprehension	.18
Pattern Analysis	.58*
Copying	.51
Memory For Sentences	.19
Partial Composite SAS	.31

Note: The two Down's Syndrome subjects were not involved in correlational analyses reported in this table. * $p < .05$. (two-tailed).

³ However, it should be noted that the correlations between the number of items passed and the Standard Age Score were very high (ranged from $r = .996$ to 1). A close examination of correlational analyses revealed that there were no substantial differences between either correlations based on the raw score of the Stanford-Binet or correlations involving Standard Age Score and infancy measures.

6-3-7 Prediction of later MDI and IQ from other measures obtained from the information processing task

Table 6-10 presents the correlations between the 16-month exposure time and later Stanford-Binet scale (subtests and Partial Composite SAS) at 27 months of age. There was also a moderate negative correlation between exposure time and the Bayley Scales. This was true for both the analysis which included the children with Down's Syndrome (correlations were $r = -.68$ and $r = -.69$, $ps < .01$, for both versions of Bayley Scales) and the analysis that excluded these two infants ($r = -.55$, $p < .05$ and $r = -.52$, $p < .06$). The relationship between exposure time and later Stanford-Binet test (Down's Syndrome subjects were not involved) was also negative for all subtests but only marginally significant for Copying and Memory For Sentences. However, there was no relationship between the number of fixations and later outcome, either in terms of the MDI or Stanford-Binet scores.

Table 6-10

Correlations Between Exposure Time at 16 Months and Stanford-Binet at 27 Months
(N = 13)

Stanford-Binet subtests	<i>r</i>
Vocabulary	-.31
Comprehension	-.07
Pattern Analysis	-.23
Copying	-.50
Memory For Sentences	-.52
Partial Composite SAS	-.37

6-4 DISCUSSION

The present study was undertaken to determine if an infant's ability to differentiate between a novel and a previously seen target was predictive of later intelligence beyond the boundaries suggested by the literature. The results of the study did not clearly support this prediction. Although preference for novelty was related to later developmental scale scores, the correlations were substantially influenced by individual data. With the exclusion of intellectually disabled infants, the correlations for the developmental scale declined and became nonsignificant. Concerning the relationship between attentional measures at 16 months and Stanford-Binet test at 27 months, the data indicated no consistent results for predictability from the novelty task. The only evidence of correlation between preference for novelty and the later test came from one factor of the Stanford-Binet - Pattern Analysis. The nonsignificant correlations between general IQ and novelty preference are disconcerting. Another measure of infant attention, exposure time, was found to be a predictor of later MDIs. However, the predictability of this measure with psychometric intelligence was not consistent across all factors of the Stanford-Binet test.

Before further discussion of the results there are several points that should be considered. First, the magnitude of preference for novelty seems to be an age-independent factor, since there was no indication of increasing novelty scores either beyond 12 months or during the second half of the first year of age. Using a different system for scoring the percentage of preference for novelty did not increase the magnitude of novelty scores. However, the "timed test" procedure eventually resulted in a relatively uniform performance across all novel problems. Alternatively, this may have been due to the better performance of more mature babies. In the present study there was only one item from the abstract pattern category that did not exceed the chance level, probably due to the fact that this type

of stimulus may not be sufficiently complex for older infants and may therefore fail to capture their attention.

~~Second, unlike the novelty scores for which no developmental trend was found, exposure time and the number of fixations showed significant change in older infants. Exposure time for 16-months old infants was significantly shorter than for younger infants, an indication of more concentrated attention and a lower proportion of time spent looking away from the stimuli by older infants. They are more self-regulated and more capable of concentrating on and engaging in a task of this kind than younger babies. Similar reasoning explains the significant decrease in the number of fixations.~~ The low proportion of "time-off" task and the lower number of fixations to stimuli at 16 months of age, indicate that the infants may have entered a specific stage of development during which looking behaviours become more systematic. The statistically meaningful relationship found between exposure time and number of fixations ($r = .52, p < .05$) supports this suggestion. This implies that older infants, possibly due to maturation of the central nervous system, are better able to involve themselves in tasks, fixate more steadily on stimuli and spend less time looking at surroundings than when they were younger.

Third, the present study did not find the relationship between number of fixations and novelty scores as reported by Colombo et al. (1988) for 4 to 7 months old infants. They took this relationship to favour short-looking infants, and concluded that an infant with short fixation time should gain a high score on a novelty task.

Finally, concerning the concurrent validity of attentional measures with regard to developmental scores during infancy, the results of the present study were not consistent. Although MDI scores for the sample as a whole were significantly predictable from novelty scores, both measures were less dependent on each other with Down's Syndrome infants

excluded. However, even with the exclusion of these two infants, the relationship between novelty scores and PDI remained significant. The low novelty scores for Down's Syndrome subjects are an indication of the relative concurrent validity of the novelty task, at least in relation to the non-normal population.

6-4-1 Explanations for low predictive validity of novelty task

The interpretation of the predictive validity of novelty scores in 16-months old infants in terms of later Bayley Scales outcome at 27 months is complicated by the inclusion of two congenitally abnormal infants in the present sample. More specifically, the results indicate that preference for novelty, to some extent, is representative of a later MDI for the sample as a whole. However, the effect of outliers in correlational studies with a small sample size is well known (Shavelson, 1988). Unequivocally, the significant correlation in the current study is a product of the inclusion of outlier subjects. Generally, the predictive validity is higher for Down's Syndrome infants (Brooks-Gunn & Lewis, 1988). The psychometric intelligence test (the Stanford-Binet) was also not predicted by novelty scores. Down's Syndrome babies were not involved in the correlational analyses between the Stanford-Binet test and novelty task.

The unpredictability of the visual recognition memory test given at 16 months for later mental outcome, when congenitally disabled children were excluded, could be explained in terms of the predictor task as well as the predicted measures. For example, the low number of novel problems, and, to some extent, the nature of some stimuli that were used for the first time, may have contributed to the low predictability of the novelty task. Generally, more novelty items resulted in higher predictability of the task. Similarly, the nature and the level of complexity of stimuli are critical at particular ages (Sigman et al., 1986). It is possible that some stimuli, such as abstract patterns, were easy for infants at this

age to process (ceiling effect) and did not reflect the infants' performance on an information processing task.

Some plausible explanations concern the use of the predicted instruments (i.e., Bayley Scales and Stanford-Binet). For example, the low predictability may be due to the nature of the Bayley Scales; some studies have found that novelty scores can predict later Bayley Scales MDI whereas other studies have not (Slater, 1995). Furthermore, using an abbreviated form of the test may make the test results questionable because a shortened form may not be completely representative of a child's performance. The procedure used for testing in this study may also have been a source of the low predictability found for the Stanford-Binet. That is, because of the similarity between some items of the Bayley Scales and the Stanford-Binet tests (e.g., Copying and Pattern Analysis) at age 24-27 months, it was decided to administer the Bayley Scales first and the Stanford-Binet after a short break, to circumvent the effects of practice, fatigue and lowered motivation. Thus, the IQ obtained by the Stanford-Binet may have been influenced by experience (proactive interference), rendering it an unreliable measure of general cognitive ability. A few children were not initially interested in relinquishing the Bayley's materials, and, despite fulfilling the test requirements, did not become absorbed in the Stanford-Binet test. However, this argument is weakened by the existing a highly significant relationship between Bayley Scales and Stanford-Binet outcomes ($r = .92, p < .001, N = 13$).

Finally, as the literature suggests, the poor predictive validity of the novelty task for the Bayley Scales may be a function of infancy age at which the novelty task and outcome measures were administered. That is, for older infants it seems to be unlikely that one can predict later IQ based on the performance on novelty tasks (McCall & Carriger, 1993).

Exposure time is viewed as an important measure of an infant's looking behaviour and has been found to be a more significant predictor than duration of attention (e.g., Ruff et al., 1990). Exposure time could be interpreted as a measure of individual differences in memory capacity - that is, a latency period during which an infant attempts to rehearse a stimulus when memorising information. Therefore, one would expect that an infant with a lengthier exposure time would have a lower score on later mental outcomes, particularly memory ability. The results of the present study, to some extent, supported this suggestion. A significant negative association between log transformed exposure time at 16 months and the MDI score at 27 and a marginal correlation between exposure time and Memory For Sentences may suggest that infants with longer exposure times are among those whose scores in later cognitive outcomes are lower than infants with shorter exposure times. Hence, exposure time may be a more reliable measure than novelty preference for predicting later mental outcomes, at least for older infants.

In sum, the results of the current study lend little support for the suggestion that attentional measures, at least in this age range, and for normal infants, can be used to predict later cognitive functioning. However, the small and relatively homogenous sample (particularly with the exclusion of Down's Syndrome subjects) and the experimental nature of the stimuli restricts generalisation of the results. It remains possible that the novelty task could be developed as a worthwhile measure of infant cognition and the issue deserves further research. The tasks for future research are to provide more attractive and informative stimuli for older infants, as well as more interesting procedures (perhaps using cartoon drawing stimuli in novelty tasks presented on a television screen) for measuring attentional behaviours.

PART II

THE EFFECT OF EXPIRING FAMILIARISATION TIME IN 16 MONTHS ON THE PREDICTION OF LATER IQ AT 27 MONTHS OF AGE (STUDY 4)

6-5 INTRODUCTION

Study 3 found that a 10 second familiarization time was sufficient for normal infants aged 16-months (but not for Down's Syndrome babies) to process information; nine out of ten novel problems exceeded the chance level of 50% in the direction of preference for novelty, an indication of adjusted familiarization time. Although familiarization time was sufficient, there was nothing to suggest that the 10 second time period might be longer than the time that a normal infant at this age (i.e., 16-months) requires to process information. If this were true, a 16-month-old infant's capacity to process information rapidly might be underestimated. Similarly, if the preset study time is too generous, a ceiling effect might occur, and there would be no discernible improvement in preference for novelty after that time.

The next study investigated whether methodological changes in assessing visual recognition memory in older infants affected predictability of later IQ. More specifically, familiarization time is a preset time which the infant customarily has to accumulate to become familiar with the target. The predetermined time can be varied as a function of the infant's age and the complexity of the stimuli (Fagan, 1974; Rose, 1983). It is suggested that a preset time may not be desirable for older infants and may, in fact, underestimate or obscure subtle individual differences in capacity for novelty preference. There is some

evidence suggesting that the speed of processing increases with age during infancy (Richards, 1997). Moreover, making a test longer than an infant needs (because of accumulation of familiarization time) may result in a boring test in which an infant rapidly loses interest. This is an important point, because Study 3 indicated that 16 month-old infants had significantly less exposure time and looked at stimuli more steadily than younger babies. It is possible, therefore, that older infants, because of developmental change, have more intense concentration in acquiring information and probably do not need to meet the requirements of accumulating the preset time. If this were the case, one would expect that an infant at this age could orient her/his attention toward a stimulus and gather as many details about that stimulus during a specific period of expiring time as s/he needed for processing the information.

On the other hand, if infants do not complete processing of the stimuli because of short duration of exposure to stimuli, it would still be expected that they would show a preference for familiarity, because either preference for familiarity or novelty reflects the processing of available information. In this case, preference for familiarity will negatively correlate with later outcome. Although earlier studies (e.g., Weizmann, Cohen & Pratt, 1971) indicated that younger infants were more likely to respond to familiar rather than novel stimuli, other studies (e.g., Hunter, Ross, & Ames, 1982; Rose et. al., 1982), using both habituation and visual recognition memory procedures, have shown that preference for familiar stimuli is not an age-related phenomenon and can occur in infants as old as 12 months. In fact, the type of preference (familiarity or novelty) depends on the stage in which a stimulus is being processed. That is, novelty preference is a substitute for familiarity preference when the processing becomes more complete (Wagner & Sakovitz, 1986). In practice, this approach to familiarization time, if successful, is more feasible for

older children than the traditional accumulation of familiarization time. Older infants tend to tire prematurely of such a task, and terminate the assignment before a meaningful result is obtained.

The aim of the Study 4 was to explore the possibility that older infants (16 month-olds) were capable of monitoring their attention toward a visual recognition task which would result in preference for novelty during the familiarization period without the requirement of accumulating that time.

6-6 METHOD

6-6-1 Subjects

The sample for this study was composed of 16 (7 female, 9 male) infants with an average age of 16.3 months and a range of 15.1 to 17.7 months ($SD = .9$). The sample size and the ages at which infancy tests and follow-up tests were administered were comparable to those for Study 3. There was one infant with a gestation age of 32 weeks. The other subjects were full-term babies with a gestation age of 37 weeks or more. Nine infants were first born babies and the remainder were second born offspring. Only one subject (not included in the sample size) could not be tested as a result of crying, uncooperative behaviour and restlessness, all of which made the recording of corneal reflection of the stimuli difficult. All subjects were volunteered by their parents. Subjects were recruited, as for the other groups, by means of an advertisement in a local paper. Demographic characteristics of all subjects are shown in Table 6-11.

The age of the infants for the first visit and follow up testing were corrected for prematurity using estimated weeks of gestation. The mean socioeconomic status for both

parents was 3.5 ($SD = .50$), indicating that the sample came from middle class parents. The mean years of schooling for parents was 13.8 years ($SD = 2$). A series of t tests and Chi square were performed to determine any differences between the two groups involved in Study 3 and 4. There were no significant differences (t ranged from 1.19 to 1.25 all $ps NS$) on any demographic characteristic except for ethnicity. Study 3 had a significantly higher proportion of subjects from non-English speaking families.

Table 6-11

Birth Status and Demographic Characteristic of the Sample Tested in Study 4

Birth status and socioeconomic variables	N	Mean	SD	Range
Gestation age (weeks)	16	38.6	2.1	32 - 41
Birthweight (grams)	13	3190	540	2010 - 3880
Maternal age (years)	16	32.6	3.4	26 - 41
Maternal education (years)	16	14.3	2.4	11 - 18
Maternal job prestige scale	11	3.4	.55	2.4 - 4.1
Paternal age (years)	16	34.1	5.3	23 - 42
Paternal education (years)	15	13.5	2.2	10 - 18
Paternal job prestige scale	16	3.5	.50	2.55 - 4.30

6-6-2 Measures and stimuli during the infancy test and follow-up testing

All measures and stimuli (infancy tasks) have been fully described in relation to Study 3. The same materials were also used in the present study. The only differences between the two studies lay in the apparatus and the different scoring system for the visual recognition memory task as will be described below. All subjects were followed up when

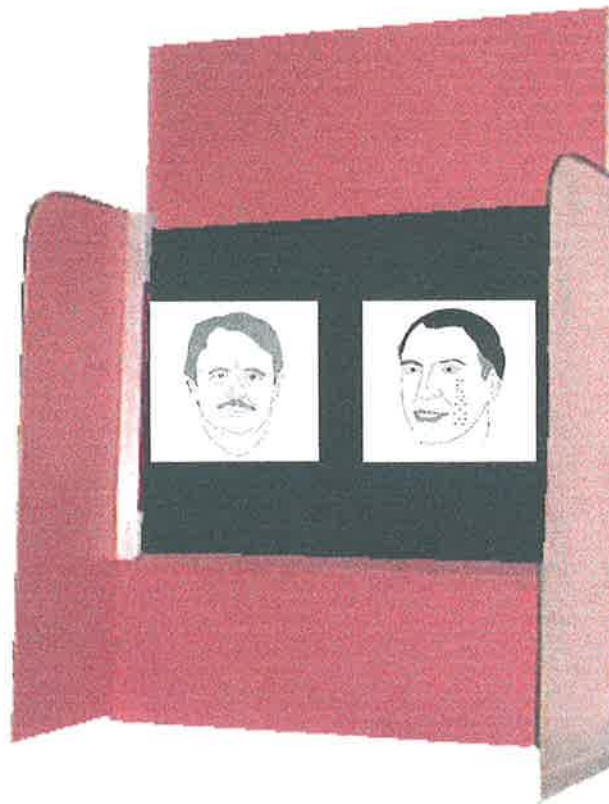
they reached 27 months of age and a developmental assessment, the Bayley Scales (both original and revised versions) and the Stanford-Binet test, were administered.

6-6-3 Apparatus

The same laboratory and equipment described for Study 3 were used for the sample in the present study. However, a different visual presentation stage was designed (Figure 6-4). One problem found with the Fagan portable presentation stage was the distraction to the infant of the repeatedly revolving and opening stage when the experimenter changed the stimuli. For older infants, who are more active than young babies, it is important to keep their concentration.

Figure 6-4

The Portable Stimulus Presentation Stage



As the equipment had to be portable, it was made from aluminium sheet 1mm thick. The unit stood 590 mm high with a detachable screen on the top, which extended the total operational height to 820 mm. The top screen was designed to hide the experimenter from the subject's view. On the subject's side of the "working" screen, there was a horizontal sliding cover of black cardboard. This was moved in and out sideways by the experimenter to expose the stimuli to the subject. Two pictures (identical for the familiarization phase but different in the test phase) were mounted on a piece of cardboard. There were 15 pieces of cardboard, all of the same width, but the height varied from 405 mm to 620 mm in 15 mm steps. This difference in height was necessary so as to assist with removal of the pictures when exposing them to the subject. On the horizontal centre line of the pictures and in the middle of the cardboard sheet, a peephole was cut. The aperture was common to, and in alignment with all of the 15 sheets, as well as the fixed back screen, so that the experimenter could look through all, or any number of the screens into the subject's eyes.

With the subject appropriately placed and the experimenter sitting behind the screen, the individual sets of pictures could be drawn vertically up and out of the equipment. At the same time as the pictures were withdrawn vertically, the sliding screen on the right hand side of the experimenter was slid into place in order to hide the next set of pictures from the subject until the experimenter was ready to observe the subject's reactions.

6-6-4 Procedure

All subjects were tested first on the novelty task and then with the Bayley Scales. In order to prevent the effects of fatigue and boredom, the number of stimuli, the sequence of presentation and the familiar or novel targets used in Study 3 were retained. However, the presentation of the stimuli, and the way in which a subject accumulated the familiarization time and the test time were different.

Testing started with the subject seated on her or his caregiver's lap (or on an adjustable high chair) in front of the testing stage. When an infant was gauged to be ready for testing, the novelty task was administered. Each novelty problem started with a familiarization phase during which the infant was exposed to two identical stimuli for 10 seconds, beginning with the first fixation to either of the stimuli and terminating when a tone sounded (after 10 seconds) from the computer. A fixation was judged when the stimulus was reflected in the infant's cornea. After the familiarization period, two 6 second immediate recognition tests were given, each consisting of a previously exposed picture paired with a new one from the same category. Again, each test started with the first fixation and expired when the 6 second preset time elapsed. In each test trial the position of the novel and familiar stimuli were reversed. All stimuli, familiar and novel, were already fixed on the black cardboard, and, when the sliding screen was removed they were presented successively, one at a time, in the same order as in Study 3.

6-7 RESULTS

6-7-1 Preliminary analyses

(i) *Infancy data*: Preliminary analyses examined the level of performance on the visual recognition memory task. Preference for novelty across each novel problem was tested against chance performance using a two tailed *t*-test. Only two novel problems (nos. 6 and 9) significantly exceeded the chance level of 50%⁴ (Table 6-12).

A comparison between novelty scores obtained here and those obtained in Study 3 for subjects of the same age revealed that the mean of novelty preference for subjects

⁴ Since each set of five novelty problems consisted of two novelty trials an alternative novelty score can also be computed by averaging the means of two novelty trials in each set. The resulting statistic indicated that only novelty scores for Set 3 (e.g., Novelty item 5 + Novelty item 6 divided by 2) exceeded the chance level of 50%.

involved in the current study was significantly lower ($t(30) = 2.02, p < .05$) than those in Study 3 (55.8% vs 63.8%), an indication of an inadequate familiarization time. The mean of actual fixation time in the familiarization phase was 6.7 seconds ($SD = 1.6$), which was probably lower than an infant of this age needs to discriminate between novel and familiar targets. In other words, since similar subjects in Study 3 significantly preferred all but one novel problem, the chance looking observed here could be attributed to the different procedure, “expiring familiarization time”, which resulted in a familiarization time too brief to allow the subjects to process the familiar targets. The means and standard deviations for the developmental indices are presented in Table 6-13.

Table 6-12

Means and Standard Deviations of Novelty Scores for all Novel Items, With t Values for the Significance of Exceeding Chance Level (N =16)

Novelty Items	Mean	SD	t -value
Novelty item 1	48.4	30.4	-.22
Novelty item 2	54.3	22.7	.75
Novelty item 3	59.4	29.2	1.03
Novelty item 4	45.4	28.2	-.66
Novelty item 5	60.3	30.1	1.37
Novelty item 6	62.5	21.5	2.33*
Novelty item 7	54.7	31.8	.59
Novelty item 8	59.2	34.2	1.08
Novelty item 9	65.3	26.3	2.30*
Novelty item 10	48.7	26.7	-0.20
Mean novelty score	55.8	11.8	1.97

Note: The t values refer to the significance level of the mean of each novelty score exceeding the chance value of 50% using a two-tailed t -test. * $p < .05$ (two-tailed).

Table 6-13

Means and Standard Deviations for all Infancy and Childhood Measures (N = 16)

A: Infancy Measures

	Mean	SD	Range
Infancy measures			
Bayley Scales (1969)			
Mental Development Index (MDI)	107.7	10.5	94 - 124
Psychomotor Development Index (PDI)	100.6	11.8	81 - 122

B: Childhood measures

	Mean	SD	Range
Childhood measures			
Bayley Scales			
MDI (1969)	120.8	12.3	98 - 139
MDI (1993)	101.3	10.1	86 - 117
Stanford-Binet Test			
Vocabulary	106.7	11.1	84 - 126
Comprehension	107.2	9.2	96 - 132
Pattern Analysis	108.0	4.4	100 - 118
Copying	102.1	10.2	69 - 124
Memory For Sentences	105.2	10.3	96 - 128
Partial Composite SAS (IQ)	106.3	8.0	93 - 122

Note: The number of subjects for Memory For Sentences, Vocabulary and Comprehension is 12, 15 and 15 respectively.

(ii) Childhood cognitive performance: The results of follow-up cognitive assessments at 27 months are also presented in Table 6-13. The MDI scores for the original version and revised edition of the Bayley Scales were 120.8 and 101.3 respectively. While all subjects were administered the Bayley Scales and the Stanford-Binet, four subjects were not given and/or had no scores on the Memory For Sentences test of the Stanford-Binet. This was due to refusal of the test ($n = 2$) or an inability to complete the test ($n = 2$). The table illustrates that the subjects' performance on the Stanford-Binet scale was in the vicinity of 106, a result similar to that obtained from subjects in Study 3. There were no significant differences for sex on the outcome measures.

6-7-2 Prediction of later MDI and IQ from the visual recognition memory task in infancy

A major aim of the study was to investigate if novelty scores obtained with this type of familiarization time would predict later intellectual assessments. Table 6-14 presents the bivariate correlations between novelty scores and later mental assessments at 27 months of age. As is evident from the table, no significant (positive or negative) relationships were found between any measure of outcome and infant preference for novel scores suggesting no predictive validity for the novelty task, either in terms of familiarity preference or novelty preference. However, the direction of correlations were, for the most part, negative.

Table 6-14

Longitudinal Correlations Between Novelty Scores and Later Childhood Tests

	N	Novelty scores <i>r</i>
<u>Bayley Scales</u>		
MDI (1969)	16	-.39
MDI (1993)	16	-.25
<u>Stanford-Binet</u>		
Vocabulary	15	-.25
Comprehension	15	.36
Pattern Analysis	16	-.42
Copying	16	.09
Memory For Sentences	12	-.07
Partial Composite SAS (IQ)	16	-.16

6-8 DISCUSSION

Even for 16-month old infants, the results from Study 4 suggested that familiarization time for the novelty task should be accumulated for a designated preset time.

Infants in this age range are still not capable of consistently attending to a task, essential behaviour if they are to gather sufficient information to process novel stimuli within a designated expiring time. Due to more mature brain functioning, it was assumed that older infants would concentrate more steadily on a task, and process the information required for differential preference, without spending much time looking away from the stimulus.

However, the results did not support this assumption, and indicated that if the direction of preference is in favour of novel stimuli, familiarization time has to be accumulated. The novelty scores obtained during infancy did not predict later mental development, perhaps because of only partial processing of stimuli.

The chance performance for 8 (out of 10) novel problems, and a large standard deviation for preference scores, indicated that an infant's attention span is too limited for her/him to concentrate continually for even as short a period of time as 10 seconds. The average familiarization time for the infants was only 6.7 seconds, possibly too short for them to be able to process the familiar stimuli and move to the next phase of differentiating between novel and familiar targets. The marginally significant relationship between the amount of actual fixation time and novelty scores ($r = .47, p = .06$) implied that infants with somewhat longer fixation times during the familiarization phase scored higher on the novelty task. The low novelty scores from the present study suggest that the infants may have been in the transition period between processing familiar targets and discriminating between familiar and novel targets (Rose et al., 1982). In such a situation, preference would be depressed and chance responding would emerge instead.

The chance responding observed here was similar to the results obtained by Rose et al. (1982) in one of their experiments with younger infants. They reported chance preference when the familiarization time was intermediate. Rose et al. (1982) interpreted

the chance responding as problematic, since "... it is not clear whether (this) represents a transitional phase in preference shifts, a phase in memory development, or a disruption in memory" (p. 712). More recently, Richards (1997) has found that, for infants aged 3 to 6 months, a brief exposure to visual stimuli resulted in familiarity preference and preference shifted towards novelty stimuli with more exposure durations. However, Freese et al. (1993) and Lasky (1980), reported that a short familiarization time for younger infants resulted in chance responding (not familiarity preference), and shifted gradually to novelty preference after a longer familiarization time. Apparently, as was the case in the present study, older infants are also subject to chance performance if familiarization time is neither too brief to reflect familiarity preference nor adjusted to elicit novelty preference.

Presumably, if the familiarization time was sufficiently brief, preference for familiarity would be observed across all novel problems. Generally, infants' responsiveness to a novelty task follows a uniform progression from familiarity preference to chance responding to preference for novelty with increasing familiarization time (Richards, 1997). Familiarity preference will more often take place when the stimuli are complex or familiarization time is very short (Wagner & Sakovits, 1986).

One important implication of this experiment could be the improvements made to the procedure used in administering the novelty task. If the minimum familiarization time required for familiarity preference is given to infants, instead of the time necessary for novelty preference, it is possible (at least in terms of the theoretical background) to predict later intelligence from familiarity preference. Familiarity preference may be used for older infants, where the establishment of novelty preference is difficult because of higher levels of activity and less endurance for testing. The task for further research remains, however, to

define an appropriate "familiarization time" during which older infants (i.e., 16 months) would have access to an amount of time which is appropriate for familiarity preference.

CHAPTER 7

PREDICTION AT 16 MONTHS OF AGE OF COGNITIVE DEVELOPMENT AT 27 MONTHS OF AGE FROM INFANT SENSORIMOTOR SCALES AND SOCIOECONOMIC FACTORS

7-1 INTRODUCTION

As outlined in Chapter 2, available longitudinal data bases provide little evidence for the predictive validity of infancy developmental tests for psychometric intelligence tests at later ages, at least for normal infants. For some intellectually disabled infants, however, the predictive validity of sensorimotor based tests may become significant (e.g., Honzik, 1983; Largo et al., 1990). Not until infants reached 18 to 24 months of age or more could developmental scales such as the Bayley MDI predict later mental competencies (e.g., Pianta & Egeland, 1994). Largo et al. (1990), for example, have investigated the predictive validity of the Griffiths test in normal, preterm and retarded infants at 9-24 months to later intellectual assessment at 7 years. They reported that the correlations between the Developmental Quotient (DQ) at 9 to 24 months and 7-year IQ were significant for preterm and retarded infants. However, for normal infants, no prediction to later IQ emerged until 18 months of age. Generally, it has been argued that developmental scales administered in early infancy bear little conceptual relation to psychometric tests of intelligence administered in childhood (Bornstein & Sigman, 1986). Nevertheless, there is some evidence suggesting that this generalization may not be correct.

