



## REFERENCE ONLY

## UNIVERSITY OF LONDON THESIS

Degree **phd**

Year **2006**

Name of Author **ALJUNTEZ SMITH**

### COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

### COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

### LOANS

Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

### REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

- A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).
- B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.
- C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.
- D. 1989 onwards. Most theses may be copied.

**This thesis comes within category D.**



This copy has been deposited in the Library of

UC



This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.



# THE ASSESSMENT OF THE SPECIAL EDUCATIONAL NEEDS OF CHILDREN WITH AUTISM IN SINGAPORE

SHARIFAH MARIAM B. H. ALJUNIED

DEPARTMENT OF PSYCHOLOGY  
UNIVERSITY COLLEGE LONDON  
26 BEDFORD WAY  
LONDON WC1E 6BT  
UNITED KINGDOM

THESIS SUBMITTED TO THE FACULTY OF LIFE SCIENCES AT THE UNIVERSITY  
COLLEGE LONDON FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SEPTEMBER 2005

UMI Number: U592607

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.

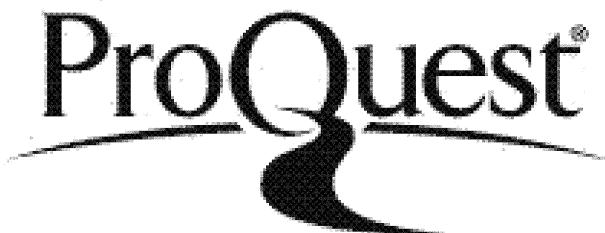


UMI U592607

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

*I owe a lot to many,  
but this thesis is dedicated  
to those that matter the most,  
my family,  
Mohd Najiib and Sara Nazirah.*

## ACKNOWLEDGEMENT

I would like to express sincere appreciation to my supervisors, Prof. Norah Frederickson and Prof. Uta Frith, for their guidance and insights throughout the research.

I would like to also thank other faculty members, who have provided assistance in various aspects of the research, namely Dr. Robin Murphy, Dr. Phil Reed, Dr. Simon Williams and Ms. Sarah White.

I am indebted to the following organisations and their staff for their cooperation and support:

- Autism Resource Center (Singapore)
- Grace Orchard School (Singapore)
- Helen Alison School (UK)
- Metta School (Singapore)
- National Autistic Society (UK)
- Pathlight School (Singapore)
- Rainbow Center (Singapore)

This thesis was made possible through the generous sponsorship of the Ministry of Education (Singapore). I express my deep appreciation to officers from the Staff Training Branch and Education Programmes Division (EPD) who have supported me throughout the duration of my studies; and in particular, to the current and previous directors of EPD, Ms. Sum Chee Wah, Ms. Ho Peng and Mrs. Mok Choon Hoe, for their leadership and guidance.

Lastly, but in no sense the least, I am thankful to my family and friends for their never fading encouragement and support. Special thanks go to Zarinah Mohamed and Laura Cockburn for their invaluable assistance and support throughout my studies.

## ABSTRACT

In Singapore, there is a high reliance on IQ scores as the basis for deciding children's access to special educational provisions. Children with disabilities remain in the mainstream if they are perceived to be able to cope with the demands of the mainstream schools. On the other hand, if children were seen to require intensive support, referral to special schools would be initiated. This thesis aims to evaluate the validity of measures of intelligence and other selected indicators of special educational needs (SEN) for children with autism in Singapore.

The first phase of the thesis involved identifying an independent measure of SEN. Results of Study 1, which involved interviews with the parents of 40 children with autism, provided support for the International Classification of Functioning Disability and Health (ICF: WHO, 2001) as an adequate independent measure of SEN.

The second phase involved the evaluation of selected indicators of SEN that can be used alongside the ICF, namely measures of intelligence, theory of mind, executive function, central coherence and cognitive modifiability. These were evaluated based on their psychometric and treatment validity, as defined in educational contexts. For evaluations of psychometric validity, two criteria were used: firstly, the extent to which the indicators were able to predict children's SEN level; and secondly, the extent to which the indicators were able to distinguish children with autism who can cope with mainstream schools, from those that require special schools. This involved individual assessments with 52 children with autism and interviews with their parents (Study 2). For evaluations of treatment validity a qualitative approach was adopted to obtain practitioners' views on the extent to which the indicators of SEN were able to provide information that can be used to plan interventions (Study 3).

The findings indicated that it was the combination of indicators that accounted for the greatest variance in the SEN levels of children with autism. However, depending on the purpose of testing and types of sub-group of children with

autism, different indicators proved to have different validity strength. When the treatment validity of these measures was evaluated, measures of theory of mind showed the strongest treatment validity. The findings are discussed in terms of their implications for SEN assessments in Singapore, and the assessment of children with autism in general.