7-1-1 The predictive power of developmental scales

Several researchers investigated or compared the predictability of developmental measures with other infant tests, and found that these types of tests, even during the early months of age, were valid precursors of infants' later intelligence for both normal and at-risk infants (Bradley et al., 1989; Crowe et al., 1987; Gannon, 1968; Goffeney, Henderson & Butler, 1971; Kaplan-Estrin, Jacobson, & Jacobson, 1994; Plomin & DeFries, 1983; Ramey, Campbell & Nicholson, 1973; Rubin & Balow, 1979; Siegel, 1981, 1983; see also review by Siegel, 1989). In earlier studies, Goffeney et al. (1971) found that the 8-month Bayley MDI correlated with the 7 years Full Scale IQ and Verbal and Performance scores on the WISC ($r_s = .11 - .21, N = 626$). However, they reported that when the sex factor was controlled, correlations were generally more often significant in favour of girls and, when ethnic background was compared, correlations were more often significant for white than for black children. The study conducted by Ramey et al. (1973) found that the Bayley MDI administered at 9-12 months predicted the Stanford-Binet at 36 months ($r = .71, p < .01, N = 11$) and became more accurate when children were reared in a relatively homogeneous environment. In a similar vein, Rubin and Balow (1979) have reported that low scores on the 8-month MDI (but not average scores or high scores) were associated with low performance on an IQ test and academic achievement at 4-9 years of age. Finally, in a collaborative investigation from six sites recruiting 931 subjects, Bradley et al. (1989) found that there was a moderate stability of $r = .53$ between 12 and 24 months for the Bayley MDI.

In addition to traditional developmental scales, some longitudinal studies have found that later cognitive development in childhood can also be predicted by Piagetian based tests (that measure schemes, means-end relationship, concept of space, object permanence, imitation and causality) administered during infancy (Rose et al., 1991a; Siegel, 1981;

Uzgiris, 1989; Wachs, 1975). Siegel (1981), for instance, found that the Uzgiris-Hunt Scales (Uzgiris & Hunt, 1975) given at 4 to 18 months predicted later language developmental at 2 years of age, and differentiated between infants who were cognitively delayed and those who were not.

This evidence raised the possibility that conventional infant tests have been discarded as predictors of subsequent intelligence too hastily (Seigel, 1989). Seigel's opinion was based on a review of longitudinal studies, which found that there was moderate stability between developmental tests in infancy and psychometric intelligence test scores in childhood. Seigel maintains that the exact score on an IQ test is not important. For educational purposes, only the range in which an infant's score may fall is important and it is possible to predict such a range from infancy tests.

Humphreys and Davey (1988) used the simplex model and analysed the intercorrelations of measures of infant developmental scales with childhood intelligence tests between the ages of 9 months and 9 years. They concluded that "continuity in intellectual development from 12 months to 9 years along a dimension appropriately called general intelligence cannot presently be rejected" (p. 195), but suggested that the 9 months scores were questionable, and that scores obtained from earlier ages were not reliable for prediction of later IQ.

The availability of developmental test scores from all samples of the previous studies made it possible to test the Humphreys and Davey's (1988) hypothesis for infants in the age range from 6 to 16 months. Bearing in mind the contradictions in the literature, the main concern of the present analysis was to investigate the predictive value of the MDI and PDI obtained from infants aged 6, 7, 9, 12 and 16 months, to later MDI/IQ scores in children

aged between 18 and 27 months, and to compare the results with those obtained from a visual recognition memory task.

7-1-2 Predicting later IQ from socioenvironmental factors

Following the demonstrated generally low predictability of sensorimotor based tests to later childhood intelligence, developmentalists turned to other measures that were not dependent on the infant. As an alternative, the relationship was assessed between independent factors, such as parental IQ, education and occupational prestige, and the infant's later mental capacities. On occasions, some researchers also used other factors, such as birth order and maternal age. These factors do not rely on the infant's responsiveness and therefore could be measured more precisely.

There are a number of studies that attempted to estimate infant intelligence by means of parental IQ, and most found that there is a genetic contribution to infants' test scores (e.g., Cardon & Fulker, 1991; Plomin & DeFries, 1983; Thompson et al., 1985). A recent longitudinal study by Scarr, Weinberg and Waldman (1993) found that the correlations between parents and their biological offspring, and also between parents and their adopted children, reliably deviate from zero, suggesting the importance of both genetic and environmental effects.

Given that the influence of genetic factors is not perfect (DeFries, Plomin, & Labuda, 1987), then the remaining shared variance in intelligence from infancy to adulthood could be attributed to a substantial environmental contribution. Many studies have reported a significant correlation between parental education or SES and infants' intellectual capacities (e.g., Largo et al., 1990; Laucht et al., 1994; O'Connor et al., 1984). In an early study, Bronman, Nichols and Kennedy (1975 cited by Siegler & Richards, 1982) examined the relationship between the Stanford-Binet and 169 prenatal and postnatal predictors for

more than 26,000 4-year old subjects. They found that race, SES and maternal IQ were the best predictors, but none of these factors accounted for more than 20% of the variance. Laucht et al. (1994) have reported that parental education was the best predictor throughout the first 54 months of age. However, it seems that early development is more influenced by maturation than by polygenic and environmental factors (Thompson, 1990) because the relationship between SES and later outcome increased as infants became older. Bradley et al. (1989), for example, reported that the relationship between social status and mental test scores for their sample increased in magnitude between the ages of 1 and 3 years. Nevertheless, it is not unusual to find inconsistent results in the literature. A further objective of the present analyses was to examine the role of the mediating influence of socioenvironmental factors in determining later MDI/IQ for infants during the first 27 months of life.

7-2 METHOD

7-2-1 Subjects

The present investigation used data taken from all subjects in Study 1, Study 3 and Study 4. No new subjects were involved in this study. A comprehensive demographic description of subjects has previously been reported in chapters 4 and 6. Briefly, the entire cohort for the current study included 108 healthy infants (50 females, 58 males) who successfully completed all infancy tests¹. The subjects, depending on the infant's age at the first testing, were assigned to one of five age groups. The first four age groups correspond to the four age-related procedures of the FTII (i.e., 27, 29, 39 and 52), and the final group consisted of infants with an average age of 69 weeks at the first testing. There were then 18

¹ The data for two Down's Syndrome infants from the 69 week-old subjects were excluded from the analyses.

infants aged 27 weeks, 15 aged 29 weeks, 22 aged 39 weeks, 23 aged 52 weeks and 30 infants at 69 weeks of age. As a whole, the sample was derived from a middle class neighbourhood, and the majority (74.1%, $n = 80$) was of Australian parentage.

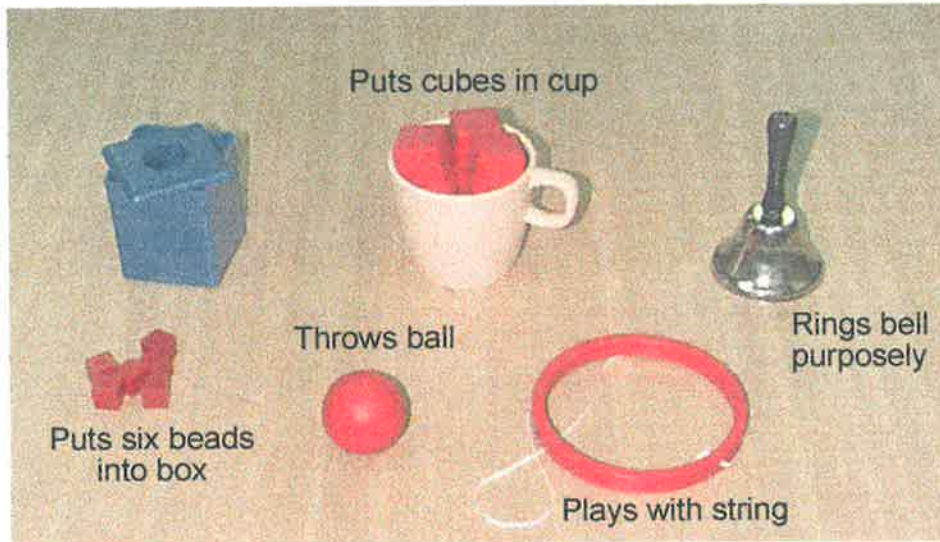
7-2-2 Measures and procedure

The design and procedure for both infancy tests and childhood tests have been fully described in relation to the previous studies. Briefly, all infants were seen in infancy and were administered the Bayley Scales of Infant Development (Bayley, 1969) - from which two different indices (i.e., MDI and PDI) for an infant's developmental status were drawn (see Figure 7-1 for sample of some test items from the Bayley Scales used in this study). The first four age groups were followed up one year later, as well as when they reached two years of age, and they were retested with both the original version of the Bayley Scales (Bayley, 1969) and the revised version of the same scale (Bayley, 1993). At two years of age, an abbreviated version of the Stanford-Binet was also administered to the subjects (see Study 2). The oldest group (i.e., 69 week-olds) was also followed up at 27 months and retested with the same measures as the previous groups. All tests were given and scored at the child's corrected age for prematurity rather than chronological age. To obtain other demographic characteristics for subjects and their parents, a background information sheet was completed for each subject at the first testing. Birth status and health information were also obtained through the infant's health record completed at birth by medical staff.

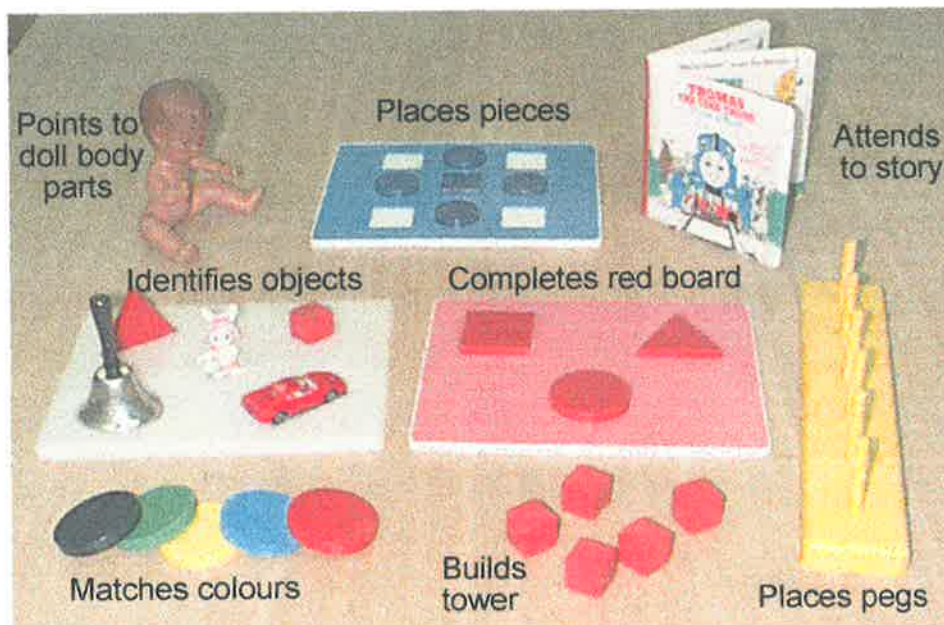
(i) ***The Standard Progressive Matrices***: As one aim of the study was to investigate the relationship between IQs for infants and their mothers, all mothers were tested with a general intelligence test, the Standard Progressive Matrices (Australian Standardization). The Standard Progressive Matrices is a non-verbal intelligence test designed to provide a measure of general reasoning ability (Spearman's g factor) and is held to be a culture-fair test. The test has five sets, each consisting of 12 items. Each item consists of a matrix of

Figure 7-1

Some Test Material from the Bayley Scales of Infant Development



Infant Test Items



Childhood Test Items

abstract figures with one piece of information left blank. On the bottom of each matrix six or eight patterns from the same category are given as multiple choice answers. A subject is expected to select the best figure that logically matches the rest of the matrix in order to complete it. All sets, and the items within each set, are arranged in order of difficulty. This test was chosen to assess the intelligence of the subjects' mothers because the results of some studies (e.g., Thompson et al., 1985) suggested that infant developmental scores tended to be correlated with the *g* factor of the parents, rather than with specific cognitive abilities. The concept of fluid intelligence and *g* factor, as suggested by Eysenck (1988), is clearly defined by tests such as Raven matrices. Moreover, the subjects' mothers had different levels of education, and approximately 25% of the mothers were from different cultural backgrounds to the major Australian culture. The raw score, rather than a standard score, was used in this study. However, in accordance with the Australian Adaptation version, a timed procedure (20 minutes limit) was chosen.

7-3 RESULTS

7-3-1 Preliminary analysis

(i) *Infancy measures*: The means and standard deviations for the developmental test scores during infancy are presented in Table 7-1. The means of the MDI varied from 110.8 to 115.8, and the means of the PDI from 105.1 to 113.8. In all ages but 52 weeks, the MDI and PDI significantly correlated with each other (r s from .52 to .86 all p s < .01). In most age groups, girls scored slightly higher on the mental index than boys and boys scored slightly higher than girls on the psychomotor index, but the differences were not statistically significant (t ranged from 1.02 to 1.15, all p s *NS*). The only difference observed here was the sex difference in MDI for 16-month old subjects, which was not evident when data for

the two groups involved in Study 3 and Study 4 were reported separately. When the MDI scores for both groups were combined, girls scored approximately 10 points (116 vs 106) higher than boys and the difference was significant ($t(28) = 2.57, p < .05$). There was no significant difference for ethnic background on both indices (t ranged from .87 to 1.79, all ps NS).

Table 7-1

Means and Standard Deviations of the Bayley Scales During Infancy

Bayley Scales (1969)	N	MD	SD	Range
27-week MDI	18	113.1	15.0	91 - 139
27-week PDI	18	113.8	11.1	92 - 129
29-week MDI	15	110.9	12.8	86 - 128
29-week PDI	15	105.2	8.5	87 - 117
39-week MDI	22	112.6	12.6	89 - 140
39-week PDI	22	105.6	16.3	72 - 139
52-week MDI	23	115.8	8.4	93 - 131
52-week PDI	23	106.2	9.9	86 - 122
69-week MDI	30	110.8	11.5	92 - 130
69-week PDI	30	105.1	10.9	81 - 124

Both the MDIs and PDIs were higher than the normative data. The higher mean MDIs found in the present study are consistent with other research which has found similar scores for normal infants in this age range (e.g., Bathurst & Gottfried, 1987; Bradley et al., 1989; Coates & Lewis, 1984; Field, et al., 1981; Kaplan-Estrin et al., 1994). Coates and Lewis, for example, found that the MDI for their subjects was 122 and 114 at 3 and 12 months of age and Bradley et al. (1989) have reported a mean of 117.1 ($SD = 10.8$) for one

year old infants from one site of their collaborative study in North America. However, it should be noted that the Bayley Scales (Bayley, 1969) overestimate infants' mental and motor competencies (Campbell et al., 1986; Tasbihsazan, Nettelbeck & Kirby, 1997). This will be discussed further in Chapter 8.

(ii) *Childhood measures*: In the follow-up assessment, data for five subjects were lost due to absence and/or withdrawal from the project (11 subjects for whom two follow-up testings were scheduled also missed one testing). Table 7-2 presents the mental outcomes in the follow-up sessions. The mean IQs (Partial Composite SAS) for 24 and 27-month-olds were 102.1 and 108.2 respectively. The low standard deviation of the Stanford-Binet test is an indication of the relatively similar performance of most children, particularly on the Abstract/Visual Reasoning tests (i.e., Copying and Pattern Analysis).

The MDI scores from the revised version of the Bayley Scales (Bayley, 1993) showed significant differences in cognitive ability for gender in favour of females at 24 months ($t(66) = 2.46, p < .05$) and 27 months ($t(27) = 2.12, p < .05$) of age². However, there was no significant gender difference for the Stanford-Binet outcome. In terms of ethnic background, English-speaking children scored significantly higher than children coming from non-English speaking backgrounds in the oldest age group. That is, 27-month old children from English speaking families scored higher, both on the MDI ($t(27) = 2.15, p < .05$)³ and partial Composite SAS ($t(27) = 2.12, p < .05$), than subjects from non-English background families.

² The difference between MDI scores for the first version of the Bayley Scales (Bayley, 1969) was also significantly higher ($t(66) = 2.17, p < .05$) at 24-month-old and 27 months of age ($t(27) = 2.48 = p < .05$).

³ The difference was significant only for the First Edition of the Bayley Scales and not for the Second Edition.

Table 7-2

Means and Standard Deviations for Bayley Scales and Stanford-Binet Test

Bayley Scales Age	MDI (1969)			MDI (1993)		
	M	SD	Range	M	SD	Range
18 months (n = 27)	117.7	13.6	81 - 150	98.6	10.6	77 - 121
21 months (n = 20)	116.7	15.0	93 - 150	100.0	11.1	80 - 120
24 months (n = 68)	122.5	19.4	76 - 150	102.6	14.2	68 - 134
27 months (n = 29)	122.8	13.9	98 - 145	104.5	11.2	86 - 129

Stanford-Binet test	24 months			27 months		
	M	SD	Range	M	SD	Range
Vocabulary	101.6	10.6	84 - 130	108.4	10.7	84 - 126
Comprehension	103.6	9.0	84 - 132	104.7	10.7	84 - 132
Pattern Analysis	103.2	7.5	88 - 118	106.3	8.3	88 - 126
Copying	98.9	6.9	96 - 132	105.7	11.3	96 - 124
Memory For Sentences	104.9	7.7	96 - 120	108.9	10.5	96 - 128
General IQ	102.1	7.9	89 - 125	108.2	10.1	93 - 128

Note: The number of subjects for the 24-month Stanford-Binet test varied between 53 and 61, except for Memory For Sentences for which only 30 children succeeded in achieving a score. The number of subjects for the same test at 27 months were 25 for Memory For Sentences and between 28 and 29 for the other tests.

Of particular interest was a substantial increase in cognitive abilities during the second year of life (see Table 7-3). Comparing means of the MDI obtained in infancy and childhood with a paired-sample *t* test, it was found that the MDIs obtained during infancy were significantly lower than those obtained in childhood. That is, there was an increase of 9.9 points in the MDI between 6 and 16 months and 24 and 27 months of age.

7-3-2 Intercorrelations between the MDI and SES

Table 7-4 presents the intercorrelations between the MDI obtained from subjects in infancy and parental demographic characteristics. It is clear from the table that there was (with one exception) no significant correlation between infancy MDI and environmental

factors, nor was there a significant correlation between the same measures when the effect of age was removed. This implies that, at least at this stage, Bayley scores are independent of parents' socioenvironmental factors. One might expect a higher correlation between SES and infant test scores as the children get older (Wilson & Matheny, 1983).

Table 7-3

Means and *t* Values for MDI Obtained at Infancy and Childhood

Infancy age	N	Infancy measure MDI (1969)	Childhood measure MDI (1969)	<i>t</i> -values
27-Week	16	113.2	116.3	.65
29-Week	12	111.7	132.2	2.64*
39-Week	19	111.7	120.2	2.08*
52-Week	21	116.4	123.8	2.54*
69-Week	29	110.9	122.8	5.98***
Total sample	97	112.7	122.6	-

p* ≤ .05 (two-tailed). * *p* ≤ .001 (two-tailed).

Table 7-4

Correlations Between MDI at Infancy and Parental Demographic Characteristics

Demographic variables	27 weeks MDI	29 weeks MDI	39 weeks MDI	52 weeks MDI	69 weeks MDI
Maternal age	.47	.14	.11	.12	-.17
Maternal education	.20	-.30	.29	-.26	.15
Maternal job status	-.12	.28	-.09	.06	-.11
Maternal IQ	-.04	.09	.15	.04	.31
Paternal education	.01	-.04	.02	-.02	.32
Paternal job status	.08	-.06	.11	.29	-.41*

**p* ≤ .05 (two-tailed).

7-3-3 Reliability of mental and psychomotor indices

The internal consistency (Cronbach's alpha) of the mental index of the Bayley Scales was sufficiently high to make the test psychometrically reliable. In contrast to the information processing task (FTII), the coefficient alphas of Bayley Scales for all ages were substantially higher. The internal consistency of the MDI was .81, .76, .74, .76 and .85 at 27, 29, 39, 52 and 69 weeks of age respectively, which replicates normative data from the Bayley Scales.

The stability of the MDI over short-term (two weeks) and relatively longer term (two to three months) was also assessed for the the infants who were tested twice. Resulting coefficients are displayed in Table 7-5. As can be seen from the table, with the exception of low test-retest reliability for both the MDI and the PDI between the ages of 29 and 39 weeks, the two scales were highly reliable across different age spans. The MDI from the Bayley Scales during childhood was also highly reliable. The correlation coefficients between 18 and 24 months and between 21 and 24 months of age were, $r = .70, p < .001, n = 25$, and $r = .84, p < .001, n = 17$, respectively.

Table 7-5

Test-retest Reliability of MDI and PDI During Infancy

Age groups	N	MDI	PDI
27/29 weeks	12	.94***	.91***
29/39 weeks	12	.40	.45
39/52 weeks	13	.86***	.74**

** $p \leq .01$ (two-tailed). *** $p \leq .001$ (two-tailed).

7-3-4 Bivariate correlations between infancy MDI and later MDI at childhood

Pearson product-moment correlations were computed between the MDI and the PDI in infancy, and later developmental outcomes at 18, 21, 24, and 27 months of age (Table 7-6). The 27 and 29 week MDIs did not predict later Bayley Scales scores either at 18 months or at 24 months. However, as the table clearly shows, the MDIs at 39, and particularly at 52 and 69 weeks, significantly predicted subsequent mental indices at most ages. The magnitude of the correlations increased as the age at the early testing increased.

The PDI, a measure of motor development at 27 weeks, predicted later Bayley Scales scores only at 18 months. For 39-week-old infants, however, the relationships between the PDI and subsequent MDI were significant at both 21 and 24 months. The correlations between the PDI in infancy with later MDI for other groups were not consistent, leaving the PDI at 39 months the best predictor of subsequent functioning. Generally, in earlier ages, the PDI and the MDI correlate well but correlations decline from 1 year to about 2 years (Plomin et al., 1988)⁴. Thus, the possibility of increasing the predictability of the Bayley Scales was assessed using the average score of the MDI and the PDI for all age groups. The results were essentially the same, and no predictability was found for the 27 and 29 week tests.

7-3-5 Correlation between MDI at infancy and later IQ

Correlations between earlier MDI and later Stanford-Binet test are presented in Table 7-7. The pattern of correlations between the MDI and the Stanford-Binet resemble the pattern of correlations between the MDI and later Bayley Scales. That is, correlations between the MDI in infancy and specific cognitive factors of the Stanford-Binet in childhood increased as the age (from 27 to 69 weeks) at which the Bayley Scales were

⁴ In the present study the correlations between MDI and PDI were .66, .71 and .83 for 27, 29 and 39-week olds respectively. The correlation decreases to .27 at 52 weeks of age, however.

Table 7-6

Bivariate Correlations Between Earlier Bayley Mental and Motor Scales (MDI and PDI) and Subsequent MDI in Childhood

Infancy measures	18 months		21 months		24 months		27 months	
	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)	MDI (1969)	MDI (1993)
27 weeks MDI and PDI (n = 15 to 16)	.50 .57*	.48 .49			.37 .29	.21 .07		
29 weeks MDI and PDI (n = 12)	-.11 -.57	-.22 -.59			-.54 -.39	-.39 -.57		
39 weeks MDI and PDI (n = 19 to 20)			.58** .58**	.40 .45*	.55* .61**	.51* .57**		
52 weeks MDI and PDI (n = 21)					.65** .11	.61** .13		
69 weeks MDI and PDI (n = 29)							.66*** .27	.81*** .46**

Note: MDI is the Mental Development Index obtained from the Bayley Scales (1969 & 1993). Correlations in **bold** are bivariate relationships between PDI and later Bayley Scales. *p < .05. (two-tailed). ** p < .01. (two-tailed). *** p < .001. (two-tailed).

administered increased. Once again, all correlations between the 27 and 29 week MDIs and later Stanford-Binet were nonsignificant, and some were negative. The number and the magnitude of the correlations between 39 week MDIs and Stanford-Binet factors were still low, but all correlations were positive, and some of them significant at 52 and 69 weeks. The relationship between the MDI and Partial Composite SAS was also significant at the two oldest ages. However, the relationship between the PDI and later Stanford-Binet (not reflected in the table) was low and nonsignificant for all age groups.

Table 7-7

Correlations Between MDI at Infancy and Stanford-Binet Test at Childhood (**bold**) and the Number of Subjects in Each Correlation

Infancy age	27 weeks	29 weeks	39 weeks	52 weeks	69 weeks
	MDI	MDI	MDI	MDI	MDI
Stanford-Binet Test					
Vocabulary	.14	-.28	.37	.58**	.67***
	12	9	15	20	28
Comprehension	.26	-.13	.23	.01	.35
	12	9	12	20	28
Pattern Analysis	.50	-.36	.66**	.48*	.43*
	14	11	15	21	29
Copying	-.11	.44	.31	.14	.53**
	12	9	15	20	29
Memory For Sentences	.16	-.67	.51	.49	.53**
	7	8	9	10	25
Partial Composite SAS	.29	-.47	.45	.61**	.72***
	14	11	15	21	29

Note: Correlations between the 27, 29, 39 and 52 week MDI and the Stanford-Binet were based on administration of the Stanford-Binet at 2 years, and correlations between the 69 week MDI and the Stanford-Binet were based on the Stanford-Binet scale administered at 27 months of age.
 * $p \leq .05$. (two-tailed). ** $p \leq .01$. (two-tailed). *** $p \leq .001$. (two-tailed).

7-3-6 Multivariate analyses

The results of Study 2 indicated that MDI from Bayley Scales and Vocabulary and Pattern Analysis scores from the the Stanford-Binet, all at 24 months of age, were predictable from the 52-week FTII. Similarly, the 52-week MDI, as described above (see Table 7-6 and Table 7-7) did predict later outcome for both Bayley Scales and Stanford-Binet at 24 months. In order to determine the contribution of each predictor alone and the power of the combination of both predictors, the MDI and FTII scores at 52 weeks were entered into regression equations as independent variables. Dependent variables were the MDI from the Bayley Scales and Vocabulary and Pattern Analysis from the Stanford-Binet, all at 24 months of age.

The MDI and FTII at 52 weeks, each alone, explained 37% and 24% of variance in the 24-months MDI scores, respectively. Explained variance increased to 44% when these two variables were combined but the increase from 37% was not statistically significant ($\alpha = .05$). Thus, when the FTII was added to the regression, its contribution to the prediction amounted to only 7% of explained variance.

Separate multiple regression analyses were made, with the 52-week MDI and FTII as independent variables and with Vocabulary and Pattern Analysis scores from the Stanford-Binet test as dependent variables. These two subtests were selected because, as already discussed, they were significantly correlated with MDI and FTII scores during infancy. It can be seen from Table 7-8 that, once again, the results from multiple regression did not support any significant contribution from the FTII over the MDI (Appendix H).

Table 7-8

Multiple Regression Predicting Later MDI and Stanford-Binet Subtests at 2 Years of Age

Dependent Variables	Independent Variables				F Change
	MDI		MDI + FTII		
	R ²	Adjusted R ²	R ²	Adjusted R ²	
24-month-MDI	.37**	.34	.44**	.38	2.21 (<i>ns</i>)
Vocabulary	.33**	.30	.39**	.32	1.72 (<i>ns</i>)
Pattern Analysis	.22*	.18	.33	.25	2.58 (<i>ns</i>)

* $p \leq .05$. (two-tailed). ** $p \leq .01$. (two-tailed).

7-3-7 Correlational analyses of socioenvironmental indices and later mental outcomes

The relationships between the most frequently used environmental factors, such as maternal education and job status, mother's IQ, birth order and maternal age and later mental outcomes were assessed. Table 7-9 presents correlations between socioenvironmental measures and later mental outcome at 24 and 27 months of age⁵. As can be seen from the table, maternal job status was the only predictor of later outcome. The magnitude of correlations for both age groups was the same. However, due to the smaller sample size of the 27-month old infants, the correlation for this group was not significant. Maternal job status accounted for 9% to 13% of the variance in later mental outcome. The significant negative correlation between job scale and later outcome indicates that infants from lower SES strata are less likely to achieve high scores on intellectual functioning than their peers of higher socioeconomic status. No significant relationship was found between birth order and later outcome or between maternal age and later mental functions, either in terms of a psychometric *g* factor or MDI scores.

⁵ The correlations reported in this table are based on the subsequent scores from the second version of the Bayley MDI. Another correlational analysis was performed using the MDI from the first version. The results were consistent for both analyses.

Table 7-9

Bivariate Correlations Between Birth Status and Parental Socioenvironmental Factors and Later Cognitive Outcome

Demographic variables	24 months outcome				27 months outcome			
	MDI (1993)		Stanford-Binet		MDI (1993)		Stanford-Binet	
	<i>r</i>	N	<i>r</i>	N	<i>r</i>	N	<i>r</i>	N
Birth order	.09	68	.05	61	-.01	29	.11	29
Maternal age	.20	66	.13	59	.02	29	.02	28
Maternal education	.21	68	.07	61	.02	29	.16	29
Maternal job status	-.36**	48	-.32*	43	-.36	21	-.30	21
Paternal education	-.07	64	.00	57	-.05	28	-.11	28
Paternal job status	.01	60	.09	54	.11	26	.20	26

* $p < .05$. (two-tailed). ** $p < .01$. (two-tailed).