## TABLE OF CONTENTS

ACKNOWLEDGEMENT	iii
ABSTRACT	iv
LIST OF TABLES	xii
LIST OF FIGURES	xv
<b>CHAPTER 1: INTRODUCTION</b>	
1.1 CONTEXT	1
1.2 OVERVIEW	5
<b>CHAPTER 2: CONCEPTUAL FRAMEWORK FOR SPECIAL EDUCATIONAL NEEDS</b>	
2.1 CONCEPTUAL FRAMEWORK FOR SPECIAL EDUCATIONAL NEEDS	7
2.1.1 Focus on individual differences	7
2.1.2 Focus on environmental demands	8
2.1.3 Interactional analysis of SEN	9
2.2 METHODOLOGY FOR EVALUATING THE VALIDITY AND RELIABILITY OF SEN	12
<b>ASSESSMENTS</b>	
2.2.1 Definition of validity	13
2.2.2 Types of validity evidence as described in the <i>Standards</i>	14
2.2.3 Measurements of reliability	20
2.2.4 Summary	22
2.3 REVIEW OF THE INTERNATIONAL CLASSIFICATION OF FUNCTIONING, DISABILITY AND HEALTH	23
2.3.1 Background and Rationale	23
2.3.2 Description of the ICF	24
2.3.3 Adequacy of the ICF model	27
2.3.4 Feasibility	29
2.3.5 Applicability for specific populations	29
2.3.6 ICF checklist	29
2.3.7 Reliability and validity of the ICF checklist	30
2.3.8 Critique of the ICF	32
2.3.9 Summary and conclusion	38
2.4 CONCEPTS AND THEORIES OF AUTISM	41
2.4.1 Concept of autism	41
2.4.2 Theories of autism	46
2.4.3 Cognitive modifiability in children with autism	53
2.4.4 Conclusion	59
2.5 ASSESSMENT APPROACHES FOR CHILDREN WITH AUTISM	60

2.5.1	Behaviour observation scales	61
2.5.2	Parental reports	61
2.5.3	Investigator-based interviews	63
2.5.4	Comparison of the ADI and the DISCO	66
2.5.5	Summary on investigator-based interviews	70
2.6	SUMMARY AND CONCLUSION FOR CHAPTER 2	71

<b>CHAPTER 3: EVIDENCE BASED ON CONTENT AND RESPONSE PROCESSES FOR THE INVESTIGATOR-BASED INTERVIEW IN PARENTS AND HIGH-PERFORMING CHILDREN WITH AUTISM</b>		73
---	--	----

3.1	INTRODUCTION	73
3.1.1	Evidence based on content	75
3.1.2	Evidence based on response processes	76
3.1.4	Evidence of reliability	78
3.1.5	Evidence based on external criteria	78
3.1.6	Evidence based on treatment validity	79
3.1.7	Objectives of present study	79
3.2	METHODOLOGY	81
3.2.1	Participants	81
3.2.2	Measures	83
3.3.3	Procedures	86
3.3	RESULTS	87
3.3.1	Results for the first criterion	87
3.3.2	Results for the second criterion	90
3.3.3	Results for the third criterion	92
3.3.4	Summary of results	94
3.4	DISCUSSIONS	95
3.4.1	Evaluation of the first criterion	95
3.4.2	Evaluation of the second criterion	96
3.4.3	Evaluation of the third criterion	96
3.4.4	Methodological issues	97
3.4.5	Conclusion	99

<b>CHAPTER 4: ASSESSMENT OF INTELLIGENCE AND EXECUTIVE FUNCTION IN HIGH-PERFORMING CHILDREN WITH AUTISM</b>		100
---	--	-----

4.1	MEASURE OF INTELLIGENCE	100
4.1.1	What is intelligence; how is it measured?	101
4.1.2	Measurement issues	102
4.1.3	Description and review of the WISC-III (Singapore)	106
4.2	MEASURE OF EXECUTIVE FUNCTION	111

4.2.1	What is executive function; how is it measured?	111
4.2.2	Measurement issues	112
4.2.3	Review of the BADS-C	115
4.3	MEASURE OF CENTRAL COHERENCE	121
4.3.1	What is central coherence; how is it measured?	121
4.3.2	Measurement issues	123
4.3.3	Review of the Children's Embedded Figures Test (CEFT)	124
4.4	MEASURE OF COGNITIVE MODIFIABILITY	127
4.4.1	What is cognitive modifiability; how is it measured?	127
4.4.2	Review of dynamic assessments	128
4.4.3	Review of the Cognitive Modifiability Battery (CMB)	134
4.5	MEASURE OF THEORY OF MIND ABILITIES	141
4.5.1	What is theory of mind; how is it measured?	141
4.5.2	Measurement issues	143
4.5.3	Theory of Mind Battery	144
4.6	SUMMARY AND CONCLUSION FOR CHAPTER 4	152