To investigate the importance of maternal general IQ in predicting an infant's potential intelligence later in life, the subjects' mothers' Raven test scores were correlated with the childhood test scores. The correlation between the children's developmental index and their mothers IQ was not significant at any age. However, there were significant correlations between mothers' scores on the Standard Progressive Matrices and Pattern Analysis and Copying from the Stanford-Binet test at 24 months (Table 7-10).

7-4 DISCUSSION

The major focus of the present study was to examine the predictive power of the conventional infant developmental scales, in addition to some of the most frequently used socioenvironmental factors. The study provided some evidence of the value of sensorimotor scales in predicting potential mental capacities of children when they were given at 9 months or later, but not when administered at approximately 7 months of age or

Table 7-10

Correlations Between Children's Stanford-Binet Scores at 24 and 27 Months and Their Mothers' Standard Progressive Matrices Scores

Outcome measures	24 months		27 months	
	N	<i>r</i>	N	<i>r</i>
Bayley MDI (1993)	64	.06	27	.29
Stanford-Binet test				
Vocabulary	54	.19	26	.19
Comprehension	51	.09	26	.27
Pattern Analysis	58	.43*	27	.14
Copying	55	.44*	27	.13
Memory For Sentences	29	.07	23	.14
Partial Composite SAS	59	.34	27	.20

* $p < .05$. (Two-tailed).

earlier. The study could not support a consistent predictability for some socioenvironmental factors such as birth order, maternal age and parental education, at least in early childhood. The only factor that was a reasonable predictor of infants' later mental functioning was maternal job status, which predicted childhood test scores at 24 months of age only. In addition, the current study did not find strong support for prediction of later IQ from maternal IQ scores.

Before discussing the results of the study, there are a number of comments that can be made on several general findings concerning infant and childhood mental scores. First, a comparison of the MDI in infancy with early childhood MDI has disclosed an increase of about 10 points in the infants' cognitive functioning between the first assessment and the subsequent assessments at 24 and 27 months of age, where, possibly because of increased age, the children were more exposed to the home environmental influences. Generally, it

has been concluded that the social class of parents does not affect infants' cognitive skills in early infancy, because the MDI scores for low class infants have been found to be similar to infants from middle class parentage (e.g., Kaplan et al., 1994 ; Widemayer et al., 1990). However, starting from the second year of life, the impact of environmental factors is incremental and more effective. For example, the increasing mental scores observed here contrast with the substantial decline of the MDI during the second year of age for infants from lower class parentage. Kaplan et al. (1994), Rose and Wallace (1985) reported a significant decline of between 10 and 22 points on the Bayley MDI between 12 and 25 months, although their subjects' MDI scores fell within the normal range of infancy tests. Their subjects came from families that were predominantly of a lower socioeconomic status. These data also acknowledged the proposition put forward by McCall (1981a) that cognitive development during early infancy is canalised and follows "the species-typical path". That is, the early developmental function for all members of species is largely determined by maturation.

Second, there has been a tendency toward gender differences in MDI, particularly for older infants (i.e., 24 to 27 months). However, female performance on the Stanford-Binet test was not significantly higher than the males. Such difference as exists in other developmental tests could be a consequence of better performance by girls on vocalisation items. The gender difference found may be attributable to a peculiarity in language development, but may also reflect a difference in temperament. Generally, girls are more cooperative and become more engaged in a task than boys. This may result in a higher score in favour of girls.

The primary objective of the current study was to establish the first age range during which a developmental scale (i.e., the Bayley Scales) could predict childhood mental

capacities. The results of the study suggest that skills exhibited by an infant, from 9 months to some extent, and consistently from 12 months onwards, can predict their later capacity to solve problems typical of a developmental scale or an intelligence test. While acknowledging the inconsistencies in the literature, the present study supported studies that found some predictive validity for infancy developmental scores obtained after 9 months of age. Nevertheless, the 12 month developmental test scores predicted later outcome more precisely than the 9 month MDI. This supports Humphreys and Davey (1988) who questioned the predictability of the 9-month-MDI, and suggested that the starting point for prediction of the construct of general intelligence is 12 months.

When comparing the predictive power of the MDI with the novelty scores obtained from the FTII, the results of multiple regression clearly showed a superiority for the predictive validity of the Bayley Scales. At two years of age, a higher portion of explained variance was accounted for by the MDI than by the FTII. Overall, between 22% and 33 % of the variance was uniquely accounted for by the MDI. When the FTII was added to the regression, its contribution to the prediction never exceeded more than 11% of the explained variance and in all instances the amount of change was nonsignificant. However, it should be noted that the age range studied here was relatively restricted and the results of the predictive validity of the MDI should be interpreted cautiously.

Studies that have been designed to assess the genetic contribution to intelligence have found evidence of a genetic influence on cognitive abilities. However, the influence of heritability becomes more prominent with increasing age (e.g., Baker, DeFries, & Fulker, 1983; Plomin & DeFries, 1983; Thompson et al., 1985). The present study was not designed to assess the genetic relationship between infants and parents, but the low relationship found between infants' MDI and Stanford-Binet scores and their mothers'

general IQ, suggests that the shared variance in maternal intelligence that can be explained by the MDI/IQ of offspring at 24 to 27 months of age is very low.

The influence of the home environment on cognitive development is generally indicated by socioeconomic status of parents. In the past, several socioenvironmental factors, including birth order, maternal age and parental socioeconomic status (SES), have been found to affect a child's mental development. In the present study the correlation between maternal job prestige and mental outcome for children aged 24 months was significant ($r = -.36, p < .01$ and $r = -.32, p < .05$ for Bayley Scales and Stanford-Binet respectively). However, paternal job prestige did not predict later outcome at 24 and/or 27 months (perhaps due to small sample size). Since the Australian scale of occupational prestige (Daniel, 1983) rates the relative status of each job downward (the top score is 1 and is given to the highest social standing for job and the lowest score is 7 for the least social standing job), the negative correlation implies that infants from higher social classes have higher subsequent MDI/IQ. It should be noted that using parental occupation as a measure of social standing is a controversial issue, particularly when one score for only one parent is taken into account. However, it seems to be one of the best factors representing SES (Mueller & Parcel, 1981). Nevertheless, it is not reasonable to assume that only female work or male work determine the family social standing (Muller & Parcel, 1981). One suggestion is that when both parents are working, social standing should be measured by the more prestigious occupation (Haug, 1973).

In general, SES is operationally defined by either the level of parental education or parental occupation (e.g., Plomin & DeFries, 1983). The level of parents' education and paternal occupation prestige are highly correlated (Bradley et al., 1989), as was found in the present study ($r = -.50, p < .001, N = 102$). However, the number of schooling years of

parents did not correlate with later outcome in childhood. This pattern of inconsistency in the intercorrelations between socioenvironmental influences and childhood outcome, particularly at an early age, is not exclusive to the present study (see Bayley, 1970).

The present study did not find any predictability of later cognitive development for birth order and/or maternal age. It is possible, since the majority of subjects (92 out of 108) in the present study were first or second born babies, that the sample could not capture the relative importance of birth order, resulting in no significant relationship between birth order and IQ. A similar explanation could be argued for the lack of a relationship between mother's age and later MDI/IQ. As is clear from the demographic table (7-1), the standard deviation for maternal age is relatively low, and 73.3% of the subjects' mothers clustered around 26-35 years.

Furthermore, it is not inconceivable that birth order and maternal age, like other demographic characteristics, needs a longer passage of time to correlate with intellectual functioning. The correlations between infant test scores and home environmental influences appear to be greater later in development (Baker et al., 1983; Johnson et al., 1993). For example, the relationship between SES and later environment factors was found to be very weak at 6 and 24 months, but after 3 years of age it improved, to reach .66 at 6 years of age (Wilson & Matheny, 1983). It seems that the relatively short interval of time studied here is one important factor that obscures the predictability of socioeconomic influences.

CHAPTER 8

INCREASING MENTAL DEVELOPMENT INDEX IN AUSTRALIAN CHILDREN¹

8-1 INTRODUCTION

Historically, the first psychometric attempts to measure infants' functioning date back to the 1920s when the first infant intelligence test was published (Gesell, 1925). This was based on motor and sensory activities which were easy to measure and thought to be a reliable reflection of infants' mental competencies. All subsequent infant developmental scales including the Bayley Scales of Infant Development (Bayley, 1969) were not only based on this construct but also borrowed many items from the first test (Honzik, 1983).

8-1-1 Reliability of the Bayley Scales of Infant Development (Bayley, 1993)

The Bayley Scales (Bayley, 1969) constitute one of the best known tests and it has been used frequently in research and clinical appraisal with thousands of children of different cultural backgrounds. Bearing in mind that Nancy Bayley started her first longitudinal study with the view of constructing an infant test in the 1930s, it is clear that the normative data from this test are dated. On the other hand, given the nature of sensorimotor activities sampled by the procedures involved, it is not immediately obvious that normative data would change.

Recently, the second edition of the Bayley Scales became commercially available (Bayley, 1993). The test was published with a number of specific goals in mind. First, normative data have been updated. The new version has been standardised on 1700

¹ An article based on this chapter has been published in: Tasbihsazan, R., Nettelbeck, T., & Kirby, N. (1997). Increasing Mental Development Index in Australian children: A comparative study of two versions of the Bayley Mental Scale. *Australian Psychologist*, 32 (2), 120-125 (see Appendices).

children, representing an American population of infants from 1 to 42 months of age. This sample consisted of 100 children in each of 17 age groups which, in practical terms, extended the upper age of the first version of the test from 30 to 42 months of age. Second, despite the majority of common items (76%) across the two mental scales and an identical scoring system to the previous edition, many new items have been included to improve the content coverage; 29 items from the mental scale of the original version have also been removed. Other goals were to improve the materials and psychometric properties, as well as the clinical utility of the scales. In particular, providing a 3-points cutoff score for infant behaviour during testing has made the Behaviour Rating Scale more meaningful for clinical use. Consequently, the new edition is intended to reflect an extensive range of mental domains, including infant attention, problem solving, number and counting, language and personal/social development and sensory-motor integration (Bayley, 1993). In addition, items assessing an infant's memory and ability for processing information were added to the new edition, to increase the predictive validity of the test. New evidence is promoting the acceptance that this type of task can predict IQ later in childhood (Bornstein, 1989b; Rose, 1988). This is consistent with Anderson's (1992) argument that cognitive processes underlying individual differences in conventional IQ are different to those that have commonly been included in tests of infant development.

The first edition of the Bayley Scales (Bayley, 1969) was in use world wide and it is intended that the second edition will be similarly widely used. However, the test is standardised on American children and the normative data may not be applicable to other infants from different cultural backgrounds. This is an extremely important point, since the new version is heavily dependent on the vocabulary items for infants aged 18-27 months (although it is less dependent on verbal items beyond that). The preponderance of verbal

items may render the test culturally biased and inappropriate for infants from a non-English background. In the present study, 25% of the sample were from different ethnic backgrounds, making it possible to investigate whether the new version of the Bayley Scales is culture fair.

To date, no independent study has compared the original and new versions of the Bayley Scales or has investigated the reliability, stability and concurrent validity of the new version. The current report used data from Studies 2, 3 and 4 to pursue its first aim, namely, to investigate the psychometric properties of the second edition of the Bayley Scales (Bayley, 1993). Although gender and parental education levels of the sample recruited for the present study approximated the Bayley standardisation sample, no attempt was made to recruit a representative sample. The 18 to 27 months age range was chosen for investigation because the majority of items added to, or deleted from, the Mental Scale were relevant to this age range. In addition, this particular age range is critical in infant testing because language is starting to become established and a sharp increase in vocabulary occurs around this age (Reznick & Goldfield, 1992).

8-1-2 MDI gains

During the last few decades, there have been rapid economic growth and considerable changes in lifestyle and social welfare in most cultures, but particularly in Northern American and Western European contexts. Changes have included an increase in parental education and improved nutrition. These, combined with the conviction that a highly stimulating and enriched environment and early experience are important to a child's development, have increased and accelerated children's mental processing in recent years. For example, Lynn, Hampson and Mullineux (1987) reported an average IQ gain of 12 points for 9 to 11 year old English children. Emanuelsson and Svensson (1990) also

reported an increase in spatial and reasoning intelligence among 13 year-old Swedish children during a 25 years period. Based on adult scores from intelligence tests in 14 nations, Flynn (1987) concluded, that "... the generational gains range from 5 to 25 points and the median is 15 points, or a full standard deviation" (p. 184). Recent data have indicated that test score gains are continuing over time (Teasdale & Owen, 1989).

In the field of intelligence assessment of young children, however, there is little information available regarding IQ gain. In one study, Fuggle, Tokar, Grant and Smith (1992) found that after nearly 20 years 5-year-olds demonstrated an average IQ gain of 8 points in a full scale WPPSI (Wechsler, 1967) IQ score. Campbell, Siegel, Parr and Ramey (1986) reported that the Bayley Scales overestimated their sample's performance by about 12 points. However, it is still unknown whether the gains are identical for both genders. The possibility of higher IQ gains for girls, for example, has never been considered, but seems likely, given that girls score higher on early childhood tests than boys, possibly because of their ability on average to speak at an earlier age and their larger vocabulary span. Despite some evidence for IQ gains at different ages in childhood, little is known of IQ gains for infants aged 27 months and younger. A second aim of the present study, then, was to focus on IQ gains, if any, in an Australian infant sample by the simultaneous administration of the original and revised versions of the Bayley Scales.

8-2 METHOD

8-2-1 Subjects

A cohort consisting of 97 subjects (50 males, 47 females) in four different age groups participated in the present study. There were 28 subjects 18 months of age, 20 subjects 21 months old, 22 subjects 24 months old and 27 subjects 27 months of age. The

numbers of females and males in each group were approximately equal. In order to compare the scores on the Bayley Scales (Bayley, 1993) with scores on the fourth edition of the Stanford-Binet test (Thorndike et al., 1986), 42 of the subjects who participated at 18 and 21 months of age were again visited at two years of age when the first age level of the Stanford-Binet test is applicable, and tested with both the Stanford-Binet and the Bayley Scales. This increased the 24-month old sample size to 64. In addition, to assess the test-retest reliability of the Bayley MDI between 24 and 27 months, 12 children who initially were tested at 24 months were retested at 27 months of age.

All subjects were normal, healthy infants with no prenatal or obstetrical complications as indicated by their personal health records. Table 8-1 presents the demographic and birth status of these subjects. A full description of the demographic characteristics of the 97 infants has been reported in Studies 1, 3 and 4.

Table 8-1

Demographic Characteristics and Birth Status of the Sample

Variables	N	M	SD
Gestational age (weeks)	97	39.5	1.4
Birth weight (kilograms)	89	3.6	.54
Maternal age (years)	95	31.1	4.4
Maternal education (years)	97	13.6	2.4
Paternal education (years)	92	14.1	2.9
Status of occupation (mother)	66	3.6	.87
Status of occupation (father)	85	3.7	1.1
Birth status and ethnicity	N	%	
Male	50	51.5	
Female	47	48.5	
First born	57	58.8	
Later born ^a	40	41.2	
English speaking families	74	76.3	
Non-English speaking families ^b	23	23.7	

Note: ^a Second born (N = 24), Third born (N = 12), Fourth born (N = 4).
^b Iranian, Greek, Polish, Sri Lankan, Yugoslavian.

8-2-2 Measures and procedure

The original version of the Bayley Scales of Infant Development (Bayley, 1969) and the Second Edition of the Bayley Scales (Bayley, 1993) were administered to infants at age 18, 21, 24 and 27 months. For the purpose of the present study, only the Mental Scales from both versions were administered. The Stanford-Binet test was also administered to all children when they reached two years of age, as well as to the 24 and 27-month-old subjects. Testing for all subjects was scheduled within two weeks of the child attaining 18, 21, 24 and 27 months, corrected for prematurity. Five tests from the Stanford-Binet were selected for the subjects in this study. They were Vocabulary, Comprehension, Pattern Analysis, Copying and Memory for Sentences. For some cases a single session proved to be too tiring and a second session was therefore required. During testing the parents were instructed to facilitate their infant's performance but not to interfere or assist the examiner unless they were requested to do so. For some children from ethnic backgrounds, limited parental assistance was provided as required for language items. All measures and the procedure were fully described in Study 2.

8-3 RESULTS

8-3-1 IQ gains

The means and standard deviations for the two versions of Bayley MDI and Stanford-Binet IQ scores are shown in Table 8-2. The average means for the four selected age groups were 119.5 and 101.3 for MDI (Bayley, 1969) and (Bayley, 1993) respectively, indicating that today's Australian infants have an IQ (MDI) gain of about 18 (1.2 standard deviation) points when compared to the 1969 norms. This gain was higher than the 12 points reported by Campbell et al. (1986). The differences in gains according to gender and

Table 8-2

Summary Table of Sample, Mean Scores and Standard Deviations on Each Measure

Age Group	N			Bayley MDI (1969)		Bayley MDI (1993)		Stanford-Binet (1986)		
	Initial	Repeat	Total	Mean	SD	Mean	SD	N	Mean	SD
18 months	28	-	28	118.5	15.0	99.0	11.5	-	-	-
21 months	20	-	20	113.3	14.1	97.6	11.0	-	-	-
24 months	22	42	64	122.8	16.9	104.3	11.5	57 ^a	102.7	8.0
27 months	27	12	39	122.3	13.7	103.9	11.1	27	107.1	9.7
Total sample	97	54	151					84		

Stanford-Binet Subtests	N ^b	24 months		27 months		
		M	SD	N	M	SD
Vocabulary	53	102.2	10.9	26	107.3	10.7
Comprehension	50	103.4	9.2	26	104.3	11.1
Pattern Analysis	56	103.9	7.5	27	104.9	8.0
Copying	53	99.1	7.0	27	104.6	11.1
Memory For Sentences	29	104.8	7.8	23	108.0	10.5

Note: ^a - Seven 24-month old subjects did not achieve a score on any subtests of the Stanford-Binet either due to uncooperativeness or inability to perform the test.

^b - Variation in N for the different subtests of the Stanford-Binet occurred because not all children performed all 5 different subtests.

ethnicity background are shown in Table 8-3. Rising mental functioning in females was marginally higher than for males ($t(95) = 1.85, p < .10$) and it was significantly higher for Australian infants than for those children from an ethnically different background ($t(95) = 2.2, p < .05$). One-way ANOVA indicated that the observed differences in IQ gains within the four age groups were similar. However, the increase for higher scoring infants (i. e., 132-150) was higher than for infants from the normal range including infants with IQs one standard deviation below the mean ($F(3, 96) = 22, p < .001$; Appendix I).

Table 8-3

Comparison Between Two Versions of Bayley MDI (IQ Gains) for Total Sample and According to Gender and Ethnic Background

Sample	N	Bayley MDI (1969)		Bayley MDI (1993)	
		Mean	SD	Mean	SD
Total sample (age partialled)	97	119.5	15.1	101.3	11.5
Gender					
Female	47	123.9	15.1	104.7	11.7
Male	50	115.3	14.0	98.4	10.7
Ethnicity Background					
English speaking families	74	121.1	14.3	102.1	11.1
Non-English speaking families	23	114.3	16.8	98.6	12.4

8-3-2 Comparison between two mental scales (MDI)

It is evident from Table 8-2 that the MDI from the first edition were consistently higher than those from the second edition. A series of t tests confirmed significant differences for all four age groups (t ranged from 11.7 to 17.3; all p s $< .001$). A series of t tests were also calculated to establish if there were any differences between gender and ethnic background. The results demonstrated that, for this age range, females scored

consistently higher than males on both versions. However, the differences were significant only at 18 months for both the first edition ($t(26) = 2.3, p < .05$) and the second edition ($t(26) = 2.2, p < .05$) of the scale. Infants from English speaking families and those from ethnic backgrounds performed at comparable levels (t ranged from .68 to .14; all ps NS). Correlations obtained at 18, 21, 24 and 27 months between both versions were highly significant for all groups, ranging from .84 to .93. The magnitude of these correlations was higher than those reported in the manual of the test ($r = .63$) mainly because of the wider age range reported in Bayley's study.

8-3-3 Reliability and stability of the MDI

The internal consistency of the new MDI was explored using Cronbach's alpha (Table 8-4). The test was highly reliable, alpha ranging from .81 to .85 and thereby supporting the reliability coefficient and the standard error of measurement obtained by Bayley (1993). A longer-term stability of the MDI scores was computed, defined as consistency of the scores from 18 to 24 months, from 21 to 24 months, and from 24 to 27 months of age. As can be seen from Table 8-4 (lower portion), the new version was highly stable across 6 months. The magnitude of the correlations increased with age but this was confounded with shorter intervals of time. This was true for both versions of the Scale, and was consistent with literature in the field (e.g., Lewis et al., 1986).

8-3-4 Comparison of the Bayley Scales with the Stanford-Binet Test

The relationship between the MDI (Bayley, 1993) and the Stanford-Binet test was investigated by administering both measures to all subjects either at 24 or 27 months of age (Table 8-5). With the exception of Memory for Sentences at 24 months, which was not significant, all correlations between the two tests were highly significant. Generally, correlations increased in all cognitive areas at older ages.

Table 8-4

Reliability Coefficients (α), Standard Error of Measurement (SEM) and Test-retest Stability of the Mental Scale of Four Age Groups

Reliability	18 months	21 months	24 months	27 months
α	.85	.83	.84	.81
SEM	4.45	4.53	4.60	4.83
N	28	20	22	27

Stability (across months)	18/24	21/24	24/27
Bayley MDI (1969)	.63***	.82***	.76**
Bayley MDI (1993)	.79***	.86***	.92**
N	24	18	12

** $p < .01$ (two-tailed). *** $p < .001$ (two-tailed).

Table 8-5

Correlation Between MDI (1993) and Stanford-Binet Scores at 24 and 27 Months of Age

Stanford-Binet subtests	24-months Bayley MDI (1993)		27-months Bayley MDI (1993)	
	N	r	N	r
Vocabulary	53	.80***	26	.82***
Comprehension	50	.62***	27	.51**
Pattern Analysis	56	.43***	27	.62***
Copying	53	.42**	27	.66***
Memory For Sentences	29	.18	23	.66***
Composit score	57	.76***	27	.88***

Note: Variation in N for the different subtests of the Stanford-Binet occurred because not all children scored all 5 different subtests. ** $p < .01$ (two-tailed). *** $p < .001$ (two-tailed).

8-4 DISCUSSION

The second edition of the Bayley Scales (Bayley, 1993) was standardized on a sample representative of the American infant population aged between 1 and 42 months with a multistep process. The test materials have been successfully redesigned, and

directions for administration, scoring rules, and basal and ceiling levels are sufficiently clear. It has been found that the MDI of the new version has high internal consistency with a low standard error of measurement, suggesting that the items are measuring a common entity. The stability of the scale across 18 to 27 months of age was also high, indicating that scores are stable over time. This result is consistent with other research, which has shown greater stability over a short interval of time, and confirms the proposition that intelligence becomes increasingly stabilised around two years of age (Lewis et al., 1986; McCall et al., 1972).

The correlation found in this study between the first version and the second version of Bayley mental scale was higher than that reported by the author of the test (Bayley, 1993), as the current study recruited subjects in a very narrow age range (18 to 27 months) in comparison to the wider age range (1 to 42 months) tested in Bayley's study. Within-subjects comparisons of MDI scores from both versions of the scale have also shown that the early version of the Bayley Scales no longer provides a reliable measure of an infant's intellectual activities and use of the older instrument should be abandoned, at least for infants aged between 18 and 27 months.

A tendency for gender differences has repeatedly been reported by other researchers using the first version of the MDI. However, this is the first study to find significant differences for 18 month old infants with the second version. One possible explanation for such a difference lies with the composition of the scale itself. Both versions of the MDI are heavily dependent on vocalisation and, since language development begins earlier in girls (Goldfield & Reznick, 1990; Reznick & Goldfield, 1992), the significant results may have been due to a different level of vocabulary development rather than intellectual ability.

The high correlations between the Stanford-Binet test and MDI from the new Bayley

test supported the concurrent validity of the mental scale. The highest correlation was found between Partial Composite Standardised Age Score and MDI, indicating that the mental scale is highly related to crystallised, fluid and memory abilities in 27-month-old children. Similar results were reported by Bayley (1993) when the mental scale was correlated with the WPPSI-R. Although further analysis suggested that the MDI was more related to a vocabulary factor than to comprehension, these results must be interpreted cautiously since the number of subjects is small and the sample relatively homogenous. The means of the MDI at 24 and 27 months were close to the Partial Composite SAS, from the Stanford-Binet test, suggesting that the MDI is measuring a theoretical construct similar to that assessed by the Stanford-Binet, presumably a *g* factor.

The present research has extended findings by Flynn (1984, 1987), Fuggle et al.(1992) and Kitchen et al.(1989), pointing to increasing IQ to the infant population. Based on a comparison of two MDI scores normalised nearly 25 years apart, Australian infants showed substantial MDI (IQ) gains of between 4 and 35 points (Mean = 18). This was the case for infants whose scores ranged from approximately one standard deviation below, to two standard deviations above the sample's mean. Previously some have found that gains are more concentrated among infants with lower intellectual ability (Lynn & Hampson, 1989; Teasdale & Owen, 1989), although Spitz (1989) reported that IQ gains were evident only within the average range and diminished gradually for both below and above the average range. However, the results of the present study suggested that the IQ gains were evident for all levels of intelligence but most marked in subjects with higher mental functioning. The gains were evident for both genders, and for infants from different ethnic groups and of different socioeconomic status. However, they were significantly higher for children born to Australian parents than for those born to families from an ethnic

background.

The IQ gains obtained here for Australian infants were higher than those published by Bayley (1993) for American infants aged 1 - 42 months. They also differed from those in Campbell and colleagues' study (1986). The latter difference supports the work of Lynn and Hampson (1989), because Campbell and colleagues' subjects were tested 10 years ago, and in the intervening period IQ gains may have continued to increase. Considering Campbell and colleagues' data, and the data from this study, it is recommended that the Bayley scale should be restandardised after a maximum of 8 to 10 years have passed in order to be reliable. These data are also consistent with those of Kitchen et al.(1989) who reported a mean WPPSI IQ score of 114.6 for older Australian children. The present study assessed one sample with two different normalised tests, and the differences observed in intelligence scores may be attributed to increasing levels of specific skills. Intervening factors such as socioeconomic status and parental lifestyle were kept constant in this study, and the method is obviously preferable to that of comparing two different samples over a specific period of time, when intervening variables may confound the results.

While the increase in intelligence scores for adults has been attributed to changes in the education system (Emanuelsson & Svensson, 1990; Teasdale & Owen, 1987, 1989), to changes in the curricula in school (Lynn & Hampson, 1989), and improvements in living standards and nutrition (Lynn, 1989), Fuggle et al. (1992) hypothesised that a rise in IQ for children may be associated with changes in their growth pattern. It is clear that school education cannot be assumed to be a factor affecting IQ gains in preschool infants. However, environmental factors and improvements in parental lifestyle may provide a reasonable explanation for the increases reported here. These factors could include quality of nutrition, highly supervised prenatal periods, the advanced technology used in obstetrical

care for pregnancy and childbirth and quality public hygiene. The adverse effect of poor nutrition is well documented. Hendrichsen, Skinhoj and Andersen (1986), for example, found that in monozygotic twins the lower birth-weight twin has a lower performance IQ (the low birth-weight for one twin is normally due to insufficient prenatal nutrition compared with the other twin). Further, today's infants in a developed country are exposed to a more enriched environment than was the case a generation ago, through television, responsive toys and the availability of intervention programmes. The parents' level of education, at least for older children, could provide a cognitively stimulating environment for learning and early experience, thereby influencing IQ gains.

In summary, data from this study suggest that the mean IQ for the infant population becomes higher over time, and normative data should be reviewed regularly. In ongoing longitudinal studies, and in the recent studies in which the Bayley Scales (Bayley, 1969) have been used, the discrepancy between the old and new versions must be taken into consideration for any interpretation of the MDI. The mental scores derived from the original version, particularly for infants who scored more than one standard deviation above the average, are severely overestimated. For example, some subjects who scored extremely high (i.e., between 140 and +150) with the 1969 edition, could only achieve an MDI score between 114 to a maximum of 134 with the new version. Conversely, two Down's Syndrome infants in the present study (but excluded from the analysis reported here) who scored around 65 with the old version of MDI scored below 50 with the new version. All this evidence suggests that the old version of the Bayley Scales (and probably all other developmental scales that have not been updated) should no longer be in use.

CHAPTER 9

GENERAL DISCUSSION AND CONCLUSION

The current research project investigated the reliability and predictive validity of visual recognition memory tasks, in comparison with those obtained from a standard developmental test, among normal infants aged from 6 to 16 months. The main objective of the studies was to address whether, and to what extent, individual variation in infant performance on visual recognition memory tasks and other tests of perceptual/motor development could account for individual differences in mental functioning at a later point in time. Further, if infancy scores were found to predict later intellectual outcome, the research aimed to establish the age at which such prediction could be made with confidence. Within this context, following a brief discussion of the theories underlying infancy tests and a review of the studies undertaken, consideration will be given to the following issues: (i) developmental trends in the infant's cognitive skills and visual fixation measures, and intercorrelation between fixation measures; (ii) the empirical findings concerning the psychometric properties of the information processing test used in this thesis; (iii) the relationship between novelty tasks, conventional infant tests and later IQ, and the failure of infancy tests at early ages to predict later mental outcome. Limitations of the research program, suggestions for further research and final comments will complete this chapter.

9-1 THEORETICAL BACKGROUND

At least up until the early 1980s, the predominant theory of infant intelligence was based on the assumption that an infant's later cognitive development could not be predicted

from a score on sensorimotor scales made during infancy. Empirical evidence convinced researchers that the nature of intelligence in infancy is qualitatively different from that in childhood, and that in order to attain cognitive competencies analogous to the symbolic and abstract skills in older children and adults, infants necessarily pass through several distinct stages (Piaget, 1954). This led investigators to conclude that the nature of intelligence is discontinuous (Wilson, 1983). The theory of discontinuity assumed that intelligence is composed of a set of separate but related mental abilities rather than a single, general factor (Lewis et al., 1986) and, thus, a "g factor" was not applicable to infant intelligence (Lewis & McGurk, 1972). Alternatively, a g factor does exist in infancy, but its nature is subject to change across time (McCall et al., 1972, 1977).

However, the empirical evidence outlined in Chapter 2 contradicts this "well-established" view of the discontinuity of intelligence. Fagan (1984a) and Sternberg (Berg & Sternberg, 1985; Sternberg, 1981) challenged the assertion that the development of intelligence was discontinuous, and surmised that some aspects of intelligence, for example, an infant's ability to deal with novelty (preference for novelty), is continuous throughout life. This view led researchers to reconceptualize the development of intelligence in an infant's early years, and stress factors such as attention, memory and the speed of processing of information. Fagan, for example, equated variations in g with individual differences in the speed of knowledge acquisition.

The new approach is based on information processing theory. The theory assumes that speed of information processing constitutes a fundamental aspect of cognitive development during infancy and shares some common variance with later intelligence tests. One operational definition of speed of processing identifies an infant's ability to differentiate between familiar and new stimuli following her/his exposure to the familiar stimulus, within

a specific time frame. According to the continuity model, the relationship between an infant's performance on a novelty task and a childhood intelligence score is mainly mediated by the speed with which an infant encodes, abstracts and memorises the task, and solves new problems. An influential version of the information processing model has emphasized that the speed with which information is perceived and processed is biologically determined and stable during development (Anderson, 1992). According to the theory of minimal cognitive architecture, individual differences in intelligence are determined by the speed of a basic processing mechanism. Anderson hypothesized that the new infancy measures also reflect the speed of a basic processing mechanism. He concluded that "The real predictive variable is not novelty preference, but processing speed" (Anderson, 1992, p. 126).

9-2 SUMMARY OF STUDIES AND RESULTS

This thesis consisted of four studies and two supplementary examinations of data. Studies 1 and 2 were designed to test the reliability and predictive validity of the Fagan procedure. Studies 3 and 4 were exploratory in nature and investigated the possibility of extending the Fagan procedure to older infants using a different scoring system and different formulations for familiarization time. Figure 9-1 shows the observed relationships between infancy measures and later MDI/IQ across all studies.

9-2-1 Fagan Test Of Infant Intelligence (Study 1 & Study 2)

The aim of the critical review of the FTII was to examine the internal consistency and test-retest reliability of all fixation measures obtained from the FTII, and to analyse the novelty items constituting the test. The second objective of the study was to address the predictability of the FTII after a one year interval of time for all subjects, as well as a 15 month interval for 9-month-old infants and an 18 months interval for 6 to 7 month-old babies.

Figure 9-1

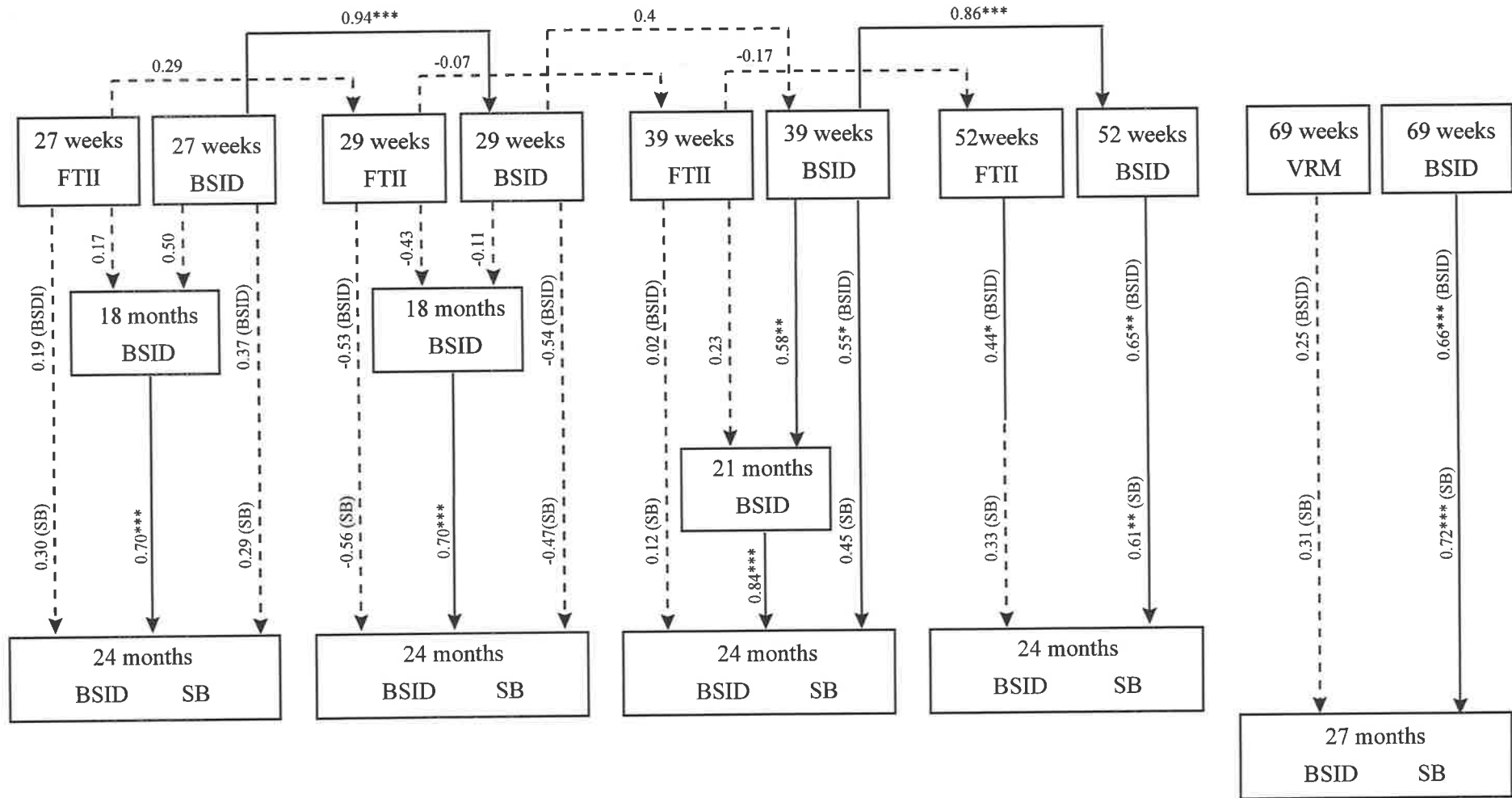


Figure 9-1 illustrates the intercorrelations between infancy measures and later MDI/IQ in childhood. (Solid lines represent significant correlations and broken lines represent non significant correlations).

Correlations reported for the BSID are based on the original version of the Bayley Scales (Bayley, 1969).

The correlations between infancy measures and the Stanford Binet are based on Composite Standard Age Score (IQ).

The results of Study 1 clearly showed that the FTII, within the age range studied here, is not a reliable test in terms of psychometric analysis. There was little indication of homogeneity between novelty items for any age-related test and the test-retest reliability model was violated. That is, the percentage of preference for novelty was not stable across ages. The item analysis conducted on the FTII confirmed that some novelty items were not discriminable and several other items were not related to the scale. The hypothesis based on the prediction of a meaningful intercorrelation between all fixation measures also failed.

The results of Study 2 indicated no predictive validity for the FTII in younger infants, and only limited evidence for the prediction of later IQ for the oldest age group (i.e., 52 week-olds). The 52 week novelty scores predicted the MDI from the Bayley Scales at 24 months, but only some tests of the Stanford-Binet at the same age; from 8 expected correlations between novelty scores and later MDI/IQ, four correlations were significant. In addition, the FTII did not predict the Composite Standard Age Scores (General IQ) obtained from the Stanford-Binet. Moreover, there was an obvious error rate in classifying infants in terms of normality. Thus, the results did not support the claim made by Fagan and his colleagues in relation to the predictive validation of the test in a normal population (Fagan & Detterman, 1992; Fagan & McGrath, 1981).

9-2-2 New Visual Recognition Memory Task (Study 3 & Study 4)

The results of the second study raised a research question for the next study. Having established that the FTII showed only a limited predictability for the oldest group of infants, Studies 3 and 4 investigated the usefulness of the novelty preference measure in predicting later IQ for 16 month old infants, using a different method for measuring novelty scores. The low internal consistency of the FTII led the author to hypothesise that using a time-fixed procedure for scoring novelty preference might result in higher internal consistency for

the task. It was suggested that the low internal consistency of the task could be a function of expiring a time during novelty trials that results in chance performance. Thus, using a time-fixed procedure, an infant would have more opportunity to choose deliberately the right stimulus (the new one) even if s/he failed in his or her first attempt and looked at the familiar target. Ultimately, this procedure would reduce chance performance and one would expect a higher internal consistency across the novel problems. Although the results did not support the hypothesis that a new scoring system would increase the internal consistency of the test, the timed-test procedure clearly resulted in a more uniform performance across different novelty items than the traditional expiring time procedure. Almost all items exceeded chance performance. Similarly, the results of Study 3 did not support the predictive validity for the novelty task for older infants. Instead, the study clearly showed that predictability was influenced by outlier subjects. That is, the prediction outcome was enhanced by the inclusion of Down's Syndrome infants.

The working hypothesis for Study 4 resulted from Study 3, in which it was observed that older infants were more self-regulated and able to monitor steadily their attention to a task. Therefore, they may not need to accumulate the total familiarization time. Moreover, accumulating familiarization time may result in a ceiling effect in an infant's performance and make the test non-representative of the infant's actual competencies. However, the results indicated that performance for most infants remained at chance level, and the novelty scores obtained from this procedure did not predict later outcome. This suggests that, although the infants processed the familiar stimulus, they did not complete processing of the new target. Hence they did not show more complete processing and were not able to discriminate between novel and familiar targets. However, the study created a working

hypothesis for further research based on using familiarity preference for prediction of later IQ, rather than novelty preference.

9-2-3 Bayley Scales of Infant Development (Two complementary examinations of data)

The primary objective of the examination of the data obtained from the Bayley Scales was to investigate the predictive value of infants' conventional developmental scales (MDI and PDI), and to determine the first age, if any, during which the Bayley Scales could predict childhood mental capacities. The study provided sufficient evidence to suggest that the Bayley Scales can predict later outcome at 2 years of age and beyond when administered at 9-12 months or later, but not when administered at approximately 6-7 months of age. This suggests that 9 months of age might be the earliest age at which prediction is possible. However, the Bayley Scales during infancy (like the FTII) did not identify any of four infants who scored around 70 MDI points later at 24 months.

Reanalysis of the data from the Bayley Scales in all studies disclosed an average MDI gain of 18 points for young Australian children during the last 25 years. Compared with their Australian counterparts, the ethnic sample gained significantly fewer MDI points (14 points). As a consequence of such a gain, it was concluded that the original version of the Bayley Scales is no longer a reliable measure for infants aged 18 to 27 months, and probably not for younger infants.

9-3 DEVELOPMENTAL TRENDS IN COGNITIVE SKILLS AND FIXATION MEASURES

Considered together, the results of Study 1 and Study 3 indicated that, while the number of fixations and exposure time underwent a developmental change during the first 16 months of life, novelty scores did not change during the same developmental period.

This is consistent with Colombo et al.'s (1988) and Jacobson et al.'s (1992) studies, which found that novelty scores did not change during the first years of life. Fagan (Fagan, 1971; Fagan & Shepherd, 1979) reported that infants, by 5 to 6 months, are able to remember faces of a different sex, different age and different orientation. Therefore, if the level of complexity of stimuli and the familiarization time for the stimuli to be remembered are adjusted in accordance with the infant's age, no developmental change is expected for the percent of preference for novelty. In this case, a relatively shorter familiarization time, or more difficult stimuli, may result in an indication of developmental change in performance on novelty tasks. For example, a significant change for novelty scores was observed in Study 4 because familiarization time was expiring, and the infants had a relatively short familiarization time compared with infants of the same age in Study 3.

The means of novelty scores, in general, have been found to be only slightly higher than chance performance and this was also the case for the the FTII and the new novelty task used in Study 3. Across all age groups, mean novelty scores were clustered around 62% to 64%. However, the percentage of looking to novel stimuli generally fluctuated across different novelty items and some novelty items were not discriminated. One possible reason for the low mean novelty scores (i.e., slightly above chance) may be that, at the level of each individual subject, novelty scores across all items (10 items in this study) are averaged and, as a consequence, those items with a preference score higher than 50% are affected by the scores less than 50%, resulting in chance performance, or slightly above, overall. Evidence for this reasoning was found in the large standard deviation in novelty preference scores (up to 36%) for each subject.

In relation to the developmental trend for the two other looking measures (number of fixations and exposure time), two important findings were reported for the first time as part

of the present project, and these warrant further consideration. The first was a developmental change evident in the number of fixations. Consistent with the research hypothesis, the number of fixations during the second 6 months of life increased significantly. That is, the length of each glance became shorter with increasing age, and 12 month old infants had the highest rate of fixations in familiarization time. However, beginning with the second year of age, and perhaps because of a transition in cognitive functioning, a reverse trend for this measure was documented. It was found that the length of each glance increased dramatically, and that 16-month-old infants had significantly fewer fixations during the familiarization phase. This implies that infants, at this age, can look steadily at a stimulus for a longer period of time.

The second noticeable developmental change in looking behaviour occurred in exposure time, the time an infant needs to accumulate the familiarization time. One hypothesis in Study 3 was that exposure time in older infants would be lower than in younger infants. This hypothesis was supported. Despite no developmental change for this measure during the second half of the first year (see Study 1), exposure time showed a sharp decrease at 16 months of age. A 16-month-old is able to concentrate more effectively on a task and spends less time exploring the context beyond the task. Observation of infants during testing showed that as soon as these infants engaged in the task they continued to process stimuli and, in general, the interruption between fixations was very short. In other words, the highest rate of exposure time was due to the latency time an infant needed to be involved in the task. This implies that a 16-month-old infant should be regarded as a more focused subject who can be more readily relied on than a younger baby to complete her/his task.

In contrast to the novelty preference score, there was an obvious developmental trend for the MDI from the Bayley Scales. The 24 month and 27 month MDIs were significantly higher than MDI during infancy. Significant differences observed between the MDI during infancy and childhood could be attributed to the environmental influences that affected the cognitive development in the second year of life. In addition, as documented by, for example, McCall et al. (1977) and Plomin et al. (1990) a critical transition occurs in an infant's cognitive ability during the second year of life. The significant decline in cognitive scores from 1 to 2 years found by Rose et al. (1989) for low SES subjects, is consistent with an environmental interpretation. As the infant enters its second year of life the influence of environmental factors becomes more important (Bradley et al., 1989).

9-3-1 Intercorrelation between fixation measures

If it is accepted that infant intelligence is validly measured by a response to novel stimuli (Fagan, 1984a, 1992), then it would be expected that there would be intercorrelations between all, or at least some, measures derived from novelty tasks. At least three different measures can be extracted from data obtained from the FTII, as well as other potential measures. For example, the length of each fixation during familiarization time, the length of the first fixation to a familiar stimulus, the pause between each fixation, and the latency time to begin the first fixation could all be recorded. However, the test scoring system in the FTII is not programmed to record these measures.

The pattern of intercorrelations between fixation measures for young infants has not always been consistent (e.g., Rose & Feldman, 1987), and the results of the present study were no exception. Nevertheless, there were two important findings in the present study that warrant more consideration. First, the present study rejected the notion that short-looking behaviour indicates faster processing. Colombo et al. (1988) reported that infants

aged 4 to 7 months who have a higher number of fixations are "brighter" (in terms of higher novelty scores) than those who show a smaller number of fixations. Colombo and Mitchell (1990) have suggested that short-looking infants may have a more integrated central nervous system. The present project found a significant correlation between the number of fixations and novelty scores only for subjects aged 39 weeks. However, this relationship was not replicated either for the sample as a whole or for other individual age groups. Similarly, there was no relationship between an infant's MDI from the Bayley Scales and the number of fixations, suggesting that infants who look briefly are not necessarily brighter infants.

Second, an interesting finding was the significant negative correlation between novelty scores and exposure time for infants at 16 months of age. Such an intercorrelation was not evident for infants aged 12 months and younger. This result implies that bright infants (in terms of having higher novelty scores) have less time-off-task and spend more time focusing their attention on task. The fact that infants at this age had less exposure time and fewer numbers of fixations than younger infants indicated that the pattern of looking behaviour became more systematic with age. As a consequence of this developmental trend, a negative correlation between novelty scores and exposure time became significant, suggesting that both exposure time and novelty scores could be viewed as indices of mental functioning for older infants.

9-4 RELIABILITY OF VISUAL RECOGNITION MEMORY TASKS

The results of Study 1 and Study 3 suggested that the novelty tasks as used here have a very low internal consistency and test-retest reliability. All novelty items are intended to measure a single dimension of cognitive activities but the results showed little evidence of that. The low reliability for this type of task is a major criticism of the procedure (Pomerlea

et al. 1989; Malcuit et al., 1988a, b). Although the low reliability for the Fagan procedure was consistent with the literature for this type of task, the low reliability is not acceptable for a commercial test. When marketed, a test is assumed to be reliable and to measure the abilities which theoretically underpin the test. An unreliable test may be effective on one occasion for one sample but not effective on a second occasion for another sample, or even for the same subjects. However, Fagan and Detterman (1992, p. 184) acknowledging the low reliability for the FTII, have argued that "an unreliable test can be as valid as any other test".

While most researchers have argued that the rapidity of developmental change during infancy accounts for the instability of attentional measures, others have maintained that the nature of stimuli used, the number of components constituting the test and the sample characteristics have contributed to the low internal consistency and test-retest reliability. It is suggested that one reason for the unreliability of the FTII may lie in the social nature of the stimuli used in the test. Despite all items in the FTII belonging to the same category (i.e., a human face) the shared variance between novelty items is very low. It is possible that the infants internalized different aspects of the social stimuli (e.g., different affective expression, distinction between age, sex) on different trials or testing sessions, leading to instability of the results within and between sessions. Fagan and Detterman (1992) have argued that, because face recognition is a multifaceted process, it is unreasonable to assume that all items measure the same cognitive process. Using color slides of individuals' faces, Colombo et al. (1986) have also found no consistent test-retest reliability for the presence or magnitude of dishabituation measures. In another study using alphabet letters (which seem more similar than human faces) Colombo et al. (1988) also

found low internal consistency for the novelty task. However, they reported significant cross-age stability ($r = .39$ to $r = .51$) for novelty scores obtained from these type of stimuli.

As discussed in Chapter 4, another aspect of the FTII that might contribute to its low reliability, is the number of novelty items that did not correlate with the total scale. The low intercorrelation between novelty items supports the notion that the FTII probably is not measuring a single common entity operationalized as novelty preference. This was also the case for the novelty task used in Study 3. It was suggested that the low internal consistency and high number of novelty items that were not discriminated might be due to the procedure with which novelty scores are measured. However, a new approach to measuring preference for novelty (accumulating test time as used in Study 3) suggested that low reliability is not exclusively a function of the methodology involved in collecting data for novelty scores.

Berg and Sternberg (1985) argue that response to novelty in early infancy involves both motivational and information-extraction components. Since infancy tests are generally affected by an infant's temperament and mood, it is not unlikely that instability in results is due more to different moods and/or temperament at different times, than to the infants' ability to process information. A similar line of reasoning suggests that young infants may not have developed the ability to engage themselves in an attentional task - merely looking at the stimuli - which results in a consistent pattern of looking behaviours. As a consequence, their performance on visual recognition memory tasks would not be expected to be consistent even over short periods of time.

Yet another factor contributing to the low reliability, suggested by Rose et al. (1988a), is the social class or status of the infant's parents. They argue that attentional problems in the low SES infants may play a dominant role in the poor reliability of novelty tasks. However, the present study does not support this; the sample in the current study

were mostly infants from middle-class parentage and hence were unlikely to be affected in this way.

Considered together, the results are in agreement with DiLalla et al.'s (1990) interpretation, that the FTII measures different aspects of cognitive abilities at different points in time. It could be concluded that the measures obtained from novelty tasks are not reliable regardless of how they are measured and at which age (during early infancy) they are collected, mainly due to the fact that the fixation measures themselves are questionable, at least with the paired comparison method.

9-5 THE RELATIONSHIP BETWEEN NOVELTY SCORES AND LATER MDI/IQ

Based on the empirical evidence reviewed in Chapter 3, preference for novelty during infancy would seem to have some shared variance with later IQ in childhood. However, data from the present study have failed to find a sufficient degree of support for this hypothesis. The finding of a low predictive validity for the Fagan procedure is not exclusive to this project. Work conducted by Andersson (1996) has also found that the relationship between novelty scores from the FTII and later IQ was low and inconsistent across different mental outcomes.

The poor predictive validity of the FTII challenges both the theoretical base and technical construction of the test. On one hand, data clearly indicated that preference for novelty was achieved. On the other hand, the novelty scores did not predict later outcome. This raises the possibility that either the FTII does not sufficiently indicate preference for novelty or the measure of novelty preference obtained bears no relationship to later IQ (i.e., the theory is not a tenable proposition). As was clarified in the literature review, it appears unwise at this stage to dismiss the usefulness of the theory underpinning preferential

looking. Clearly, an infant's ability to distinguish a new target from a familiar one is a reflection of some aspect of cognitive development. Response to novelty does appear to be also an important facet of intelligence in infants and adults (Sternberg, 1985). However, the task intended to measure novelty preference may not function as intended. If this is the case, the factors that affect the accuracy of the percentage of preference for novelty need to be identified. Visual fixation measures, including preference for novelty, are generally affected by many uncontrollable factors and, therefore, may or may not be taken as a measure of novelty preference, at least in early infancy.

In actual fact, there are several procedural difficulties that may affect the novelty score measures. Some potentially inhibiting factors have been discussed in Chapter 5. One explanation may be that fixation measures are not necessarily measures of information processing. For example, the "total looking" at novel (or familiar) stimuli that is recorded by the examiner may not be representative of an infant's *attention* to the targets. At least some of this looking time may be spent just *staring* at targets (Lecuyer, 1989) without any processing involved. Using heart rate, Richards (1988) has differentiated between fixation on an object and a more attentive looking. Thus, one possible explanation for the unpredictability of the novelty task may be that, with the Fagan procedure, one cannot be confident that the novelty score reflects intentional looking and an accurate measure of the rate of information processing.

If this is the case, researchers need to search for alternative measures of information processing. In fact, most researchers who have found a relationship between information processing tests during infancy and later IQ maintain that the speed of processing mediates the predictability of later IQ. For example, Rose (1980, 1983) hypothesized that the percentage of preference for novelty is a function of the speed of processing information

from the environment. In this case, the percentage of preference for novelty is, indeed, an indirect index of the speed of encoding the information content of the stimulus to be remembered. If prediction is mediated by speed of information processing, a direct measure of speed of processing would more reliably predict later outcome. The literature suggests that measures such as reaction time and inspection time are reliable indicators of speed of information processing for older children and adults (Kirby & Nettelbeck, 1991; Anderson, 1986; Nettelbeck & Wilson, 1985; Jenkinson, 1983). While there is no direct measure of inspection time for infants, some types of RT can be measured in early infancy. For example, DiLalla et al. (1990) reported that the baseline RT obtained from the visual expectation method (as described in Chapter 3) at 8 months did, in fact, predict later Stanford-Binet at 3 years of age ($r = .47, p < .01, N = 33$; in this study RT scores were reversed).

Nevertheless, it should be noted that it is still not well documented that the predictability of novelty tasks is mediated by measures of speed of information processing. It is entirely possible that the novelty task is measuring such diverse behaviours as speed of processing, motivation, stimulus appeal (instead of preference for novel stimulus) and/or conditioned learning (Malcuit et al., 1988a). Some researchers hold that better performance on novelty tasks is related to language acquisition (Thompson, et al. 1991; Rose et al., 1991b), short-term memory (Colombo et al., 1989) or a g factor (Fagan, 1984a, see also Fagan [1984b and c] for evidence indicating that visual recognition memory is strongly related to intelligence). McCall (1994), for example, pointed out that, in addition to processing speed, it is possible that an infant's ability to inhibit his/her attention to a familiar or less salient target constitutes the basis for prediction from novelty tasks. It is even feasible that predictability is not a function of novelty preference but is mediated by other

factors such as social class of parents. Kagan (1989) argues that the middle-class infant's faster rate of processing may be due to a more conscientious, accelerating mother who provides a better context for her infant in a test situation.

Apart from the problems associated with the fixation measures, some plausible explanations can be considered for the low predictive validity of the FTII and VRM task. One factor to be considered is the infant's age at the initial and follow-up testing. At least in terms of the FTII, almost none of the novelty scores obtained before 12 months of age predicted later IQ, supporting the notion that the novelty task for young infants, at least within the paired comparison method, cannot predict later mental outcome. Looking measures for young infants may be too fragile to be relied on as measures of cognitive development. It is possible that 12-month and older infants can explore the environment more effectively and show more effectual understanding of the task. However, this possibility is not endorsed by the literature, since the relationship between infants' information processing tests and later IQ was generally found to be higher for younger infants aged between 2 and 8 months (McCall & Carriger, 1993; Slater, 1995). Similarly, the age at which prediction was made (follow-up testing) may reduce the possibility of prediction. That is, 24 or 27 months of age may not be an appropriate age at which to predict later IQ (see below) since general intelligence may still not be stable (Humphreys & Davey, 1988; Siegler & Richards, 1982).

In a similar vein of reasoning, one more plausible explanation may rest in the selection of the criterion measures (i.e., Bayley Scales and Stanford-Binet test). The question to be answered is whether and, to what extent, the Bayley Scales and Stanford-Binet at 24 and 27 months can be relied on as measures of cognitive development. From a Piagetian perspective, children of this age are transiting from one stage of development to a

higher stage in which the level of functioning will change. On the one hand, the Bayley Scales may not be an appropriate test for assessing psychometric intelligence because of its emphasis on developmental milestones. For example, it was found that the correlations between novelty scores and the Bayley Scales were generally inconsistent (in some instances the correlation was statistically significant and sometimes not). On the other hand, the Stanford-Binet test at this age may not provide a very satisfactory measure of intelligence. Not only is 24 to 27 months the first age range during which the Stanford-Binet can be administered but some children in the present studies could not achieve the minimum requirements for calculating an accurate measure of their IQ. This suggests that the Stanford-Binet scores may not be representative of the ultimate level of intelligence, and thereby render the predictability of the novelty task unreliable. Thus, it is possible that information processing tasks may provide a more meaningful predictor of later intelligence at preschool ages than they do in the first 27 months of life.

It might also be concluded that prediction from the novelty tasks is affected by the degree of normality of the population tested, and may only predict effectively with the inclusion of at-risk infants (as was the case for the new novelty task). If this is so, then one should question the usefulness of this type of task in a normal population. The point is that many infants who turn out to be mentally retarded appear initially to be normal, and cannot be diagnosed as retarded at birth or in early infancy (Hunt, 1981; Jacobson et al, 1985). To be effective, a predictor test should predict for all infants, regardless of their status as "normal" or "subnormal". In addition, it has long been documented that the sensorimotor tasks are also able to predict later IQ in a population scoring in the below average range of intelligence.