5.1	INTRODUCTION	153
5.2	CRITERION VALIDITY	153
5.3	PARTICIPANTS	154
5.4	MEASURES	155
5.4.1	Criterion measure for SEN level	155
5.4.2	Criterion for ability to cope with demands of mainstream school	158
5.4.3	Predictor variables (Indicators of SEN)	160
5.5	PROCEDURE	162
5.5.1	Individual testing	162
5.5.2	Parent interview	163
5.5.3	Feedback	163
5.6	RESULTS	164
5.6.1	Preliminary analyses	164
5.6.2	First criterion: Concurrence with an independent measure of SEN, i.e. ICF	165
5.6.3	Second criterion: Accuracy in distinguishing children who can cope with mainstream with mainstream, from those that need special schools.	168
5.6.4	Summary of results	172
5.7	DISCUSSION	173
5.7.1	Evaluation of the first criterion	173

5.7.2	Evaluation of the second criterion	174
5.7.3	Methodological issues	176
5.8	CONCLUSION	178
<b>CHAPTER 6: TREATMENT VALIDITY OF THE INDICATORS OF SPECIAL NEEDS FOR CHILDREN WITH AUTISM</b>		179
6.1	INTRODUCTION	179
6.2	TREATMENT VALIDITY	180
6.2.1	Treatment validity for the measure of intelligence	181
6.2.2	Treatment validity for the measure of cognitive modifiability	183
6.2.3	Treatment validity for the measure of executive function	185
6.2.4	Treatment validity for the measure of theory of mind	186
6.2.5	Summary	188
6.3	METHODS	190
6.3.1	Objectives of Study 3	190
6.3.2	Structure of focus groups	191
6.3.3	Field trials	192
6.3.4	Participants	192
6.3.5	Procedure	192
6.3.6	Data analysis	193
6.4	RESULTS	195
6.4.1	Frequency analysis of coded units	195
6.4.2	Treatment validity for the measure of intelligence (WISC-III)	196
6.4.3	Treatment validity for the measure of cognitive modifiability (CMB)	201
6.4.4	Treatment validity for the measure of executive function (BADS-C)	206
6.4.5	Treatment validity for the measure of theory of mind (ToM Battery)	210
6.5	DISCUSSION	214
6.5.1	Summary of FGDs	214
6.5.2	Comparison of the treatment validity of the four tests	216
6.5.3	Factors affecting evaluation of treatment validity	216
6.6	CONCLUSION	218
<b>CHAPTER 7: FOLLOW-UP FROM STUDY 2</b>		219
7.1	FOLLOW-UP FROM RESULTS OF STUDY 2	219
7.2	RESULTS	221
7.3	DISCUSSION	221
7.3.1	Focus on specific core impairments in autism	223
7.3.2	Interactional analysis of SEN	223
7.3.3	Methodological factors	224