One consequence of the poor predictive validity of FTII was the relatively high rate of error in classification of infants in terms of normality. One reason for such a misclassification rate may rest in the selection of the cut-off score for detecting at-risk infants. The novelty preference cut-off score of < 53% as an indication of subnormality is based on earlier versions of the FTII administered to younger infants (4 to 7 months), and may not be applicable for the current marketed version used here. Both the inability of the test to identify a Down's Syndrome baby and the finding that it classified four normal infants as sub-average (as measured by the Bayley Scales and Stanford-Binet) are clearly unacceptable.

9-6 THE PREDICTABILITY OF THE BAYLEY SCLAES

As outlined in Chapter 7, the Bayley Scales were not capable of predicting later mental development until 9 months of age. From 9 months onwards, the test was found to be a reasonable predictor of later MDI and, at 12 months, was a reliable predictor of both MDI and IQ. One of the most accepted explanations for the failure of traditional infants tests to predict later mental functioning, is that the pool items constituting the test are not effectively measuring symbolic and abstract skills. However, an alternative interpretation is that the nature of intelligence during infancy is markedly different from that in school age children, suggesting that the measurement of general intelligence, as it exists in adults, cannot be obtained before 4 to 5 years of age (Siegler & Richards, 1982). Data from the present research were more in favour of the latter interpretation. First, neither the Bayley Scales nor the FTII predicted later outcome before 9 months of age. Although the FTII is considered to be an information processing task which, in theoretical terms, is measuring the construct of intelligence that is very close to childhood intelligence, it did not predict later

intelligence. Second, some items in the Bayley Scales administered to infants aged 9 months overlap with items used for infants aged 7 months, and it is not unusual for infants to score some items over their chronological age but not to score some items below their age. In other words, the difference between the 7-month MDI and the 9-month MDI is based on only a few more items that were administered to older infants. Moreover, there is a discrepancy in age placement of test items across different developmental scales (Brooks-Gunn & Weinraub, 1983). This suggests that the construct of test items may not be entirely responsible for low prediction. Rather, the unpredictability is perhaps due to some major transitions in the qualitative nature of infants' intelligence that occur during the first year of life.

Another possible explanation for the poor predictability before 9 months may rest in the nature of the relationship between the examiner and infants during testing. Generally, younger infants are not as cooperative as older ones. They are sensitive to the presence of a stranger and their behaviour deviates more from normal. Observation during data collection for the present research, indicated that some infants were hesitant, turned to their mothers and were more passive than active participants in their response to test items. To the extent that the examiner can facilitate the testing session and bring infants to an optimal disposition for testing, this may influence the test results for some infants. That is, not all test results may be representative of an infant's total ability. In fact, there is some evidence to indicate that infants' test scores were significantly correlated to their social responsiveness to the examiner during testing (e.g., Lamb, Garn & Keating, 1981). If this is so, a reliable prediction is unlikely to be made from scores which may not be entirely representative of an infant's competencies.

Some studies have suggested that SES and maternal education in particular, are the best predictors of later IQ (e.g., Laucht et al., 1994). The present study did not support this contention. The failure in the current study to find predictability for SES and maternal IQ is consistent with findings of other researchers who have found low predictability for environmental factors in the earlier months of life (e.g., Bradley, 1989). However, since the project was not designed to measure either the quality of the home environment and/or the interaction between infants and mothers, it is difficult to dismiss the relative importance of SES and other factors of the home environment in early infancy. The limited range of subjects in terms of social status in the current study made it difficult to evaluate the relative importance of parental social standing in predicting later cognitive development.

Finally, this study made a useful contribution to Australian data in relation to the reliability of old developmental infant scales. As a consequence of an increase in infants' mental competencies in the past two or three decades, the old version of the Bayley Scales and, possibly all infant development scales of similar age, are no longer accurate indicators of infants' mental abilities. Given rapid economic growth and a highly enriched environment compared with a few decades ago, it would be expected that infants would be more responsive to the environment and score higher on a cognitive task than they did even 25 years ago.

9-7 CONTINUITY AND DISCONTINUITY ISSUE

A full discussion of the continuity or discontinuity of intellectual behaviours is beyond the scope of this thesis. Considering that the data collected have covered only the first 27 months of the life span, it is very difficult to argue on that basis for the nature of continuity or discontinuity of intellectual development. The data were not totally consistent

with a model either of continuity or discontinuity in intelligence. Rather, they tended to support both continuity and discontinuity, an outcome which is consistent with researchers who do not support either of the two extreme dimensions (e.g., Kagan et al., 1978; Sternberg & Okagaki, 1989).

The data can perhaps be interpreted in terms of one of the models of continuity proposed by Bornstein and Sigman (1986); "continuity of developmental status" because there were some significant relationships between the developmental indices during infancy and later IQ. In this model, Bornstein and Sigman have argued that longitudinal correlations between two different performances at time 1 and time 2 may be found because these behaviours, although totally different, both follow the same developmental trajectories. This model of continuity of developmental status is similar to Kagan's original model of heterotypic continuity (Kagan, 1971). According to Kagan (1971, p. 14), heterotypic continuity refers to "stabilities between two classes of responses that are manifestly different, but theoretically related". Similarly, McCall et al. (1977) hold that the surface character of behaviour changes from one period of time to the next, but the deep structure of character and individual differences remain stable over the developmental period. The instability of individual differences in the first 9-12 months of life observed here may also partly be compatible with McCall's (1981a) proposition that early mental development is highly canalized (i.e., all infants proceed along a species-typical path in a normal environment).

The continuity and discontinuity of intelligence debate leads to the question of predictability or unpredictability of later cognitive development. It seems useful to make a distinction between continuity and predictability to understand the course of development in mental activities. There are two different but related issues that require more consideration.

The first issue is whether intelligence is continuous or not, and the second is whether intelligence is predictable or not. It is conceivable that intelligence is a continuous phenomenon from 9-12 months onward (and perhaps from birth), but not a predictable event, at least at an early age. Unpredictability does not, of course, mean that intelligence is necessarily discontinuous. The unpredictability, as outlined above, may be due to shortcomings in the available technology for assessing infant intelligence, or infant test-taking behaviour, rather than the nature of cognitive development itself. Alternatively, it may be due to the level of maturity of the central nervous system that requires a certain developmental level to be achieved before infancy tests can be reliable measures of an infant's potential capacities. Overall, the data from the present studies seem to be in agreement with Bell and Waldrop's (1989) conclusion that "... if we really thoroughly understand the process of development..., much of what we see as discontinuity would be seen as continuity. However, discontinuities will remain on account of nodal or transition points in development at which reorganization of behaviour occur" (p. 83).

9-8 LIMITATIONS OF THE RESEARCH PROGRAM

There were some limitations to the work presented here that restricted the interpretation of and generalization from the data. One major limitation of the research program was that, because of time constraints, a limited age range was studied here. In fact, the length of time frame on which prediction was based, was a maximum of 18 months for the younger group (i.e., 6½ to 7-month-olds who were followed-up at 18 and 24 months) and 12 months for older infants (i.e., 12 and 16 months old infants who were followed-up at 24 and 27 months respectively). This restricted age range may have contributed to the low predictability of the infancy tests (particularly for the FTII), because it could not assess the

possibility of the predictability of later IQ at older ages, when more symbolic intellectual functioning is expected and, perhaps, prediction is more likely to be detected. Age 3 years and later would have been preferable for testing the predictability of infancy tests. On the other hand, the poor test-retest reliability of the FTII suggests that there are major problems with this procedure.

The sample used in each age group was also small and, in general, the samples were markedly homogenous and probably nonrepresentative of the population, because of a restricted range of mental functioning. The sample was also restricted in terms of parental social class and that may have affected the generality of the results. The selection of stimuli and familiarization time in Study 3 and Study 4 were exploratory in nature, and it is possible that the nature of stimuli used and the level of complexity of each category of stimuli restricted prediction from the information processing task. Similarly, the familiarization time was based on the time recommended for 12 month old babies. While the investigator could be assured that the allotted familiarization time was sufficient for all subjects to process the stimuli, the time may have been more than the necessary time some subjects required, causing a ceiling effect.

9-9 CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

In relation to the nine major aims of the empirical work addressed in the literature review, the project can be summarized as follows, although the results of this research should not be viewed as conclusive.

(i) the examination of the psychometric properties of fixation measures: There is little evidence of internal consistency for the FTII and VRM task and no test-retest

reliability for the FTII, suggesting that not all items measure a single construct "preference for novelty".

(ii) the predictive validation of the FTII on an Australian infant population: The FTII in the earlier months of life was not found to be a valid predictor of later MDI and/or IQ. It had only limited predictability for the 52-week test in relation to the Bayley Scales and some tests of the Stanford-Binet. It is suggested that little reliance should be placed on the FTII, at least for single and double administrations of the task.

(iii) the possibility of extension of the novelty preference procedure to older infants: Information processing tasks have extended to older ages, but their usefulness in prediction of later IQ was dependent on the inclusion of congenitally disabled children.

(iv) the examination of using a time-fixed procedure in measuring novelty scores to enhance the internal consistency of the novelty task: This method of scoring preference for novelty did not enhance the internal consistency of the task, but resulted in a more uniform performance on the novelty task in terms of the percentage of novelty score.

(v) to examine the usefulness of applying an expiring familiarization time instead of accumulating time: This type of familiarization time resulted in chance performance and was not predictive of later outcome. The study suggested that using familiarity preference instead of novelty preference might be a useful procedure for older infants.

(vi) the comparison between the predictive validity of information processing tasks with those of infant developmental tests and socioeconomic factors: Overall, the Bayley Scales predicted later mental outcome at earlier ages more effectively than the FTII. It is concluded that the Bayley Scales has more power in prediction of later outcome in this age range than the FTII and the VRM task used in this project. The results from multiple

regression analyses did not support any significant contribution from the FTII over the MDI scores in predicting later outcome at 24 months of age.

(vii) the age at which the first prediction of later MDI/IQ becomes possible:

Apparently, at the present time with the available tests and technology, there is at least a waiting period of about 9 months of age (in terms of the Bayley Scales) before prediction is possible, although at this age it is still not as reliable as it would be at 12 months.

(viii) the reliability and validity of the recently revised version of the Bayley Scales in Australian infants population: The new version of the Bayley Scales (Bayley, 1993) is a psychometrically reliable test for Australian infants aged 18 to 27 months. However, it seems to favour female rather than male infants, possibly due to early language acquisition in females. The highly significant correlations obtained between the Bayley Scales (Bayley, 1993) and Stanford-Binet test indicate that the MDI is related to general IQ in terms of the Stanford-Binet .

(ix) the possibility of an MDI gain during the last 25 years: Australian infants have gained 18 points on the MDI during the last 25 years and, as a result, the original version of the Bayley Scales is no longer a valid test, at least for infants aged 18 to 27 months.

The results of the studies clearly suggest the need for further research to develop new procedures and perhaps new tests for measuring human infant cognition. Several lines of research are suggested: (i) It would be highly desirable to retest the infants in the present studies on the Stanford-Binet at or after the age of 3 to obtain a more reliable measure of IQ. This would provide more conclusive evidence concerning the predictive validity of FTII. (ii) If speed of information processing is a major component of the individual differences in intelligence at any age, further research is needed to develop a direct measure of speed of processing in infants (e.g., measures of CNS reaction time). (iii) The procedure

recommended for measuring fixation duration (recording the reflection of targets in the infant's pupil) is clearly insufficient and unreliable. Further research should attempt to develop a new procedure which is easy to use and reliably measurable. However, considering that the subjects are very young infants, it is acknowledged that developing such a procedure might be very difficult. (iv) Inattention measures seem equally important and might be developed as measures of an infant's cognitive development. (v) In relation to the FTII, attempts should be directed to item analysis of the FTII to enhance the reliability of the test. This might include changing the scoring system and eliminating the novelty items that are unreliable for each age group. It would be a worthwhile project to assess the FTII in a large random sample of Australian infants, in order to test data from this study.

9-10 A FINAL COMMENT

The empirical evidence from the present research program suggests that measures of visual recognition memory as indexed by the FTII appear to be poor predictors of later mental functioning. Thus, data did not support a stable theoretical link between preference for novelty during infancy and later IQ. Alternatively, it is possible that the Fagan procedure does not reflect a reliable measure of an infant's disposition for preference for novelty. Rather, it may measure speed of information processing, an infant's ability to sustain attention, short-term memory or even some behaviours related to temperament. Whatever the skills, functions or behaviours measured by novelty tasks, it seems that they are not stable during the first two years of development. It is possible that the relationship observed by Fagan and reported in the normative data was influenced by data from five at-risk infants, or possibly by the multiple administration of the task. This is an important point, since the results of the present research clearly indicated that prediction was affected

by outlier subjects. It is safe to suggest that, within the Fagan procedure, relying only on recording fixation measures, and with the administration of a single test or two tests (at two different ages), there is little opportunity for prediction of later IQ, at least for infants not at-risk. The FTII and its procedure clearly need some refinement to make the test more useful. Thus, it remains possible that novelty tasks including the FTII could be developed as a worthwhile measure of an infant's current cognition and later mental outcome (at least for some age groups), as the data do suggest some amount of shared variance (although not significant) between an infant's ability to prefer a novel target and her/his cognitive skills later in life.

The present research substantially agrees with the literature that the predictive value of developmental tests (Bayley Scales) increases as the infant's age at the initial test increases. Almost no prediction is possible from environmental factors such as SES and maternal education or IQ during the first two years of life, probably because the effects of the interaction of the infant's competencies with environmental influences takes time to emerge. The present data suggest that, as some developmentalists have proposed, infant development is characterized by a period of time (perhaps the first 9-12 months of life) during which cognitive development undergoes a developmental change rendering prediction very difficult, if not impossible.

It is acknowledged that developmentalists are still far from having a reliable tool for predicting an infant's cognitive development later in life. Using any existing type of test in the early months to predict later IQ seems unwarranted at the present time, at least for normal infants.

LIST OF APPENDICES

Appendix A	A Summary of Raw Data	248
Appendix B	Background Information Form	260
Appendix C	A Computerised Print out of the Results of the FTII	262
Appendix D	Bar Graphs of Infant's Performance on Novelty Items	264
Appendix E	Oneway ANOVA (Preference for Novelty)	266
Appendix F	Oneway ANOVA (Comparing Means of Number of Fixations)	267
Appendix G	Oneway ANOVA (Comparing Means of Exposure Time)	268
Appendix H	Multiple Regression for Variables Predicting later IQ	269
Appendix I	Oneway ANOVA (Comparing Means of IQ Gains Between Higher Scoring and Normal and Lower Scoring Infants)	272
Publication	Paper presented in 24 th Annual Experimental Psychology	
Publication	Paper published in Australian Psychologist, 32 (2), 120-125	

Appendix A

Summary Data (Infancy Measures)

Age Group 1 (27-week-old)

No.	Code	Sex	FTII 27 weeks	FTII 29 weeks	Ex. T. 27weeks	Ex. T. 29 weeks	NF 27 weeks	NF 29 weeks	MDI 27 weeks	PDI 27 weeks	MDI 29 weeks	PDI 29 weeks
1	6	2	70.0	.	8.91	.	3.40	.	110	116	.	.
2	7	1	60.8	.	10.15	.	2.67	.	116	104	.	.
3	10	1	63.1	68.8	6.06	7.58	1.73	2.06	100	92	102	91
4	21	2	66.6	.	11.25	.	3.13	.	124	124	.	.
5	22	2	65.3	65.8	6.48	7.78	2.90	3.26	96	100	94	95
6	23	1	57.6	.	7.03	.	2.77	.	124	116	.	.
7	24	2	62.0	.	7.47	.	1.93	.	120	129	.	.
8	43	1	79.1	60.3	7.71	7.43	2.40	2.63	139	120	136	106
9	46	2	76.7	69.9	6.77	7.75	2.33	3.02	129	124	136	117
10	50	2	65.1	64.3	6.55	8.67	3.07	3.15	100	112	111	102
11	51	2	62.7	59.6	6.96	7.75	2.23	3.15	100	124	102	110
12	58	2	65.6	.	6.53	.	3.57	.	116	120	.	.
13	62	1	71.0	66.3	8.31	7.81	2.25	3.15	94	104	94	91
14	67	2	58.9	64.6	8.96	6.01	2.93	2.08	116	108	111	106
15	71	1	62.2	69.5	7.66	10.23	2.27	2.78	134	124	123	110
16	72	2	63.7	62.0	6.35	8.04	2.53	3.12	129	127	132	110
17	80	2	53.3	59.6	5.19	6.11	2.73	2.17	91	100	86	91
18	81	2	59.9	57.8	7.56	10.34	3.30	4.24	98	104	99	91

Note: FTII = Novelty Scores from the FTII. Ex. T. = Exposure Time. NF = Number of Fixations. MDI = Mental Development Index. PDI = Psychomotor Development Index.

Appendix A

Summary Data (Childhood Measures)

Age Group 1 (27-week-old)

No.	BSIS (1) 18 months	BSID (2) 18 months	BSID (1) 24 months	BSID (2) 24 months	Vocab 24 months	Comp 24 months	Pat. A. 24 months	Copy 24 months	Memory 24 months	IQ24 24 months
1	107	91	116	104	102	102	106	110	108	107
2	115	93	137	112	98	102	106	96	96	99
3	.	.	102	92	94	84	94	.	.	90
4	107	85	93	80	.	.	106	96	.	101
5	107	89
6	124	113	150	114	110	110	112	.	108	112
7	117	101	123	98	84	102	100	96	.	94
8	111	97	137	112	106	102	118	96	100	105
9
10	105	91	123	100	98	110	94	96	100	99
11	119	99	127	100	94	102	106	96	.	99
12	119	89	89	70
13	113	95	119	100	98	96	106	96	100	99
14	111	93	137	110	94	110	106	96	.	102
15	111	95	100	86	84	102	94	96	.	92
16	134	111	132	112	102	96	112	96	104	103
17	81	77	76	68
18	.	.	100	98	.	.	94	96	.	94

Note: BSID1 & 2 = Bayley Scales Of Infant Development (1969, 1993).

Vocab = Vocabulary. Comp = Comprehension. Pat .A. = Pattern Analysis. Copy = Copying. Memory = Memory For Sentences.

IQ = Partial Composite SAS.

Appendix A

Summary Data (Infancy Measures)

Age Group 2 (29-week-old)

No.	Code	Sex	FTII 29 weeks	FTII 39 weeks	Ex. T. 29weeks	Ex. T. 39 weeks	NF 29 weeks	NF 39 weeks	MDI 29 weeks	PDI 39 weeks	MDI 29 weeks	PDI 29 weeks
1	13	2	61.6	62.7	12.81	6.46	2.72	3.13	102	106	126	119
2	16	1	63.4	68.5	6.62	6.10	2.40	5.03	119	113	120	119
3	29	2	64.7	56.6	4.97	5.34	1.98	4.43	99	106	114	92
4	30	2	70.5	.	5.14	.	2.08	.	123	102	.	.
5	31	1	62.1	76.8	7.66	6.32	3.42	5.30	128	106	126	92
6	32	2	69.9	63.9	6.83	7.56	2.63	3.48	111	102	103	90
7	33	2	61.3	50.4	8.73	5.20	2.53	3.60	128	113	114	119
8	35	2	62.3	65.5	9.65	8.48	2.93	3.60	107	110	114	105
9	37	2	69.8	63.8	6.21	5.83	2.78	3.15	111	110	114	105
10	38	1	52.0	65.1	7.39	8.94	2.18	3.82	86	95	111	100
11	41	2	59.4	69.4	6.10	8.12	2.87	5.30	115	117	123	105
12	47	1	66.5	64.2	5.47	6.35	2.73	4.15	119	110	117	96
13	53	1	58.7	.	6.86	.	2.62	.	94	87	.	.
14	57	1	60.4	61.6	6.61	7.54	3.00	3.47	123	110	133	119
15	82	2	57.8	.	10.34	.	4.24	.	99	91	.	.

Note: FTII = Novelty Scores from the FTII. Ex. T. = Exposure Time. NF = Number of Fixations. MDI = Mental Development Index. PDI = Psychomotor Development Index.

Appendix A

Summary Data (Childhood Measures)

Age Group 2 (29-week-old)

No.	BSIS (1) 18 months	BSID (2) 18 months	BSID (1) 24 months	BSID (2) 24 months	Vocab 24 months	Comp 24 months	Pat. A. 24 months	Copy 24 months	Memory 24 months	IQ24 24 months
1	109	95
2	113	89	116	102	.	96	118	96	.	102
3	117	97	143	112	98	110	112	96	.	105
4
5	146	119	151	134	118	128	100	132	112	121
6	.	.	114	98
7	117	97	109	86	.	96	106	96	.	98
8	121	109	150	112	106	110	106	96	.	106
9	121	99	127	104	98	.	106	.	.	99
10	130	111	150	114	110	110	112	96	118	113
11	132	109	137	108	110	110	106	96	96	103
12	113	91	106	88	94	102	88	96	.	94
13	151	121	151	132	130	114	118	96	112	117
14	128	107	132	102	98	.	118	.	.	118
15

Note: BSID1 & 2 = Bayley Scales Of Infant Development (1969, 1993).

Vocab = Vocabulary. Comp = Comprehension. Pat. A. = Pattern Analysis. Copy = Copying. Memory = Memory For Sentences.
IQ = Partial Composite SAS.

Appendix A

Summary Data (Infancy Measures)

Age Group 3 (39-week-old)

No.	Code	Sex	FTII 39 Weeks	FTII 52 Weeks	Ex. T. 39 Weeks	Ex. T. 52 Weeks	NF 39 Weeks	NF 52 Weeks	MDI 39 Weeks	PDI 39 Weeks	MDI 52 Weeks	PDI 52 Weeks
1	3	1	66.9	61.8	14.90	10.92	4.08	3.47	117	119	126	111
2	4	2	62.0	62.6	9.31	9.59	2.85	4.53	111	100	122	122
3	8	1	66.2	70.3	6.02	8.70	3.32	4.13	129	139	126	115
4	14	2	59.0	65.2	5.31	6.13	2.27	3.70	114	105	112	122
5	15	2	56.6	58.4	7.95	7.93	2.32	3.80	117	105	106	105
6	17	1	60.2	78.3	5.71	5.31	2.42	4.60	126	111	122	111
7	18	1	62.8	69.4	8.52	6.01	2.48	2.83	111	111	119	98
8	20	1	51.5	.	6.40	.	2.65	.	103	92	.	.
9	28	1	72.0	56.1	5.97	7.82	2.38	4.57	114	119	112	111
10	34	1	52.7	.	8.38	.	2.75	.	89	76	.	.
11	40	2	56.4	64.0	6.38	9.19	3.07	4.27	140	134	140	128
12	42	2	64.5	67.5	6.99	6.09	3.13	4.23	103	92	91	92
13	45	1	59.4	57.2	5.36	5.85	4.43	3.87	107	111	112	105
14	63	2	58.5	.	7.37	.	3.18	.	118	111	.	.
15	64	1	66.9	56.1	7.66	5.00	4.82	3.97	99	100	98	105
16	65	1	77.8	63.1	13.85	9.62	3.82	3.62	99	72	89	66
17	68	1	59.5	.	6.93	.	2.75	.	120	111	.	.
18	69	2	64.9	.	6.10	.	4.20	.	99	90	.	.
19	73	2	56.4	.	4.82	.	2.97	.	99	90	.	.
20	75	2	68.6	.	6.06	.	4.14	.	107	111	.	.
21	76	2	65.7	.	10.73	.	4.10	.	126	119	.	.
22	78	1	63.6	.	6.32	.	3.82	.	129	105	.	.

Note: FTII = Novelty Scores from the FTII. Ex. T. = Exposure Time. NF = Number of Fixations. MDI = Mental Development Index
PDI = Psychomotor Development Index.

Appendix A

Summary Data (Childhood Measures)

Age Group 3 (39-week-old)

No.	BSID (1) 21months	BSID (2) 21months	BSID (1) 24 months	BISD (2) 24 months	Vocab 24 months	Comp 24 months	Pat. A. 24 months	Copy 24 months	Memory 24 months	IQ24 24 months
1	106	86	106	94	94	96	106	96	.	98
2	111	98
3	150	120	150	122	122	132	112	110	120	125
4	104	90
5	113	98	114	98	84	.	106	96	96	93
6	146	114	137	114	102	110	106	96	100	103
7	106	88	104	84	98	90	100	96	.	95
8	122	112	127	110	102	110	106	96	96	102
9	122	112	150	118	110	110	106	110	114	113
10	93	80	96	78
11	122	106	150	114	106	102	106	96	.	103
12	.	.	81	68
13	120	102	132	112	102	102	94	96	100	98
14	109	92	112	104	102	96	112	96	.	102
15	108	98	119	108	94	.	100	96	100	97
16	.	.	106	96	98	102	94	96	.	97
17	134	110	151	124	118	110	106	110	108	112
18	111	102	119	94	94	.	94	96	.	93
19	95	86	98	86
20	127	110	132	114	118	114	106	96	108	111
21	108	90	100	92
22	127	104

Note: BSID1 & 2 = Bayley Scales Of Infant Development (1969, 1993).

Vocab = Vocabulary. Comp = Comprehension. Pat. A. = Pattern Analysis. Copy = Copying. Memory = Memory For Sentences.

IQ = (Partial Composite SAS).

Appendix A

Summary Data (Infancy Measures)

Age Group 4 (52-week-old)

No.	Code	Sex	FTII 52 Weeks	Ex. T. 52 Weeks	NF 52 Weeks	MDI 52 Weeks	PDI 52 Weeks
1	1	1	63.0	10.82	4.50	115	105
2	2	2	65.0	8.83	2.40	112	111
3	5	1	54.6	10.62	3.92	109	86
4	9	2	63.6	7.60	3.55	93	92
5	11	2	60.0	8.78	2.80	106	111
6	12	1	60.9	10.42	3.55	112	105
7	25	1	68.8	4.54	2.92	119	117
8	26	2	73.6	11.91	4.80	119	111
9	27	2	49.6	8.79	3.33	112	98
10	39	1	60.2	7.16	3.37	112	105
11	44	1	79.6	8.19	4.27	117	117
12	48	1	60.8	6.02	3.87	112	111
13	49	2	65.0	6.96	3.62	122	105
14	54	2	64.1	5.07	2.63	126	92
15	55	2	64.7	5.88	4.50	109	111
16	56	1	62.6	9.69	4.17	131	122
17	59	1	73.4	5.40	4.38	128	92
18	60	2	66.2	6.96	3.37	122	117
19	66	2	72.9	5.77	4.10	115	117
20	70	1	69.5	9.11	4.62	126	117
21	74	1	64.2	5.59	3.03	122	98
22	77	1	61.1	9.24	5.55	109	105
23	79	2	61.2	7.48	3.78	115	98

Note: FTII = Novelty Scores from the FTII. Ex. T = Exposure Time. NF = Number of Fixations. MDI = Mental Development Index. PDI = Psychomotor Development Index.

Appendix A

Summary Data (Childhood Measures)

Age Group 4 (52-week-old)

No.	BSID (1) 24 months	BISD (2) 24 months	Vocab 24 months	Comp 24 months	Pat. A. 24 months	Copy 24 months	Memory 24 months	IQ24 24 months
1	143	122	102	110	106	110	104	108
2	91	82	84	.	100	110	.	94
3	114	96	90	96	100	96	.	94
4
5	106	100	98	102	88	96	100	96
6	123	104	98	114	100	.	.	104
7	132	114	110	102	106	110	.	109
8	137	116	102	96	100	110	100	102
9	96	84	84	96	88	96	.	89
10	132	106	.	110	.	96	.	103
11	137	110	102	102	106	96	96	100
12	106	88	84	102	94	96	.	92
13	150	116	126	102	112	96	112	112
14	127	108	102	110	106	110	96	105
15	116	98	98	96	100	96	.	97
16	150	122	110	110	100	96	118	111
17	132	114	110	90	100	96	.	99
18	116	98	98	96	94	96	.	95
19	123	110	114	96	100	96	100	101
20
21	132	110	106	90	106	96	118	107
22	104	96	90	90	94	96	.	90
23	132	110	114	110	94	96	108	106

Note: BSID1 & 2 = Bayley Scales Of Infant Development (1969, 1993).

Vocab = Vocabulary. Comp = Comprehension. Pat. A. = Pattern Analysis. Copy = Copying. Memory = Memory For Sentences.

IQ = (Partial Composite SAS).

Appendix A

Summary Data (Infancy Measures)

Age Group 5-1 (69-week-old)

No.	Code	Sex	VRM 69 Weeks	NF 69 Weeks	Ex. T. 69 Weeks	MDI 69 Weeks	PDI 69 Weeks
1	103	1	65.70	5.40	11.78	109	102
2	104	1	69.80	6.25	12.85	107	110
3	106	1	65.40	5.60	11.22	130	115
4	107	1	59.90	4.60	10.76	112	108
5	109	1	40.50	4.40	13.24	65	63
6	110	1	40.40	9.20	35.26	65	50
7	111	1	74.20	5.40	10.60	130	114
8	112	1	72.20	5.80	14.22	126	124
9	117	1	74.50	6.40	11.16	121	114
10	101	2	58.30	10.00	15.10	92	102
11	102	2	67.80	5.00	11.64	109	116
12	105	2	72.00	4.20	11.94	100	117
13	108	2	68.30	6.40	16.60	118	100
14	113	2	66.70	5.60	12.64	112	102
15	114	2	59.00	9.00	11.30	105	106
16	115	2	65.30	6.40	11.10	130	114

Note: FTII = Novelty Scores from the FTII. Ex. T. = Exposure Time. NF = Number of Fixations.
MDI = Mental Development Index. PDI = Psychomotor Development Index.

Appendix A

Summary Data (Childhood Measures)

Age Group 5-1 (69-week-old)

No.	BSID (1) 27 months	BSID (2) 27 months	Vocab 27 months	Comp 27 months	Pat. A. 27 months	Copy 27 months	Memory 27 months	IQ27 27 months
1	114	97	98	84	94	110	96.000	95
2
3	139	117	122	90	100	110	114.000	111
4	133	110	114	114	100	110	124.000	118
5	64	49
6	63	49
7	139	121	122	110	118	124	128.000	127
8	126	110	114	110	100	110	108.000	111
9	145	125	114	110	126	124	118.000	123
10	104	91	102	90	94	96	100.000	96
11	107	97	94	90	88	96	108.000	96
12	107	104	98	90	106	110	104.000	102
13	110	99	106	110	106	96	108.000	107
14	139	115	118	110	112	124	120.000	123
15	120	99	106	96	94	96	108.000	101
16	145	125	126	120	118	124	124.000	128

Note: BSID1 & 2 = Bayley Scales Of Infant Development (1969, 1993).

Vocab = Vocabulary. Comp = Comprehension. Pat .A. = Pattern Analysis. Copy = Copying. Memory = Memory For Sentences.
IQ = Partial Composite SAS.

Appendix A

Summary Data (Infancy Measures)

Age group 5-2 (69-week-old)

No.	Code	Sex	VRM	NF 69 Weeks	MDI 69 Weeks	PDI 69 Weeks
1	201	2	49.7	3.80	98	81
2	202	1	75.5	4.40	100	102
3	203	2	48.3	3.80	112	122
4	204	2	51.5	3.80	118	97
5	205	2	51.9	4.60	98	93
6	206	2	50.3	5.40	94	102
7	207	2	37.5	6.80	104	100
8	208	1	33.0	4.80	107	98
9	209	1	52.7	5.00	119	117
10	210	1	58.3	5.80	119	84
11	211	2	76.5	5.40	95	97
12	212	1	60.4	6.80	124	118
13	213	1	54.4	6.80	104	90
14	214	2	64.5	5.60	95	93
15	215	2	66.2	4.20	119	106
16	216	1	62.4	5.40	117	110

Note: VRM. Visual Recognition Memory. NF = Number of Fixations.
MDI = Mental Development Index. PDI = Psychomotor Development Index.

Appendix A

Summary Data (Childhood Measures)

Age group 5-2 (69-week-old)

No.	BSID (1) 27 months	BSID (2) 27 months	Vocab 27 months	Comp 27 months	Pat. A. 27 months	Copy 27 months	Memory 27 months	IQ27 27 months
1	120	102	110	102	112	124	104	112
2	120	102	98	132	106	96	96	105
3	133	106	110	114	106	96	114	111
4	126	117	106	110	118	110	104	111
5	133	97	118	102	106	96	.	107
6	114	88	94	96	100	96	.	96
7	110	99	110	102	112	96	104	106
8	133	104	106	102	106	96	96	100
9	133	117	114	114	106	96	128	119
10	139	110	106	102	106	110	100	105
11	98	86	84	96	114	96	100	93
12	126	115	122	114	112	110	100	113
13	107	91	98	102	106	96	.	100
14	104	88	.	.	106	96	.	101
15	110	99	126	110	106	124	120	122
16	126	99	98	110	106	96	96	100

Note: BSID1 & 2 = Bayley Scales Of Infant Development (1969, 1993).

Vocab = Vocabulary. Comp = Comprehension. Pat .A. = Pattern Analysis. Copy = Copying. Memory = Memory For Sentences.
IQ = Partial Composite SAS.

APPENDIX B

Background Information

Dear (parent)

I would like to request your participation in a project which is currently being conducted in the Department of Psychology, University of Adelaide. The purpose of this project will be to monitor the development of some aspects of infant intelligence, from age 6 months onwards. What is proposed has been considered and approved by the Ethics Committee at both the University and at CAHFS. At this stage we are not certain how long it will be useful to continue monitoring development. However, the intention is to test each child initially on one or two occasions between 6 and 12 months of age. A follow-up test will be planned for when your child is about 2 years of age.

The tests involved at this time will be the Bayley Scales of Infant Intelligence and the Fagan Test of Infant Intelligence. The Fagan test is a newly available procedure and we have only just purchased it from its authors in the U.S. The Fagan test is based on the observation that young infants spontaneously direct attention to novel aspects in their field of vision. It measures "preference for novelty". Essentially, it consists of a portable viewing stage, fourteen face pictures and a computerised scoring system. The potential advantage of this test over the earlier sensori-motor kinds rests on its theoretical foundation, according to which a preference for novelty reflects cognitive functioning, present in the infant, and upon which longer-term cognitive development will build. Thus, preference for novelty during infancy should predict subsequent intellectual development during later childhood and beyond. It is a very simple assessment and is carried out in a relaxed manner, with the infant held by the parent on her (his) lap. The infant is first shown a picture, for example, a woman's face. A new picture is then introduced and the length of time for which

the infant looks at the novel picture is recorded by the tester, who sits behind a small screen, observing the infant's visual fixation. The testing involved in this project can be done anywhere, at the convenience of the family involved - within the home, at the university, at a clinic - because materials are readily transportable. The test usually takes 15-20 minutes.

The Bayley Scales provide a sensori-motor test, designed to evaluate a child's development status in the first 30 months of age. It has three parts by which mental abilities, psychomotor activities and the nature of the child's social and objective orientation toward the environment are assessed. Only items consistent with the specified age involved will be administered and the average total time required is about 45 minutes.

The work is being carried out by Reza Tasbihsazan, a psychologist and university lecturer from Iran who is completing a Ph.D. in this department, under supervision by Dr Ted Nettelbeck and Dr Neil Kirby. While the purpose behind this proposal is concerned essentially with the validity of the new Fagan test in predicting future cognitive development, we expect that the programme will be of value in a broader sense. For example, it may well be useful in a case where an infant is at risk for future cognitive development, because the results could have implications for future training programmes with that child.

All tests results will be kept in confidence but information about your child will be available to you on your request. At the conclusion of the project you will be provided with a general summary of the results of testing. Of course, participation is entirely voluntary and you are free to withdraw from the project at any time.

Appendix C

INFANTEST CORPORATION'S
FAGAN TEST OF INFANT INTELLIGENCE
(C) JOSEPH F FAGAN 1986/1987/1988.

BABY'S NAME:
HOSPITAL NUMBER: Ex1_001
TESTER'S NAME: Reza

SEX: F
BIRTH GESTATIONAL AGE: 38 WEEKS
BIRTH DATE: 8/5/92
TEST DATE: Wed Aug 11 11:17:04 1993
FAGAN TEST ADMINISTERED: 92 WEEK
MOTHER EDUCATION: 17

RACE: W
BIRTH WEIGHT: 3680 GRAMS
BIRTH ORDER: 1
POSTNATAL AGE AT TEST: 53 WEEKS
FATHER EDUCATION: 16

REASON FOR REFEFFAL:
Experiment No.1

T E S T R E S U L T S

1st	TEST ITEM:	23.9%
2nd	TEST ITEM:	100.0%
3rd	TEST ITEM:	44.4%
4th	TEST ITEM:	54.8%
5th	TEST ITEM:	68.0%
6th	TEST ITEM:	69.4%
7th	TEST ITEM:	92.9%
8th	TEST ITEM:	44.7%
9th	TEST ITEM:	31.6%
10th	TEST ITEM:	100.0%

MEAN: 63.0%
ST DEV: 27.8%

STATUS: LOW RISK - RETEST OPTIONAL, BUT ADVISED.

	Counts		Times		
	L	R	L	R	All
FAMILIARIZATION	4	4	4.8	5.2	17.1
NOVELTY TRIAL # 1	1	2	0.3	1.8	3.1
NOVELTY TRIAL # 2	1	1	0.9	0.8	3.1
FAMILIARIZATION	3	5	6.2	3.8	23.8
NOVELTY TRIAL # 1	0	2	0.0	2.6	3.1
NOVELTY TRIAL # 2	2	0	2.5	0.0	3.1
FAMILIARIZATION	2	3	3.5	6.5	17.3
NOVELTY TRIAL # 1	1	1	1.4	1.4	3.1
NOVELTY TRIAL # 2	1	1	1.6	1.0	3.1
NOVELTY TRIAL # 1	0	1	0.0	3.1	3.1
NOVELTY TRIAL # 2	0	2	0.0	2.0	3.1
NOVELTY TRIAL # 1	0	2	0.0	2.5	3.1
NOVELTY TRIAL # 2	1	2	0.9	1.6	3.1
NOVELTY TRIAL # 1	1	2	1.3	1.5	3.1
NOVELTY TRIAL # 2	1	2	1.2	0.4	3.1
FAMILIARIZATION # 1	8	0	6.0	0.0	51.3
FAMILIARIZATION # 2	0	3	0.0	6.0	11.6
NOVELTY TRIAL # 1	1	1	0.1	2.8	3.1
NOVELTY TRIAL # 2	1	1	2.4	0.3	3.1
FAMILIARIZATION # 1	1	3	0.2	1.8	9.3
FAMILIARIZATION # 2	4	0	2.0	0.0	10.7
NOVELTY TRIAL # 1	1	1	1.0	1.7	3.1
NOVELTY TRIAL # 2	0	2	0.0	1.1	3.1
NOVELTY TRIAL # 1	1	1	2.3	0.7	3.1
NOVELTY TRIAL # 2	1	1	0.5	0.3	3.1
FAMILIARIZATION # 1	0	4	0.0	2.0	13.8
FAMILIARIZATION # 2	4	0	2.0	0.0	10.7
NOVELTY TRIAL # 1	0	1	0.0	2.5	3.1
NOVELTY TRIAL # 2	1	0	1.1	0.0	3.1

Appendix D

Figure D-1

The Percentage of Preference for Novelty (27-week-olds)

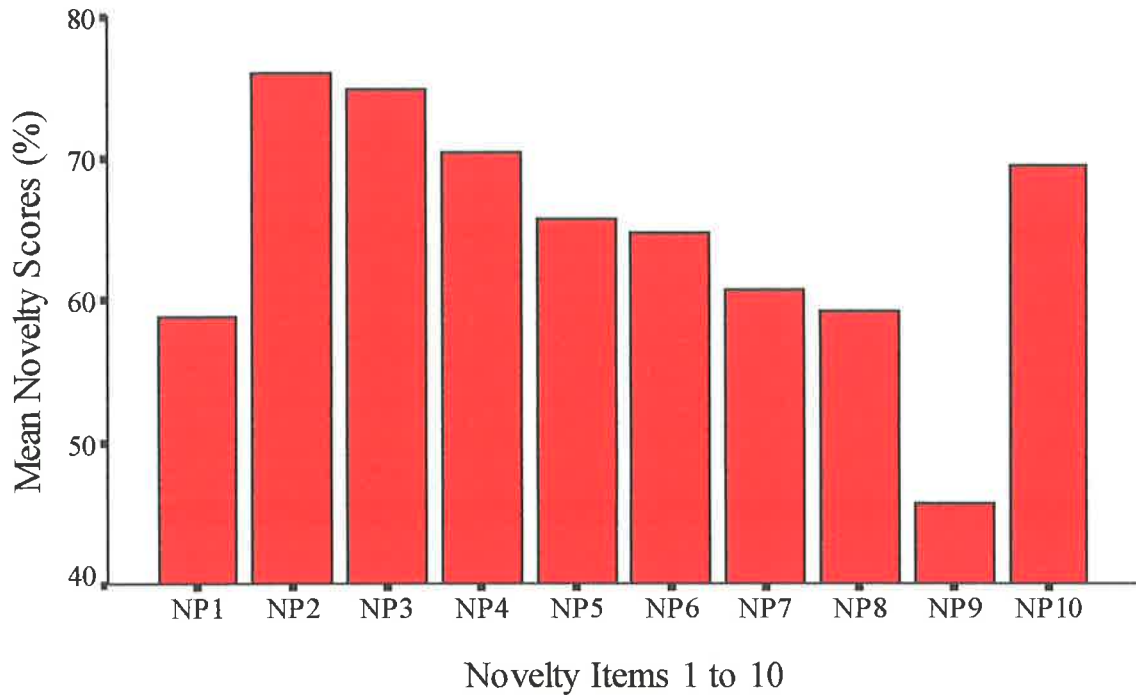


Figure D-2

The Percentage of Preference for Novelty (29-week-olds)

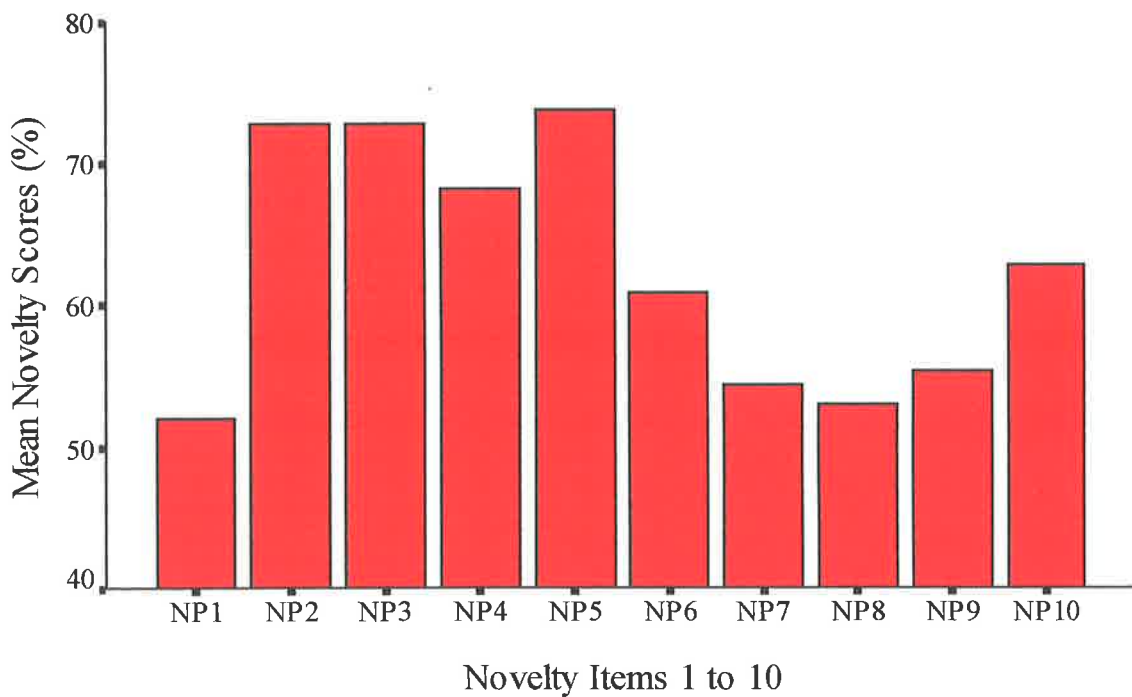


Figure D-3

The Percentage of Preference for Novelty (39-week-olds)

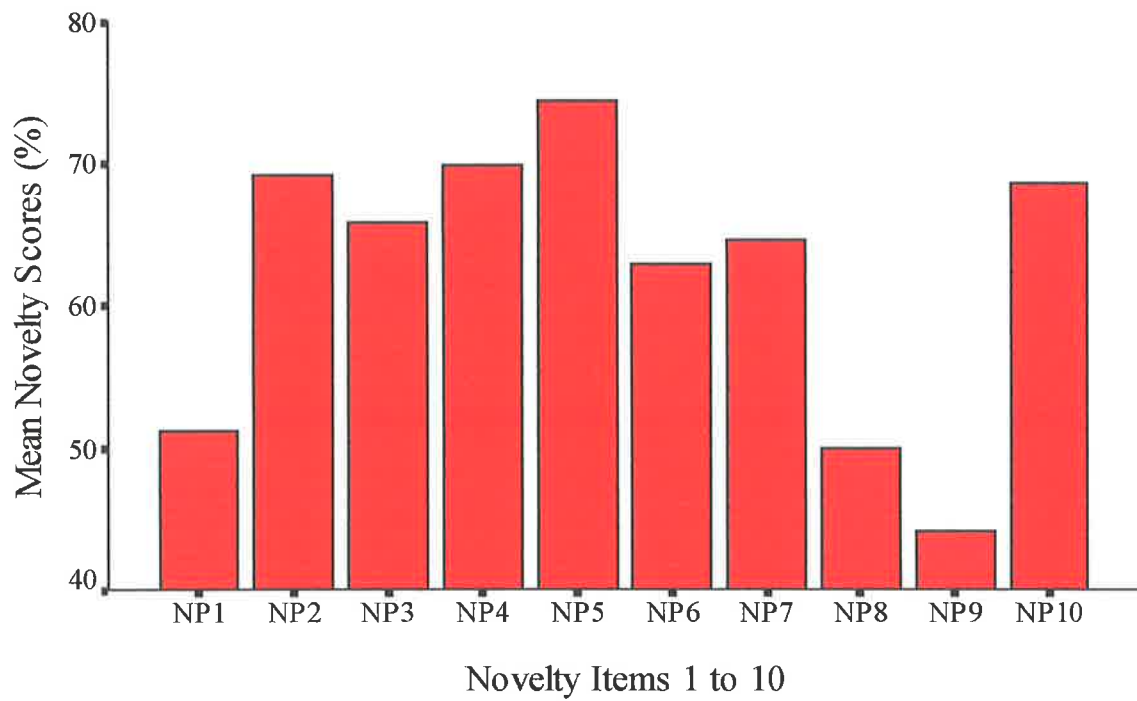
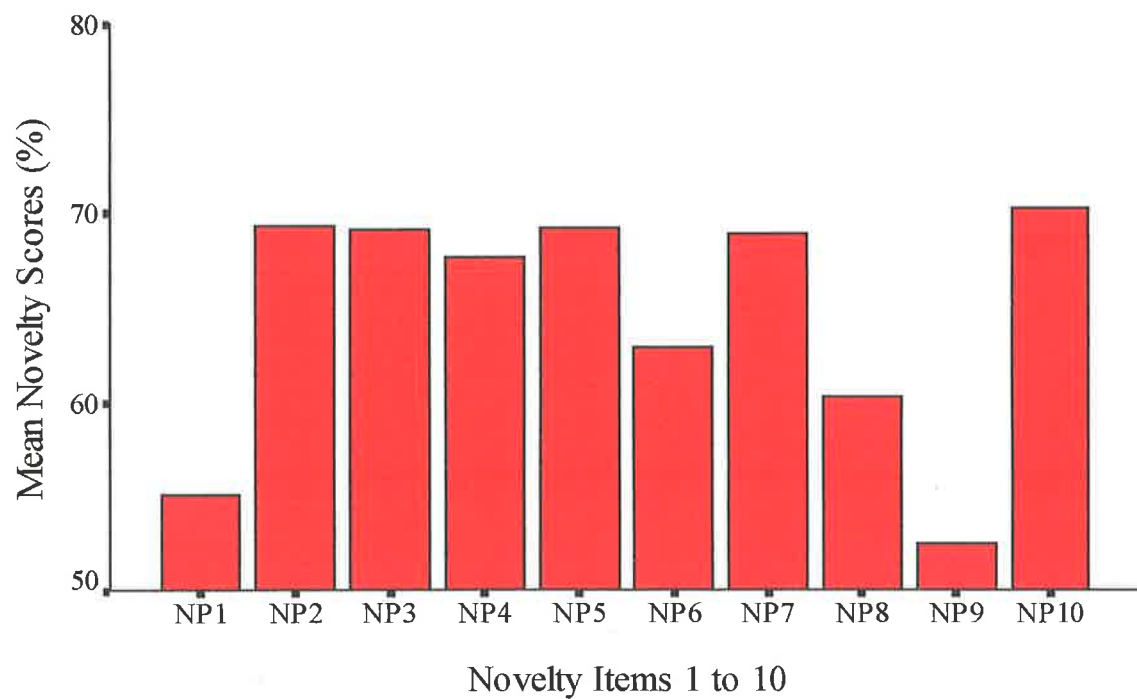


Figure D-4

The Percentage of Preference for Novelty (52-week-olds)



Appendix E

Oneway Analysis of Variance

Comparing Differences between Means of Novelty Preference Scores for Four Age Groups

Source	Degree of Freedom	Sum of Squares	Mean Squares	F Ratio	F Probability
Between Groups	3	86.01	28.68	.76	.52
Within Groups	74	2789.56	37.70		
Total	77	2875.56			

Appendix F

Oneway Analysis of Variance

Comparing Differences between Means of Number of Fixations for Four
Age Groups

Source	Degree of Freedom	Sum of Squares	Mean Squares	F Ratio	F Probability
Between Groups	3	15.98	5.33	11.47	.001
Within Groups	74	34.39	.465		
Total	77	50.36			

Post Hoc Comparisons for Number of Fixations (Bonferroni test)

Age Groups	Mean Number of Fixations	Significance
Group 1 (27 weeks)	2.6744	
Group 2 (29 weeks)	2.7407	
Group 3 (39 weeks)	3.2705	*
Group 4 (52 weeks)	3.7809	**

Note: The mean difference is significant at the .05 level.

Appendix G

Oneway Analysis of Variance

Comparing Differences between Means of Exposure Time for Four Age Groups

Source	Degree of Freedom	Sum of Squares	Mean Squares	F Ratio	F Probability
Between Groups	3	4.41	1.47	.29	.83
Within Groups	74	378.14	5.11		
Total	77	382.55			

Appendix H

Multiple Regression Analyses for Variable Predicting Later MDI/IQ

(1) Dependent Variable: Bayley Scales (MDI) at 24 months

Variable Entered on Step Number 1: Bayley Scales (MDI) at 52 weeks

Analysis of Variance

Source	Degree of Freedom	Sum of Squares	Mean Squares	F	F Significance
Regression	1	986.87974	986.87974	11.36357	.0032
Residual	19	1650.07264	86.84593		

Variable Entered on Step Number 2: FTII at 52 weeks

Analysis of Variance

Source	Degree of Freedom	Sum of Squares	Mean Squares	F	F Significance
Regression	2	1166.95891	583.47946	7.14468	.0052
Residual	18	1469.99347	81.66630		

Variables in the Equation

Variable	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>Sig T</i>
<u>Block 1</u>				
Bayley Scales (MDI)	1.034052	.306750	.611760	.0032
<u>Block 2</u>				
Bayley Scales (MDI)	.831382	.327279	.491857	.0205
Novelty Scores (FTII)	.493232	.332155	.287519	.1549

(2) Dependent Variable: Vocabulary (Stanford-Binet) at 24 months

Variable Entered on Step Number 1: Bayley Scales (MDI) at 52 weeks

Analysis of Variance

Source	Degree of Freedom	Sum of Squares	Mean Squares	F	F Significance
Regression	1	45.55587	45.55587	9.01163	.0077
Residual	18	90.99413	5.05523		

Variable Entered on Step Number 2: FTII at 52 weeks

Analysis of Variance

Source	Degree of Freedom	Sum of Squares	Mean Squares	F	F Significance
Regression	2	53.92693	26.96346	5.54783	.0140
Residual	17	82.62307	4.86018		

Variables in the Equation

Variable	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>Sig T</i>
<u>Block 1</u>				
Bayley Scales (MDI)	.224634	.074830	.577599	.0077
<u>Block 2</u>				
Bayley Scales (MDI)	.182043	.080229	.468084	.0366
Novelty Scores (FTII)	.106774	.081358	.270735	.2068

(3) Dependent Variable: Pattern Analysis (Stanford-Binet) at 24 months

Variable Entered on Step Number 1: Bayley Scales (MDI) at 52 weeks

Analysis of Variance

Source	Degree of Freedom	Sum of Squares	Mean Squares	F	F Significance
Regression	1	4.76668	4.76668	5.30177	.0335
Residual	18	16.18332	.89907		

Variable Entered on Step Number 2: FTII at 52 weeks

Analysis of Variance

Source	Degree of Freedom	Sum of Squares	Mean Squares	F	F Significance
Regression	2	6.90214	3.45107	4.17631	.0335
Residual	17	14.04786	.82634		

Variables in the Equation

Variable	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>Sig T</i>
<u>Block 1</u>				
Bayley Scales (MDI)	.072663	.031557	.476997	.0335
<u>Block 2</u>				
Bayley Scales (MDI)	.051151	.033081	.335782	.1405
Novelty Scores (FTII)	.053929	.033547	.349103	.1263

Appendix I

Oneway Analysis of Variance

Compares mean of IQ gains between higher scoring infants and infants from the normal IQ range

Source	Degree of Freedom	Sum of Squares	Mean Squares	F Ratio	F Probability
Between Groups	3	1738.8230	579.6077	21.9665	.0000
Within Groups	93	2453.8987	26.3860		
Total	96	4192.7216			

Post Hoc (Tukey Test)

Bayley Scales (1969)	IQ Gains	Significance
Group 1 (MDI, 84-100)	11.4286	
Group 2 (MDI, 101-116)	15.5789	
Group 3 (MDI, 117-132)	18.1333	*
Group 4 (MDI, 133-150)	25.4545	***

The Fagan test of infant intelligence: Predictive validation in an Australian sample of two-year-olds

R. TASBIHSAZAN, T. Nettelbeck, and N. Kirby

Psychology, Adelaide University

psyr-tas@arts.adelaide.edu.au

A new generation of infant tests has made questionable the well-established view of the discontinuity of intelligence during infancy introduced by Bayley nearly 50 years ago. During the 1970s and 1980s Fagan and his colleagues conducted several studies in which they investigated infants' recognition memory. On the basis of these studies the current version of the Fagan Test of Infant Intelligence was developed and it is now commercially available. The predictive validity of the Fagan test was examined on 78 Australian infants at four different age groups (ie, 27, 29, 39 and 52 weeks of age). All subjects were followed up one year later and again when they reached 24 months of age. No evidence of predictive validity to later mental outcome was found for infants who were initially tested at two younger ages. The Fagan test-score at 52 weeks correlated significantly with MDI of Bayley scales and some factors of Stanford-Binet at 2 years of age. The FTII administered at 9 months was only related to Memory for Sentences, a factor of Stanford-Binet test at 2 years. The inconsistency of results and low predictive validity of the Fagan test and its utility as a screening device in nonrisk infants is discussed.

Tasbihsazan, R., Nettelbeck, T., and Kirby, N., (1997) Increasing mental development index in Australian children: a comparative study of two versions of the Bayley Mental Scale.
Australian Psychologist, v. 32 (2), pp. 120-125.

NOTE:

This publication is included in the print copy of the thesis held
in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1080/00050069708257364>

REFERENCES

- Anderson, M. (1986). Inspection time and IQ in young children. *Personality and Individual Differences*, 7 (5), 677-686.
- Anderson, M. (1992). *Intelligence and development: A cognitive theory*. Oxford, UK: Blackwell.
- Andersson, H.W. (1996). The Fagan test of infant intelligence: Predictive validity in a random sample. *Psychological Reports*, 78, 1015-1026.
- Bahrack, L.E., & Pickens, J.N. (1995). Infant memory for object motion across a period of three months: Implications for a four-phase attention function. *Journal of Experimental Child Psychology*, 59, 343-371.
- Baker, L.A., DeFries, J.C., & Fulker, D.W. (1983). Longitudinal stability of cognitive ability in the Colorado Adoption Project. *Child Development*, 54, 290-297.
- Barrera, M.E., & Maurer, D. (1981). Recognition of mother's photographed face by the three-month-old infant. *Child Development*, 52, 714-716.
- Bayley, N. (1933). *The California First-year Mental Scale*. Berkeley: University of California Press.
- Bayley, N. (1936). *The California Infant Scale of Motor Development*. Berkeley: University of California Press.
- Bayley, N. (1955). On the growth of intelligence. *American Psychologist*, 10, 805-818.
- Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: The Psychological Corporation.
- Bayley, N. (1970). Development of mental abilities. In P.H. Mussen (Ed.), *Carmichael's manual of child psychology* (3rd ed.). Vol. 1. (pp. 1163-1209). New York: John Wiley & Sons.
- Bayley, N. (1993). *Bayley Scales of Infant Development (Second Edition)*. New York: The Psychological Corporation.
- Beckwith, L., Cohen, S.E., Kopp, C.B., Parmelee, A.H., & Marcy, T.G. (1976). Caregiver-infant interaction and early cognitive development in preterm infants. *Child Development*, 47, 579-587.
- Beckwith, L., & Parmelee, A.H. (1986). EEG patterns of preterm infants, home environment, and later IQ. *Child Development*, 57, 777-789.

- Bee, H.L., Barnard, K.F., Eyres, S.J., Gray, C.A., Hammond, M.A., Spietz, A.L., Snyder, C., & Clark, B. (1982). Prediction of IQ and language skill from perinatal status, child performance, family characteristics, and mother-infant interaction. *Child Development*, 53, 1134-1156.
- Bell, R.Q., & Waldrop, M.F. (1989). Achievement and cognitive correlates of minor physical anomalies in early development. In M.H. Bornstein & N.A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives*. (pp. 63-85). New Jersey: Lawrence Erlbaum Associates.
- Benasich, A.A., & Bejar, I.I. (1992). The Fagan test of infant intelligence: A critical review. *Journal of Applied Developmental Psychology*, 13, 153-171.
- Benson, J.B., Cherny, S.S., Haith, M.M., & Fulker, D. W. (1993). Rapid assessment of infant predictors of adult IQ: Midtwin-Midparent analyses. *Developmental Psychology*, 29 (3), 434-447.
- Berg, C.A., & Sternberg, R.J. (1985). Response to novelty: Continuity versus discontinuity in the developmental course of intelligence. In H. W. Reese (Series Ed.), *Advances in child development and behaviour*: Vol. 19. (pp. 2-47). New York: Academic Press.
- Bloch, H. (1988). Interpreting the familiarization process in habituation. *European Bulletin of Cognitive Psychology*, 8 (5), 441-444.
- Bornstein, M.H. (1988). Answers to three prominent questions about habituation. *European Bulletin of Cognitive Psychology*, 8 (5), 531-538.
- Bornstein, M.H. (1989a). Information processing (habituation) in infancy and stability in cognitive development. *Human Development*, 32, 129-136.
- Bornstein, M.H. (1989b). Stability in early mental development: From attention and information processing in infancy to language and cognition in childhood. In M.H. Bornstein & N.A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives*. (pp. 147-170). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bornstein, M.H. (1992). Perception and cognitive development. In M.H., Bornstein & Lamb, M.E. (Eds.), *Development Psychology: An advanced textbook* (pp. 155-209). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bornstein, M.H., & Benasich, A.A., (1986). Infant habituation: Assessment of individual differences and short-term reliability at 5 months. *Child Development* 57, 87-99.
- Bornstein, M.H, Pecheux, M. & Lecuyer, R. (1988). Visual habituation in human infants: Development and rearing circumstances. *Psychological Research*, 50, 130-133.
- Bornstein, M.H., & Sigman, M.D. (1986). Continuity in mental development from infancy. *Child Development*, 57, 251-274.

- Bradley, R.H. (1989). The use of the home inventory in longitudinal studies of child development. In M.H. Bornstein & N.A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives*. (pp. 191-215). New Jersey: Lawrence Erlbaum Associates.
- Bradley, R.H., Caldwell, B.M., Rock, S.L., Barnard, K.E., Gray, C., Hammond, M.A., Mitchell, S., Siegel, L., Ramey, C.T., Gottfried, A.W., Johnson, D.L. (1989). Home environment and cognitive development in the first 3 years of Life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology*, 25, 217-235.
- Brazelton, T.B. (1973). Neonatal Behavioural Assessment Scale. *Clinics in Developmental Medicine*, 50, Philadelphia: Lippincott.
- Broman, S.H. (1989). Infant physical status and alter cognitive development. In M.H. Bornstein & N.A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives*. (pp. 45-62). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Brooks-Gunn, J. & Lewis, M. (1988). The prediction of mental functioning in young handicapped children. In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (331-355). Philadelphia: Grune and Stratton.
- Brooks-Gunn, J. & Weinraub, M. (1983). Origins of infant intelligence testing. In M. Lewis (Ed.), *Origins of intelligence, infancy and early childhood* (2nd. ed.), (pp. 25-66). New York: Plenum Press.
- Brown, C.J. (1974). The effect of performance for visual complexity on habituation of visual fixation in infants. *Child Development*, 45, 1166-1169.
- Bushnell, I.W.P., McCutcheon, E. Sinclair, J. & Tweedlie, M.E. (1984). Infants' delayed recognition memory for colour and form. *British Journal of Developmental Psychology*, 2, 11-17.
- Bushnell, W., McCutcheon, E., Sinclair, J. & Tweedlie, M.E. (1984). Infants' delayed recognition memory for color and form. *British Journal of Developmental Psychology*, 2, 11-17.
- Campbell, S. K., Siegel, E., Parr, C. A., & Ramey C. T. (1986). Evidence for the need to renorm the Bayley Scales of Infant Development based on the performance of a population-based sample of 12-month-old infants. *Topics in early childhood special education*, 6, 83-96.
- Cardon, L. R., & Fulker, D.W. (1991). Sources of continuity in infant predictors of later IQ. *Intelligence*, 15, 279-293.

Caron, A.J., & Caron, R.F. (1981). Processing of relational information as an index of infant risk. In S.Friedman & S. Sigman (Eds.), *Preterm birth and psychological development*. (pp. 219-240). New York: Academic Press.

Caron, A.J., Caron, R.F., Minichiello, M.D., Weiss, S.J., & Friedman, S.L. (1977). Constraints on the use of the familiarization-novelty method in the assessment of infant discrimination. *Child Development*, 48, 748-762.

Catherwood, D. (1993). The robustness of infant haptic memory: Testing its capacity to withstand delay and haptic interference. *Child Development*, 64, 702-710.

Catherwood, D., Crassini, B. & Freiberg, K. (1989). Infant response to stimuli of similar hue and dissimilar shape: Tracing the origins of the categorization of objects by hue. *Child Development*, 60, 752-762.

Catherwood, D., Green, V., Skoien, P., & Holt, C. (1995). The cat among the pigeons: Categorical “pop-out” in infant visual attention. *Australian Journal of Psychology*, 47 (1), 1-7.

Cattell, P. (1940). *The Measurement of Intelligence of Infants and Young Children*. New York: Psychological Corporation.

Charlton, M. (1996). Social risk and development in the extremely low birthweight one-year-old. *Paper presented in 9th. Australian Human Development*. April 1996, Perth, Australia.

Coates, D.L., & Lewis, M. (1984). Early mother-infant interaction and infant cognitive status as predictors of school performance and cognitive behaviour in six-year-olds. *Child Development*, 55, 1219-1230.

Cohen, L.B. (1973). A two process model of infant visual attention. *Merrill-Palmer Quarterly*, 19 (3), 157-180.

Cohen, L.B. (1979). Our developing knowledge of infant perception and cognition. *American Psychologist*, 34 (10), 894-899.

Cohen, L.B. (1981). Examination of habituation as a measure of aberrant infant development. In S.Friedman & S. Sigman (Eds.), *Preterm birth and psychological development*. (pp. 241-253). New York: Academic Press.

Cohen, L.B. (1988a). Information processing constraints on infant memory and categorization. In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 237-252). Philadelphia: Grune and Stratton.

Cohen, L.B. (1988b). The relationship between infant habituation and infant information processing. *European Bulletin of Cognitive Psychology*, 8 (5), 445-454.

Cohen, L.B., DeLoache, J.S., & Pearl, R.A. (1977). An examination of interference effects in infants memory for faces. *Child Development*, 48, 88-96.

Cohen, L.B., DeLoache, J.S., & Strauss, M.S. (1979). Infant visual perception. Q. In J. D. Osofsky, (Ed.), *Handbook of infant development*. (pp. 439-461). New York: John Wiley & Sons.

Cohen, L.B., Gelber, E.R., & Lazar, M.A. (1971). Infant habituation and generalization to differing degrees of stimulus novelty. *Journal of Experimental Child Psychology*, 11, 379-389.

Cohen, S E., & Parmelee, A.H. (1983). Prediction of five-year Stanford-Binet scores in preterm infants. *Child Development*, 54, 1242-1253.

Colombo, J., Laurie, C., Martelli, T., & Hartig, B. (1984). Stimulus context and infant orientation discrimination. *Journal of Experimental Child Psychology*, 37, 576-586.

Colombo, J., & Mitchell D.W. (1988). Infant visual habituation: In defense of an information-processing analysis. *European Bulletin of Cognitive Psychology*, 8 (5), 455-461.

Colombo, J., & Mitchell, D.W. (1990). Individual differences in early visual attention: Fixation time and information processing. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*. (pp. 193-221). New Jersey: Lawrence Erlbaum Associates.

Colombo, J., Mitchell, D.W., Coldren, J.T., & Freesman, L.J. (1991). Individual differences in infant visual attention: Are short lookers faster processors or feature processor? *Child Development*, 62, 1247-1257.

Colombo, J., Mitchell, D.W., Dodd, J., Coldren, J.T., & Horowitz, F.D. (1989). Longitudinal correlates of infant attention in the paired-comparison paradigm. *Intelligence*, 13, 33-42.

Colombo, J., Mitchell, D.W., & Horowitz, F.D. (1988). Infant visual attention in the paired-comparison paradigm: Test-retest and attention-performance relations. *Child Development*, 59, 1198-1210.

Colombo, J., Mitchell, D.W., O'Brien, M., & Horowitz F.D. (1986). The stability of visual habituation during the first year of life. *Child Development*, 57, 474-487.

Cornell, E.H. (1979). Infants' recognition memory, forgetting and savings. *Journal of Experiment Child Psychology*, 28, 359-374.

Crockenberg, S. (1983). Early mother and infant antecedents of Bayley Scales performance at 21 months. *Developmental Psychology*, 19 (5), 727-730.

Crowe, T.K., Deitz, J.C., & Bennett, F.C. (1987). The relationship between the Bayley Scales of Infant Development and preschool gross motor and cognitive performance. *The American Journal of Occupational Therapy*, 41, 374-378.

- Daniel, A. E. (1983). *Power, privilege and prestige*. Melbourne, Australia: Longman Cheshire.
- Deary, I.J. (1988). Intelligence and encoding speed in infants, adults and children. *European Bulletin of Cognitive Psychology*, 8 (5), 462-468.
- DeCasper, A.J., & Spence, M.J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infants Behaviour and Development*, 9, 133-150.
- DeFries, J.C., Plomin, R., & LaBuda, M.C (1987). Genetic stability of cognitive development from childhood to adulthood. *Developmental Psychology*, 23(1), 4-12.
- de Lemos, M. (1989). *The use of the Stanford-Binet Intelligence Scale in Australia*.
- Detterman, D.K. (1982). Does "g" exist?. *Intelligence*, 6, 99-108.
- DiLalla, L.F., Thompson, L.A., Plomin, R., Phillips, K., Fagan III, J.F., Haith, M.M., Cyphers, L.H., & Fulker, D.W. (1990). Infant predictors of preschool and adult IQ: A study of infant twins and their parents. *Developmental Psychology*, 26, (5) 759-769.
- Dirks, J., & Gibson, E. (1977). Infants' perception of similarity between live people and their photographs. *Child Development*, 48, 124-130.
- Drotar, D., Mortimer, J., Shepherd, P.A., & Fagan III, J.F. (1989). Recognition memory as a method of assessing intelligence of an infant with quadriplegia. *Child Developmental Medicine and Neurology*, 31 (3), 391-394.
- Dunn, L.M. & Dunn, L.M. (1981). *Peabody Picture Vocabulary Test-Revised: Manual for forms L and M*. Circle Pines, MN: American Guidance Service.
- Escalona, S.K. & Moriarty, A. (1961). Prediction of school age intelligence from infant tests. *Child Development*, 32, 597-605.
- Eysenck, H.J. (1986). Toward a new model of intelligence. *Personality and Individual Differences*, 7 (5), 731-736.
- Eysenck, H.J. (1988). The concept of "intelligence": Useful or useless? *Intelligence*, 12, 1-16.
- Fagan III, J.F. (1970). Memory in the infant. *Journal of Experimental Child Psychology*, 9, 217-226.
- Fagan III, J.F. (1971). Infants' recognition memory for a series of visual stimuli. *Journal of Experimental Child Psychology*, 11, 244-250.
- Fagan III, J.F. (1972). Infants' recognition memory for faces. *Journal of Experimental Child Psychology*, 14, 453-476.

- Fagan III, J.F. (1973). Infants' delayed recognition memory and forgetting. *Journal of Experimental Child Psychology*, 16, 424-450.
- Fagan III, J.F. (1974). Infant recognition memory: The effect of length of familiarization and type of discrimination task. *Child Development*, 45, 351-356.
- Fagan III, J.F. (1976). Infants' recognition of invariant features of faces. *Child Development*, 47, 627-638.
- Fagan III, J.F. (1977). Infant recognition memory: Studies in forgetting. *Child Development*, 48, 68-78.
- Fagan III, J.F. (1981). Infant intelligence. *Intelligence*, 5, 239-243.
- Fagan III, J.F. (1984a). The intelligent infant: Theoretical implications. *Intelligence*, 8, 1-9.
- Fagan III, J.F. (1984b). The relationship of novelty preferences during infancy to later intelligence and later recognition memory. *Intelligence*, 8, 339-346.
- Fagan III, J.F. (1984c). Recognition memory and intelligence. *Intelligence*, 8, 31-36.
- Fagan III, J.F. (1988a). Evidence for the relationship between responsiveness to visual novelty during infancy and later intelligence: A summary. *European Bulletin of Cognitive Psychology*, 8 (5), 469-475.
- Fagan, J. F. (1988b). Screening infants for later mental retardation: From theory to practice. In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 253-265). Philadelphia: Grune and Stratton.
- Fagan III, J.F. (1990). The paired-comparison paradigm and infant intelligence. In A. Dismond (Ed.), *The development and neural bases of higher cognitive function* (pp. 337-364) New York: Academy of Science Press.
- Fagan, J.F. (1992). Intelligence: A theoretical viewpoint. *Current Directions In Psychological Science*, 1 (3), 82-86.
- Fagan III, J.F., & Detterman, D.K. (1992). The Fagan test of infant intelligence: A technical summary. *Journal of Applied Developmental Psychology*, 13, 173-193.
- Fagan, J.F., & Knevel, C.R. (1989). The prediction of above-average intelligence from infancy. Paper presented at the Society for Research in Child Development conference, Kansas City, Mo.
- Fagan III, J.F., & McGrath, S.K. (1981). Infant recognition memory and later intelligence. *Intelligence*, 5, 121-130.

Fagan III, J.F., & Montie, J.E. (1986). Identifying infants at risk for mental retardation: A cross-cultural validation study. *Abstract presented at the annual meeting of the society for behaviour pediatrics.*

Fagan III, J.F., & Montie, J.E. (1987). Identifying infants at risk for mental retardation: A cross-cultural validation study (abstract). *Paper presented at The Annual Meeting of The Society For Behaviour Pediatrics.* May 1986.

Fagan III, J.F., & Shepherd, P.A. (1979). Infants perception of face orientation. *Infant Behaviour and Development*, 2, 227-234.

Fagan III, J.F., & Shepherd, P.A. (1982). Theoretical issues in the early development of visual perception. In M. Lewis, & L.T. Taft (Eds.), *Developmental disabilities, Theory assessment and intervention* (pp. 9-34). UK:MTP Press Limited.

Fagan III, J.F., & Shepherd, P.A. (1986). *The Fagan Test of Infant Intelligence (Training manual)*. Cleveland, OH: Infantest Corporation.

Fagan III, J.F., & Shepherd, P.A. (1991). *The Fagan Test of Infant Intelligence (IBM Manual)*. Cleveland, OH: Infantest Corporation.

Fagan III, J.F., Shepherd, P.A., & Montie, J.E. (1987). An improved screening test for infants at risk for mental retardation (Abstract). *Paper presented at the Annual Meeting of The Society For Behaviour Paediatrics.* April, 1987.

Fagan III, J.F., & Singer, L. (1983). Infant recognition memory as a measure of intelligence. In L.P. Lipsitt (Ed.) *Advances in infancy research*, Vol. 2 (pp.31-78). Norwood, NJ: Ablex.

Fagan III, J.F., Singer, L.T., & Montie, J.E. (1985). An experimental selective screening device for the early detection of intellectual deficit in at-risk infants. In W.K. Frankenburg, R.N., Emde, & J.W. Sullivan (Eds.), *Early identification of children at risk: An international perspective* (pp. 257-265). New York: Plenum Press.

Fagan III, J.F., Singer, L.T., Montie, J.E., & Shepherd, P.A. (1986). Selective screening device for the early detection of normal or delayed cognitive development in infants at risk for later mental retardation. *Pediatrics*, 78 (2), 1021-1026.

Fagen, J.W., & Ohr, P.S. (1990). Individual differences in infant conditioning and memory. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*. (pp. 157-191). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fantz, R.L. (1956). A method for studying early visual development. *Perceptual and Motor Skills*, 6, 13-15.

Fantz, R.L. (1958a). Pattern vision in young infants. *The Psychological Records*, 8, 43-47.

Fantz, R.L. (1958b). Visual discrimination in a neonate chimpanzee. *Perceptual and Motor Skills*, 8, 59-66.

Fantz, R.L. (1963). Pattern vision in newborn infants. *Science*, 140, 296-297.

Fantz, R.L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, 146, 668-670.

Fantz, R. L. (1965). Visual perception from birth as shown by pattern selectivity. *Annals of The New York Academy of Science*, 118, 793-814

Fantz, R.L., & Miranda, S.B. (1975). Newborn infant attention to form and contour. *Child Development*, 46, 224-228.

Fantz, R.L., & Nevis, S. (1967). Pattern preferences and perceptual-cognitive development in early infancy. *Merrill-Palmer Quarterly of Behaviour and Development*, 13 (1), 77-108.

Faulkender, P. J., Wright, J. C., & Waldron, A. (1974). Generalized habituation of concept stimuli in toddlers. *Child Development*, 45, 1002-1010.

Fenson, L., Sapper, V., & Minner, D.G. (1974). Attention and manipulative play in the one-year-old child. *Child Development*, 45, 757-764.

Field, T.M., Dempsey, J.R., & Shuman, H.H. (1981). Developmental follow-up of pre- and postterm infants. In S.L. Friedman & M. Sigman (Eds.), *Preterm birth and Psychological development*. (pp. 299-312). New York: Academic Press.

Field, T., Hallock, N., Ting, G., Dempsey, J., Dabiri, C. & Shuman, H. (1978). A first year follow-up of high-risk infants: Formulating a cumulative risk index. *Child Development*, 49, 119-131.

Field, T.M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expressions by neonates. *Science*, 218, 179-181.

Flavell, J.H. (1982). On cognitive development. *Child Development*, 53, 1-10.

Flavell, J.H. (1992). Cognitive development: Past, present, and future. *Developmental Psychology*, 28 (6), 998-1005.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101 (2), 171-191.

Freedman, D.G., & Freedman, N.C. (1969). Behavioural differences between Chinese-American and European-American newborns. *Nature*, 224, 1227.

- Freeseman, L.J., Colombo, J., & Coldren, J.T. (1993). Individual differences in infant visual attention: Four month-olds' discrimination and generalization of global and local stimulus properties. *Child Development*, 64, 1191-1203.
- Fuggle, P. W., Tokar, S., Grant, D. B., & Smith, I. (1992). Rising IQ in British children: Recent evidence. *Journal of Child Psychology and Psychiatry*, 33 (7), 1241-1247.
- Gesell, A. (1925). *The Mental Growth of the Preschool Child*. New York: Macmillan.
- Ginsburg, H.P., & Opper, S. (1988). *Piaget's theory of intellectual development* (3rd. ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Goffeney, B., Henderson, N.B., & Butler, B.V. (1971). Negro-white, male-female eight-month developmental scores compared with seven-year WISC and Bender test scores. *Child Development*, 42, 595-604.
- Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *Journal of Child Language*, 17, 171-183.
- Gopnik, A. & Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, 58, 1523-1531.
- Gotlieb, S.J., Biasini, F.J., & Bray, N.W. (1988). Visual recognition memory in IUGN and normal birth-weight infants. *Infant Behaviour and Development*, 11, 223-228.
- Gottfried, A.W., & Bathurst, K. (1983). Hand preference across time is related to intelligence in young girls, not boys. *Science*, 221, 1074-1076.
- Gottfried, A.W., Rose, S.A., & Bridger, W.H. (1977). Cross-model transfer in human infants. *Child Development*, 48, 118-123.
- Graham, F.K., Matarazzo, R.G., & Caldwell, B.M. (1956). Behavioural differences between normal and traumatized newborns; I. The test procedures; II. Standardization, reliability and validity. *Psychological Monographs*, 70, Nos. 427-428.
- Griffiths, R. (1954). *Griffiths Mental Development Scales*. England: Young and son.
- Griffiths, R. (1986). *Griffiths Mental Development Scales*. England: The Test Agency Ltd.
- Gunderson, V.M., Grant, K.S., Burbacher, T.M., Fagan III, J.F., & Mottet, N.K. (1986). The effect of low-level prenatal methylmercury exposure on visual recognition memory in infant crab-eating Macaques. *Child Development*, 57, 1076-1083.
- Gunderson, V.M., Grant-Webster, K.S., & Sackett, G.P. (1989). Deficits in visual recognition in low birth weight infant pigtailed monkeys. *Child Development*, 60, 119-127.
- Hack, M., & Breslau, N. (1986). Very low birthweight infants: Effects of brain growth during infancy on intelligence quotient at 3 years of age. *Pediatrics*, 77 (2), 196-202.

- Hack, M., & Breslau, N. (1988). Biologic and social determinants of 3 year IQ in very low birthweight children. In P.M. Vietze & H.G. Vaughan, Jr. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 41-52). Harcourt, Brace, Jovanovich, Philadelphia: Grune and Stratton.
- Haith, M.M. (1980). *Rules that babies look by: The organization of newborn visual activity*. Hillsdale, NJ., Lawrence Erlbaum.
- Haith, M.M., Hazan, C., & Goodman, G.S. (1988). Expectation and anticipation of dynamic visual events by 3.5 month-old babies. *Child Development*, 59, 467-479.
- Halford, G.S. (1989). Reflections on 25 years of Piagetian cognitive developmental psychology, 1963-1988. *Human Development*, 32, 325-357.
- Haug, M.R. (1973). Social class measurement and woman's occupational roles. *Social forces*, 52, 86-98,
- Hawthorn, Victoria: Australian Council for Educational Research.
- Emanuelsson, I., & Svensson, A. (1990). Changes in intelligence over a quarter of a century. *Scandinavian Journal of Education Research*, 34 (3), 171-187.
- Hendrichsen, N., Skinhoj, K., & Andersen, C. (1986). Delayed growth and reduced intelligence in 9-17 year old intrauterine growth retarded children compared with their monozygous co-twins. *Acta Paediatrica Scandinavica* 75, 31-35.
- Hindley, C.B., & Owen, C.F. (1978). The extent of individual changes in IQ for ages between 6 months and 17 years in a British longitudinal sample. *Journal Of Child Psychology and Psychiatry*, 19, 329-350.
- Honzik, M.P. (1983). Measuring mental abilities in infancy. In M. Lewis (Ed.), *Origins of intelligence : Infancy and early childhood*. (pp. 67-105). New York: Plenum Press.
- Horowitz, F., Paden, L., Bhana, W. & Self, P. (1972). An infant control procedure for studying visual fixation. *Developmental Psychology*, 7, 90.
- Hull, P. (1994). Cognitive development and intelligence. In D.R. Matsumoto (Ed.), *People: Psychology from a cultural perspective*. (pp. 101-116). Belmont, California: Brooks/Cole Publishing Company.
- Humphreys, L.G., & Davey, T.C. (1988). Continuity in intellectual growth from 12 months to 9 years. *Intelligence*, 12, 183-197.
- Hunt, J.V.(1981). Predicting intellectual disorders in childhood for preterm infants with birthweights below 1501 gm. In S.L. Friedman & M. Sigman (Eds.), *Preterm birth and Psychological development* (pp. 329-351). New York: Academic Press.

- Hunter, M.A., Ross, H.S., & Ames, E.W. (1982). Preference for familiar or novel toys: Effect of familiarization time in 1-year-olds. *Developmental Psychology*, 18, 519-529.
- Huntington, L., Hans, S.L., & Zeskind, P.S. (1990). The relations among cry characteristics, demographic variables, and developmental test scores in infants prenatally exposed to methadone. *Infants Behaviour and Development*, 13, 533-538.
- Jacobson, S.W., Fein, G.G., Jacobson, J.L., Schwartz, P.M., & Dowler, J.K. (1985). The effect of intrauterine PCB exposure on visual recognition memory. *Child Development*, 56, 853-860.
- Jacobson, S.W., Jacobson, J.L., O'Neill, J.M., Padgett, R.J., Frankowski, J.J., & Bihun, J.T. (1992). Visual expectation and dimensions of infant information processing. *Child Development*, 63, 711-724.
- Jaffa, A.S. (1934). *The California Preschool Mental Scale*. Berkeley: University of California Press.
- Jenkinson, J.C. (1983). Is speed of information processing related to fluid or to crystallized intelligence? *Intelligence*, 7, 91-106.
- Jensen, A.R., & Weng, L. (1994). What is a good g?. *Intelligence*, 18, 231-258.
- Johnson, D., & Brody, N. (1977). Visual habituation, sensorimotor development and tempo of play in one-year old infants. *Child Development*, 48, 315-319.
- Johnson, D., Swank, P., Howie, V., Baldwin, C., Owen, M., & Luttmann, D. (1993). Does HOME add to the prediction of child intelligence over and above SES. *The Journal of Genetic Psychology*, 154 (1), 33-40.
- Kagan, J. (1971). *Change and continuity in infancy*. New York: John Wiley.
- Kagan, J. (1989). Commentary, *Human Development*, 32, 172-176.
- Kagan, J., Lapidus, D.R., & Moore, M. (1978). Infants antecedents of cognitive functioning: A longitudinal study. *Child Development*, 49, 1005-1023.
- Kaplan-Estrin, M., Jacobson, S.W., & Jacobson, J.L. (1994). Alternative approaches to clustering and scoring the Bayley Infant Behaviour Record. *Infant Behaviour and Development*, 17, 149-157.
- Kirby, N.H. & Nettelbeck, T. (1991). Speed of information processing and age. *Personality and Individual Differences*, 12, 183-188.
- Kitchen, W. H.; Rickards, A., Ford, G. W.; Doyle, L., Kelly, E., & Ryan, M. (1989). Selective improvement in cognitive test scores of extremely low birthweight infants aged between 2 and 5 years. *Australian Journal of Paediatrics* 25, 288-291.

- Knobloch, H. & Pasmanick, B. (1974). Gesell and Amatruda's Developmental Diagnosis. New York: Harper & Row.
- Kopp, C.B. & McCall, R.B. (1982). Predicting later mental performance for normal, at risk, and handicapped infants. In P.B. Baltes & O.G. Brim (Eds.), *Life-span development and behaviour* (Vol. 4, pp. 33-61). New York: Academic Press.
- Kopp, C.B., & Vaughan, B.E. (1982). Differences in sustained attention at 8 months predicted later MDI at 24 months. *Child Development*, 53, 174-182.
- Kranzler, J.H., & Jensen, A.R. (1991). The nature of psychometric g: Unitary process or a number of independent processes? *Intelligence*, 15, 397-422.
- Lamb, M.E., Garn, S.M., & Keating, M.T. (1981). Correlations between sociability and cognitive performance among eight-month olds. *Child Development*, 52, 711-713.
- Largo, R.H., Graf, S., Kundu, S., Hunziker, U., & Molinari, L. (1990). Predicting developmental outcome at school age from infant tests of normal, at-risk and retarded infants. *Developmental Medicine and Child Neurology*, 32, 30-45.
- Lasky, R.E. (1980). Length of familiarization and preference for novel and familiar stimuli. *Infant Behaviour and Development*, 3, 15-28.
- Lasky, R.E., Tyson, J.E., Rosenfeld, C.R., Priest, M., Krasinski, D., Heartwell, S., & Gant, N.F. (1983). Differences on Bayley's infant behaviour record for a sample of high-risk infants and their controls. *Child Development*, 54, 1211-1216.
- Laucht, M., Esser, G., & Schmidt, M.H. (1994). Contrasting infant predictors of later cognitive functioning. *Journal child psychology and psychiatry*, 35 (4), 649-662.
- Leahy, R.L (1976). Development of preferences and processes of visual scanning in the human infant during the first 3 months of life. *Developmental Psychology*, 12 (3), 250-254.
- Lecuyer, R. (1988). Please infant, can you tell me exactly what you are doing during a habituation experiment? *European Bulletin of Cognitive Psychology*, 8 (5), 476-480.
- Lecuyer, R. (1989). Habituation and attention, novelty and cognition: Where is the continuity? *Human Development*, 32, 148-157.
- Lewis, M. (1982). Attention as a measure of cognitive integrity. In M. Lewis, & L.T. Taft (Eds.), *Developmental disabilities, Theory assessment and intervention*. (pp. 185-212). UK:MTP Press Limited.
- Lewis, M. (1983). On the nature of intelligence, science or bias. In M. Lewis (Ed.), *Origins of intelligence: infancy and early childhood*. (pp. 1-24). New York: Plenum Press.
- Lewis, M. (1989). Commentary. *Human Development*, 32, 216-222.

Lewis, M., & Baldini, N. (1979). Attentional processes and individual differences. In G.A. Hale & M. Lewis. (Eds.), *Attention and cognitive development*. (pp. 135-172). New York: Plenum Press.

Lewis, M., & Brooks-Gunn, J. (1981). Visual attention at three months as a predictor of cognitive functioning at two years of age. *Intelligence*, 5, 131-140.

Lewis, M., & Brooks-Gunn, J. (1984). Age and handicapped group differences in infants' visual attention. *Child Development*, 55, 858-868.

Lewis, M., Jaskir, J., & Enright, M.K. (1986). The development of mental abilities in infancy. *Intelligence*, 10, 331-354.

Lewis, M., & McGurk, H. (1972). Evaluation of infant intelligence: Infant intelligence scores - true or false?. *Science*, 178, 1174-1177.

Lewis, M., & Sullivan, M.W. (1985). Infant intelligence and its assessment. In B.B. Wolman (Ed.), *Handbook of intelligence: Theories, measurement and applications*. New York: John Wiley.

Lynn, R., (1989). A nutrition theory of the secular increases in intelligence: Positive correlations between height, head size and IQ. *British Journal of Educational Psychology* 59, 372-377.

Lynn, R., & Hampson, S. (1989). Secular increases in reasoning and mathematical abilities in Britain, 1972-84. *School Psychology International*, 10, 301-304.

Lynn, R., Hampson, S. L., & Mullineux, J. C. (1987). A long-term increase in the fluid intelligence of English children. *Nature*, 328, 797.

MacTruk, R.H., Vietze, P.M., McCarthy, M.E., McQuiston, S., & Yarrow, L.J. (1985). The organisation of exploratory behavior in Down Syndrome and nondelayed infants. *Child Development*, 56, 573-581.

Malcuit, G., Pomerleau, A., & Lamarre, G. (1988a). Habituation, visual fixation, and cognitive activity in infants: A critical analysis and attempt at a new formulation. *European Bulletin of Cognitive Psychology*, 8 (5), 415-440.

Malcuit, G., Pomerleau, A., & Lamarre, G. (1988b). Habituation and operant visual fixation: A comment on comments. *European Bulletin of Cognitive Psychology*, 8 (5), 539-547.

McCall, R. (1971). Attention in infant: Avenue to the study of cognitive development. In D.N. Walcher & D.L. Peters (Eds.) *Early childhood: The developmental of self-regulatory mechanism*. (pp.107-137). New York: Academic Press.

McCall, R.B. (1979a). The development of intellectual functioning in infancy and the prediction of later IQ. In J. D. Osofsky, (Ed.), *Handbook of infant development*. (pp.707-741). New York: John Wiley & Sons.

- McCall, R.B. (1979b). Individual differences in the pattern of habituation at 5 and 10 months of age. *Developmental Psychology*, 15 (5), 559-569.
- McCall, R.B. (1981a). Nature-nature and the two realms of development: A proposed integration with respect to mental development. *Child Development*, 52, 1-12.
- McCall, R.B. (1981b). Early predictors of later IQ: The search continues. *Intelligence*, 5, 141-147.
- McCall, R.B. (1982). Issues in the early development of intelligence and its assessment. In M.Lewis, & L.T. Taft (Eds.), *Developmental disabilities: Theory, assessment and intervention*. (pp. 177-184). UK: MTP Press Limited.
- McCall, R.B. (1988a). Habituation, response to new stimuli, and information processing in human infants. *European Bulletin of Cognitive Psychology*, 8 (5), 481-488.
- McCall, R.B. (1988b). Identifying developmental disabilities: Resume and future directions. In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 385-401). Philadelphia: Grune and Stratton.
- McCall, R.B. (1989). Commentary, *Human Development*, 32, 177-186.
- McCall, R.B. (1994). What process mediates predictions of childhood IQ from infant habituation and recognition memory? Speculations on the role of inhibition. *Intelligence*, 18 (2), 107-125.
- McCall, R.B., & Carriger, M.S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development*, 64, 57-79.
- McCall, R.B., Eichorn, D.H. & Hogarty, P.S. (1977). Transitions in early mental development. *Monographs of the society for research in child development*, 42 (3, serial No. 171).
- McCall, R.B., Hogarty, P.S., & Hurlburt, N. (1972). Transitions in infant sensorimotor development and the prediction of childhood IQ. *American Psychologist*, 27, 728-748
- McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*. San Antonio, TX: The Psychological Corporation.
- McDonough, S.C. (1988). Early information processing in aberrant infants. In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 318-328). Philadelphia: Grune and Stratton.
- McDonough, S.C. & Cohen, L.B. (1982). Attention and memory in cerebral palsied infants. *Infant Behaviour and Development*, 5, 347-354.
- Meltzoff, A.N., & Moore, M.K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75-78.

Meltzoff, A.N., & Moore, M.K. (1983). Newborn infants imitate adult. *Child Development*, 54, 702-709.

Millar, W.S., & Weir, C.G. (1992). Relations between habituation and contingency learning in 5- to 12-month old infants. *European Bulletin of Cognitive Psychology*, 12(3), 209-222.

Millar, W.S., Weir, C.G., & Supramaniam, G. (1991). The relationships between encoding, discriminative capacities and perinatal risk status in 4-12 month old infants. *Journal of Child Psychology and Psychiatry*, 32 (3), 473-488.

Miller, D.J., Ryan, E.B., Short, E.J., Ries, P.G., McGuire, M.D., & Culler, P. (1977). Relationships between early habituation and later cognitive performance in infancy. *Child Development*, 48, 658-661.

Miller, D.J., Sinnott, J.P., Short, E.J. & Hanis, A.A. (1976). Individual differences in habituation rates and object concept performance. *Child Development*, 47, 528-531.

Miller, L.T., & Vernon, P.A. (1992). The general factor in short-term memory, intelligence and reaction time. *Intelligence*, 16, 5-29.

Miranda, S. B. (1970). Response to novel visual stimuli by Down's Syndrome and normal infants. *The proceedings, 78th Annual Convention APA*.

Miranda, S. B., & Fantz, R.L. (1973). Visual performance of Down's Syndrome and normal infants. *Child Development*, 44, 555-561.

Miranda, S.B., & Fantz, R.L. (1974). Recognition memory in Down's Syndrome and normal infants. *Child Development*, 45, 651-660.

Moffitt, T.E., Caspit, A., Haskness, A.R., & Silva, P.A. (1993). The natural history of change in intellectual performance: Who changes? How much? Is it meaningful? *Journal Of Child Psychology and Psychiatry*, 90, 152-156.

Morrongiello, B.A. (1990). The study of individual differences in infants: Auditory processing measures. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*. (pp. 271-320). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mueller, C.W., & Parcel, T.L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development*, 52, 13-30.

Mundy, P.C., Seibert, J.M., Hogan, A.E., & Fagan III, J.F. (1983). Novelty responding and behavioral development in young developmentally delayed children. *Intelligence*, 7, 163-174.

Myers, D.G. (1986). *Psychology*. New York: Worth Publisher, Inc.

- Nelson, C.A., Henschel, M., & Collins, P.F. (1993). Neural correlates of cross-modal recognition memory by 8-month-old human infants. *Developmental Psychology*, 29 (3), 411-420.
- Nettelbeck, T (1987). Inspection time and intelligence. In P.A., Vernon (Ed.), *Speed of information processing and intelligence* (pp. 295-356). Norwood, NJ: Ablex.
- Nettelbeck, T. & Wilson, C. (1985). A cross-sequential analysis of developmental differences in speed of visual information processing. *Journal of Experimental Child Psychology*, 40, 1-22.
- Nettelbeck, T. & Young, R. (1989). Inspection time and intelligence in 6-year-old children., 605-614. *Personality and Individual Differences*, 10, 605-614.
- O'Connor, M.J.(1980). A comparison of preterm and full-term infants on auditory discrimination at four months and on Bayley Scales of infants development at eighteen months. *Child Developmental* 51, 81-88.
- O'Connor, M.J., Cohen, S., & Parmelee, A.H. (1984). Infant auditory discrimination in preterm and full term infant as a predictor of 5 year intelligence. *Developmental Psychology*, 20.(1), 159-165.
- Olson, G.M., & Sherman, T. (1983). Attention, learning and memory in infants. In M.M., Haith & J.J., Campos (volume Eds.), *Handbook of Child Psychology* (4th. ed.). Vol 2, (pp 1001-1080). New York: John Wiley.
- Pecheux, M., & Lecuyer, R. (1983). Habituation rate and free exploration tempo in 4-month-old infants. *International Journal of Behaviour Development*, 6, 37-50.
- Pecheux, M., & Lecuyer, R. (1989). A longitudinal study of visual habituation between 3, 5 and 8 months of age. *British Journal of Developmental Psychology*, 7, 159-169.
- Peterson, C. (1996). Looking forward through the lifespan: *Developmental Psychology* (3rd Ed.). Sydney: Prentice Hall.
- Philips, J.L. (1981). *Piaget's theory: A primer*. San Francisco: W.H. Freeman and Company.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Pianta, R.C., & Egeland, B. (1994). Predictors of instability in children's mental test performance at 24, 48, and 96 months. *Intelligence*, 18, 145-163.
- Plomin, R., Campos, J., Corley, R., Emde, R.N., Fulker, D.W., Kagan, J., Reznick, J.S., Robinson, J., Zahn-Waxler, C., & DeFries, J.C. (1990). Individual differences during the second year of life: The MacArthur longitudinal twin study. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*. (pp. 431-455). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Plomin, R., & DeFries, J.C. (1983). The Colorado Adoption Project. *Child Development*, 54, 276-289.
- Pomerleau, A. (1989). Commentary. *Human Development*, 32, 167-171.
- Pomerleau, A., Maitre, G., & Malcuit, G. (1989). Stability of habituation responses across multiple sessions in 4-month-old infants. *Journal of Genetic Psychology*, 150 (4), 445-448.
- Prior, M., Garino, E., Sanson, A. & Oberklaid, F. (1987). Ethnic influences in difficult temperament and behavioural problems in infants. *Australian Journal of Psychology*, 39, 163-171.
- Quinn, P.C. (1994). The categorization of above and below spatial relations by young infants. *Child Development*, 65, 58-69.
- Ramey, C.T., Campbell, F.A., & Nicholson, J.E. (1973). The predictive power of the Bayley Scales of Infant Development and the Stanford-Binet intelligence test in a relatively constant environment. *Child Development*, 44, 790-795.
- Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology*, 28, 406-413.
- Richards J.E. (1988). Heart rate responses and heart rate rhythms, and infant visual sustain attention. In P.K. Ackles, J.R. Jennings, & M.H. Coles (Eds.) *Advances in psychophysiology*, Vol. 3. (pp. 189-221) Greenwich, CT: JAI Press.
- Richards J.E. (1997). Effects of attention on infants' preference for briefly exposed visual stimuli in the paired-comparison recognition-memory paradigm. *Developmental Psychology*, 33, 1-22.
- Ritz, E.G., Woodruff, A.B., Fagen, J.W. (1984). Short- and long-term stability of habituation rate and looking time in four-month old infants. *The Journal of Genetic Psychology*, 144, 285-286.
- Roberts, K. (1988) Retrieval of a basic-level category in prelinguistic infants. *Developmental Psychology*, 24, 21-27.
- Roe, K.V., McClure, A. & Roe, A. (1983). Infant Gesell scores vs. cognitive skills at age 12 years. *The journal of Genetic Psychology*, 142, 143-147.
- Rolfe, S.A. (1994). Does assessment of cognitive functioning in infancy hold the key to early detection of developmental disabilities? A review of research. *Australia and New Zealand Journal of Developmental Disabilities*, 19 (1), 61-72.
- Rose, D. H., Slater, A., & Perry, H. (1986). Prediction of childhood intelligence from habituation in early infancy. *Intelligence*, 10, 251-263.

- Rose, S.A. (1977). Infant's transfer of response between two-dimensional and three-dimensional stimuli. *Child Development*, 48, 1086-1091.
- Rose, S. A. (1980). Enhancing visual recognition memory in preterm infants. *Developmental psychology*, 16, (2), 85-92.
- Rose, S.A., (1981). Developmental changes in infants' retention of visual stimuli. *Child Development*, 52, 227-233.
- Rose, S.A., (1983). Differential rates of visual information processing in full-term and preterm infants. *Child Development*, 54, 1189-1198.
- Rose, S.A. (1988). Predicting cognitive development from infancy measures. In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 266-281). Philadelphia: Grune and Stratton.
- Rose, S.A., & Feldman, J.F. (1987). Infant visual attention: Stability of individual differences from 6 to 8 months. *Developmental Psychology* 23 (4), 490-498.
- Rose, S.A., & Feldman, J.F. (1990). Infant cognition: Individual differences and developmental continuities. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*. (pp. 229-245). New Jersey: Lawrence Erlbaum Associates.
- Rose, S.A., Feldman, J.F., McCarton, C.M., & Wolfson, J. (1988b). Information processing in seven-month-old infant as a function of risk status. *Child Development*, 59, 589-603.
- Rose, S.A., Feldman, J.F., & Wallace, I.F. (1988a). Individual differences in infants' information processing: Reliability, stability and prediction. *Child Development*, 59, 1177-1197.
- Rose, S.A., Feldman, J.F., & Wallace, I.F. (1992). Infant information processing in relation to six-year cognitive outcomes. *Child Development*, 63, 1126-1141.
- Rose, S.A., Feldman, J.F., Wallace, I.F., & Cohen, P. (1991b). Language: A partial link between infant attention and later intelligence. *Developmental Psychology*, 27 (5), 798-805.
- Rose, S.A., Feldman, J.F., Wallace, I.F., & McCarton, C. (1989). Infant visual attention: Relation to birth status and developmental outcome during the first 5 years. *Developmental Psychology*, 25 (4), 560-576.
- Rose, S.A., Feldman, J.F., Wallace, I.F., & McCarton, C. (1991a). Information processing at 1 year: Relation to birth status and developmental outcome during the first 5 years. *Developmental Psychology*, 27 (5), 723-737.
- Rose, S.A., Gottfried, A.W. & Bridger, W.H. (1978). Cross-modal transfer in infants: Relationship to prematurity and socioeconomic background. *Development Psychology*, 14 (6), 643-652.

Rose, S.A., Gottfried, A.W., Melloy-Carminar, P., & Bridger, W.H. (1982). Familiarity and novelty preference in infant recognition memory: Implications for information processing. *Developmental Psychology*, 18 (5), 704-713.

Rose, S.A., & Wallace, I.F., (1985). Visual recognition memory: A predictor of later cognitive functioning in preterms. *Child Development*, 56, 843-852.

Rosenblith, J.F. (1979). The Graham/Rosenblith behaviour examination for newborns: Prognostic value and procedural issues. In J. D. Osofsky, (Ed.), *Handbook of infant development*. (pp.216-249). New York: John Wiley & Sons.

Ross, G., Lipper, E.G., & Auld, P.A.M. (1985). Consistency and change in the development of premature infants weighing less than 1501 grams at birth. *Pediatrics*, 76, 885-891.

Ross, G., Lipper, E.G., & Auld, P.A.M. (1986). Early predictors of neurodevelopmental outcome of very low birthweight infant at three years. *Developmental Medicine and Child Neurology*, 28 (2). 171-179.

Ross, L.E., & Ross, S.M. (1981). The visual scanning and fixation behaviour of the retarded. *International Review of Research in Mental Retardation*, (Vol. 10). (pp. 1-30). New York: Academic Press.

Rovee-Collier, C. (1988). The adaptive significance of habituation in infancy. *European Bulletin of Cognitive Psychology*, 8 (5), 503-511.

Rubin, R.A., Balow, B. (1979). Measures of infant development and socioeconomic status as predictors of later intelligence and school achievement. *Developmental Psychology*, 15, 225-227.

Ruddy, M.G., & Bornstein, M.H. (1982). Cognitive correlates of infant attention and maternal stimulation over the first year of life. *Child Development*, 53, 183-188.

Ruff, H.A. (1988). The measurement of attention in high-risk infants. In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 282-296). Philadelphia: Grune and Stratton.

Ruff, H.A. (1990). Individual differences in sustained attention during infancy. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*. (pp. 247-270). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ruff, H.A., Lawson, K.R., Parrinello, R., & Weissberg, R. (1990). Long-term stability of individual differences in sustained attention in the early years. *Child Development*, 61, 60-75.

Ruff, H.A., McCarton, C., Kurtzberg, D., & Vaughan, H. (1984). Sustained attention during exploratory manipulation as a predictor of cognitive competence in preterm infants. *Child Development*, 55, 1166-1173.

- Rushton, J.p., Brainerd, C.J., & Pressley (1983). Behavioural development and construct validity: The principle of aggregation. *Psychological Bulletin*, 94, 18-38.
- Saayman, G., Ames, E.W., & Moffett, A. (1964). Response to novelty as an indicator of visual discrimination in the human infant. *Journal of Experimental Child Psychology*, 1, 189-198.
- Sameroff, A. J. Seifer, R. Baldwin, A. & Boldwin, C. (1993). Stability of intelligence from preschool to adulthood: The influence of social and family risk factors. *Child Development*, 64, 80-97.
- Sanson, A.V., Prior, M. & Oberklaid, F. (1985). Normative data on temperament in Australian infants. *Australian Journal of Psychology*. 37, 185-195.
- Sattler, J. M. (1988). *Assessment of children*. (3rd. Ed.). San Diego: Jerome M. Sattler.
- Scarr, S., Weinberg, R.A., & Waldman, I.D. (1993). IQ correlations in transracial adoptive families. *Intelligence*, 17, 541-555.
- Schiff, W., Benasich, A.A., & Bornstein, M.H. (1989). Infant sensitivity to audiovisually coherent events. *Psychological Research*, 51, 102-106.
- Self, P.A., & Horowitz, F.D. (1979). The behavioral assessment of the neonate: An overview. In Osofsky, J. D. (Ed.) *Handbook of infant development*. (Pp. 126-160). New York: John-Wiley & Sons.
- Shapiro, B.K., Palmer, F.B., Antell, S.U., Bilker, S., Ross, A., & Caputa, A.J. (1989). Giftedness: can it be predicted in infancy? *Clinical Pediatrics*, 28 (5), 205-209.
- Shavelson. (1988). *Statistical reasoning for behavioral science*. (2nd ed.). Boston: Allyn and Bacon.
- Shepherd, P.A., Fagan III, J.F. (1981). Visual pattern detection and recognition memory in children with profound mental retardation. *International Review of Research In Mental Retardation*. (Vol. 10) (pp. 31-60). New York: Academic Press.
- Shiffrin, R.M. & Atkinson, R.C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, 76 (2), 179-193.
- Siegel, L.S. (1981). Infant tests as predictors of cognitive and language development at two years. *Child Development*, 52, 545-557.
- Siegel, L.S. (1983). Correction for prematurity and its consequences for the assessment of the very low birth weight infant. *Child Development*, 54, 1176-1188.
- Siegel, L.S. (1989). A reconceptualization of prediction from infant test scores. In M.H. Bornstein & N.A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives*. (pp. 89-103). Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegler, R.S., & Richards, D.D. (1982). The development of intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence*. (pp. 897-970).: New York: Cambridge University Press.

Sigman, M. (1988). Infant attention: What processes are measured? *European Bulletin of Cognitive Psychology*, 8 (5), 512-516.

Sigman, M.D. (1983). Individual differences in infant attention: Relation to birth status and intelligence at 5 years. In Field, T. & Sostek, A. (Eds.), *Infants born at risk: Physiological, perceptual and cognitive processes* (pp 271-293). New York: Grune and Stratton.

Sigman, M.D, Cohen, S.E., Beckwith, L., & Parmelee, A.H. (1986). Infant attention in relation to intelligence abilities in childhood. *Developmental Psychology*, 22.(6), 788-792.

Sigman, M., Cohen, S.E., & Forsythe, A.B. (1981). The relation of early infant measures to later development. In S.L. Friedman & M. Sigman. (Eds.), *Preterm birth and psychological development*. (pp. 313-327). New York: Academic Press.

Sigman, M. & Parmelee, A. H. (1974). Visual preferences of four-month-old premature and full -term infants. *Child Development*, 45, 959-965.

Siqueland, E.R. (1981). Studies of visual recognition memory in preterm infants: Differences in development as a function of perinatal morbidity factors. In S.L. Friedman & M. Sigman. (Eds.), *Preterm birth and psychological development*. New York: Academic Press.

Slater, A. (1988). Habituation and visual fixation in infants: Information processing, reinforcement, and what else? *European Bulletin of Cognitive Psychology*, 8 (5), 517-523.

Slater, A. (1995). Individual differences in infancy and later IQ. *Journal of Child Psychology and Psychiatry*, 36 (1), 69-112.

Slater, A., Cooper, R., Rose, D., & Morison, V. (1989). Prediction of cognitive performance from infancy to early childhood. *Human Development*, 32, 137-147.

Slater, A., Mattock, A., Brown, E., & Bremner, J.G. (1991). Form perception at birth: Cohen and Younger (1984) revisited. *Journal of Experimental Child Psychology*, 51, 395-406.

Slater, A., & Sykes, M. (1977). Newborn infants' visual responses to square wave grating. *Child Development*, 48, 545-554.

Sokolov, N. (1963). *Percepton and the conditioned reflex*. S.W. Waydenfeld (Translation). New York: Macmillan.

Spitz, H. H. (1989). Variations in Wechsler interscale IQ disparities at different level of IQ. *Intelligence*, 13, 157-167.

Stechler, G., Bradford, S. & Levy, H. (1966). Attention in newborn: Effect on motility and skin potential. *Science*, 151, 1246-1248.

Sternberg, R.J., (1981). Novelty-seeking, novelty-finding, and the developmental continuity of intelligence. *Intelligence*, 5, 149-155.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R.J. (1995). *In search of the human mind*. New York: Harcourt Brace & Company.

Sternberg, R.J., & Okagaki, L. (1989). Continuity and discontinuity in intellectual development are not a matter of "either-or". *Human Development*, 32, 158-166.

Struthers, J.M., & Hansen, R.L. (1992). Visual recognition memory in drug-exposed infants. *Developmental and Behavioural Pediatrics*, 13 (2), 108-111.

Swain, I.U., Zelazo, P.R., & Clifton, R.K. (1993). Newborn infants' memory for speech sounds retained over 24 hours. *Developmental Psychology*, 29 (2), 312-323.

Tamis-LeMonda, C.S., & Bornstein, M.H. (1989). Habituation and maternal encouragement of attention in infancy as predictors of toddler language, play, and representational competence. *Child Development*, 60, 738-751.

Teasdale, T. W., & Owen, D. R. (1987). National secular trends in intelligence and education: A twenty-year cross-sectional study. *Nature*, 325, 119-121.

Teasdale, T. W., & Owen, D. R. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, 13, 255-262.

Terman, L.M. & Merrill, M.A. (1960). *Stanford-Binet Intelligence Scale: Manual for the third Revision, Form L-M*. Boston: Houghton Mifflin.

Thompson, L.A., Fagan, J.F., & Fulker, D.W. (1991). Longitudinal prediction of specific cognitive abilities from infant novelty preference. *Child Development*, 62, 530-538.

Thompson, L.A., Plomin, R., & DeFries, J.C. (1985). Parent-infant resemblance for general and specific cognitive abilities in the Colorado Adoption Project. *Intelligence*, 9, 1-13.

Thompson, B., Wasserman, J. & Matula, K. (1996) The factor structure of the behaviour rating scale of the Bayley Scales of Infant Development - II. Educational and Psychological measurement, 3, 460-474.

Thorndike, R.L. (1994). g. *Intelligence*, 19, 145-155.

Thorndike, R. L., Hagan, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition. Guide for administering and scoring*. Chicago, Ill: Riverside.

- Turpin, G., & Rose, D. (1988). What is the role of habituation in malcuit et al.'s "new formulation"? *European Bulletin of Cognitive Psychology*, 8 (5), 524-530.
- Uzgiris, I.C. (1989). Transformations and continuities: Intellectual functioning in infancy and beyond. In M.H. Bornstein & N.A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives*. (pp. 123-143). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Uzgiris, I.C. & Hunt, J. McV. (1975). Assessment in infancy: Ordinal scale of Psychological development. Urbana: University of Illinois Press.
- Vernon, P.A. (1983). Speed of information processing and general intelligence. *Intelligence*, 7, 53-70.
- Vernon, P.A. (1989). The generality of g. *Personality and Individual Differences*, 10 (7), 803-804.
- Vernon, P. A., & Mori, M. (1992). Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence*, 16, 273-288.
- Vietze, P.M. (1988). Do we need another test? In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 402-410). Philadelphia: Grune and Stratton.
- Wachs, T.D. (1975). Relation of infant's performance on piaget scales between twelve and twenty-four months and their Stanford-Binet performance at thirty-one months. *Child Development*, 46, 929-935.
- Wachs, T.D., & Smitherman, C.H. (1985). Infant temperament and subject loss in a habituation procedure. *Child Development*, 56, 861-867.
- Wade, C. & Tavris, C. (1990). *Psychology*. New York: Harper & Row Publishers.
- Wagner, S.H., & L.J. (1986). A process analysis of infant visual and cross-modal recognition Sakovits memory: Implication for an amodal code. *Advances in infancy research* vol. 4, 196-217.
- Wallace, F. (1988). Socioenvironmental issues in longitudinal research of high-risk infants. In P.M. Vietze & H.G. Vaughan. (Eds.), *Early identification of infants with developmental disabilities*. (pp. 356-382). Philadelphia: Grune and Stratton.
- Weinberg, R.A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist*, 44(2), 98-104.
- Weiss, M.J.S., & Zelazo, P.R. (1991). A taxonomy of newborn attention. In M.J.S. Weiss & P.R. Zelazo (Eds.), *Newborn attention: Biological constraints and the influence of experience* (pp.466-511). NJ: Ablex.

- Weizmann, F., Cohen, L.B. & Pratt, J. (1971). Novelty, familiarity, and the development of infant attention. *Developmental Psychology*, 4, 149-154.
- Widmayer, S., Peterson, L., Lerner, M., Carnahan, J., Calderon, A., Wingerd, J. & Marshall, R. (1990). Predictors of Haitian_American infant development at twelve months. *Child Development*, 61, 410-415.
- Wilson, R.S. (1978). Synchronies in mental development: An epigenetic perspective. *Science*, 202, 939-948.
- Wilson, R.S. (1983). The Louisville Twin Study: Developmental Synchronies in behaviour. *Child Development*, 54, 298-316.
- Wilson, R.S., & Matheny, A.P. (1983). Mental development: Family environment and genetic influences. *Intelligence*, 7, 195-215.
- Wohlwill, J.F. (1980). Cognitive development in childhood. In O.G. Brim & J. Kagan (Eds.), *Constancy and change in human development*. (pp. 359-444). Massachusetts: Harvard University Press.
- Worobey, J. (1990). Behavioural assessment of the neonate. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*. (pp. 137-154). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Yang, R.K. (1979). Early infant assessment: An overview. In J.D. Osofsky (Ed.) *Handbook of Infant Development*. New York: John Wiley & Sons.
- Yarrow, L.J., Klein, R.P., Lomonaco, S., & Morgan, G.A. (1975). Cognitive and motivational development in early childhood. In B.X. Friedlander, G.M. Sterritt, & G.E. Kirk (Eds.), *Exceptional Infant* (Vol. 3, pp. 491-502). New York: Brunner/Mazel.
- Young-Browne, G., Rosenfeld, H.M., & Horowitz, F.D. (1977). Infant discrimination of facial expressions. *Child Development*, 48, 555-562.
- Younger, B., & Gotlieb, S. (1988). Development of categorization skills: Changes in the nature or structure of infant form categories? *Developmental Psychology*, 24 (5), 611-619.
- Zelazo, P.R. (1982). An information processing approach to infant cognitive assessment. In M.Lewis, & L.T. Taft (Eds.), *Developmental disabilities: Theory, assessment and intervention*. (pp. 229-255). UK: MTP Press Limited.