<b>CHAPTER 8: CONCLUSION</b>	<b>225</b>
8.1 VALIDITY ARGUMENTS FOR MEASURES OF SEN	225
8.1.1 Validity arguments for the use of the ICF	227
8.1.2 Validity arguments for the measure of intelligence	228
8.1.3 Validity arguments for the measure of cognitive modifiability	229
8.1.4 Validity arguments for the measure of executive function	230
8.1.5 Validity arguments for the measure of theory of mind	231
8.2 METHODOLOGICAL AND EPISTEMOLOGICAL ISSUES	231
8.3 IMPLICATIONS FOR RESEARCH AND PRACTICE	233
8.3.1 Implications for the SEN assessments practice and policy in Singapore	233
8.3.2 Broader implications for the assessment of and research on children with autism	234
8.4 REVISITING THE INTERACTIONAL PERSPECTIVE OF SEN	236
8.5 CONCLUSION	239
<b>APPENDIX A ICF CHECKLIST</b>	<b>241</b>
<b>APPENDIX B DSM-IV AND ICD-10 DIAGNOSTIC CRITERIA FOR AUTISM</b>	<b>252</b>
<b>APPENDIX C INVESTIGATOR-BASED INTERVIEW PROTOCOL USED IN CONJUNCTION WITH THE ICF CHECKLIST</b>	<b>255</b>
<b>APPENDIX D STUDY 1: SAMPLE INVITATION LETTER AND PARENTAL INFORMED CONSENT</b>	<b>299</b>
<b>APPENDIX E STUDY 1: TRAINING FOR ICF INTER-RATER STUDY</b>	<b>302</b>
<b>APPENDIX F STUDY 1: KAPPA VALUES (INTER-RATER AGREEMENT) FOR INDIVIDUAL ICF ITEMS</b>	<b>304</b>
<b>APPENDIX G KEY FEATURES OF THE BEHAVIOURAL ASSESSMENT OF DYSEXECUTIVE SYNDROME FOR CHILDREN (BADS-C) SUBTESTS</b>	<b>307</b>
<b>APPENDIX H FIELD-TEST OF THE BEHAVIOURAL ASSESSMENT FOR DYSEXECUTIVE SYNDROME FOR CHILDREN (BADS-C)</b>	<b>309</b>
<b>APPENDIX I FIELD-TEST OF THE CHILDREN'S EMBEDDED FIGURES TEST (CEFT)</b>	<b>314</b>
<b>APPENDIX J DESCRIPTION OF THE SIX SUBTESTS IN THE COGNITIVE MODIFIABILITY BATTERY (CMB)</b>	<b>316</b>
<b>APPENDIX K FIELD-TEST OF THE COGNITIVE MODIFIABILITY BATTERY (CMB) ANALOGIES SUBTEST</b>	<b>317</b>
<b>APPENDIX L PRELIMINARY STUDY FOR THE SELECTION OF ITEMS FOR THE THEORY OF MIND (TOM) BATTERY</b>	<b>320</b>
<b>APPENDIX M DESCRIPTION OF THE THEORY OF MIND BATTERY</b>	<b>326</b>

<b>APPENDIX N</b>	<b>STUDY 2: SAMPLE OF THE INVITATION LETTER AND INFORMED CONSENT FORMS</b>	<b>345</b>
<b>APPENDIX O</b>	<b>STUDY 3: POWER-POINT SLIDES (STIMULUS PRESENTATION) USED IN FOCUS GROUP</b>	<b>348</b>
<b>APPENDIX P</b>	<b>STUDY 3: MODERATORS' GUIDE FOR FOCUS GROUP DISCUSSIONS</b>	<b>352</b>
<b>APPENDIX Q</b>	<b>STUDY 3: SAMPLE OF INFORMED CONSENT FORM FOR PARTICIPANTS OF FOCUS GROUPS</b>	<b>353</b>
<b>APPENDIX R</b>	<b>STUDY 3: OPERATIONAL DEFINITIONS FOR ALL CATEGORY HEADINGS</b>	<b>354</b>
<b>REFERENCES</b>		<b>359</b>

## LIST OF TABLES

### TABLE

2.2.2	Framework for Validity Evidence in the Standards	15
2.3.2.1	International Classification of Functioning, Disability and Health Definitions	25
2.3.2.2	International Classification of Functioning, Disability and Health List of Human Functioning – Functional Domains in Each Component	26
2.3.6.	ICF Checklist Structure	30
2.3.7	ICF Field Trial Results – Kappa Values (WHO, 2000, Slides #43-47)	31
2.3.8.1	Comparison of the Domains Included in the ICF, Vineland and DISCO	35
2.3.8.2	Summary of the Advantages, Limitations and Unknown Aspects of the ICF, Vineland and DISCO.	38
2.4.1.1	DSM-IV Diagnostic Criteria for Autism, Rett Disorder and Asperger's Syndrome	42
2.5.3.1	Comparison of a) GARS, b) ADI & c)DISCO Items on the Use of Gestures	65
3.1.1	The ICF Functional Domains	75
3.2.1	Profile of Participants (Study 1)	82
3.2.2.1	Calibration of Anchor Points for ICF Ratings Scales	85
3.3.1.1.1	Items Excluded from Reliability Analysis	88
3.3.1.1.2	Internal Reliability (Cronbach Alpha) of ICF Component & Composite Scores	89
3.3.1.2.1	Correlation (Spearman's rho) Between Ratings of First and Second raters	89
3.3.1.2.2	Number and Percentages Of ICF Items with Adequate to Excellent Agreement	90
3.3.2	ICF Score – Group Means	91
3.3.3.1	Correlation Matrix (Pearson's r) between ICF Functioning & Disability and Total DISCO Scores by Groups	93
3.3.3.2	Levene's Test for Equality of Variance	94
4.1.1.2	Measures of Intelligence for Children	103
4.1.2.2	Comparison of the WISC-III Scores of a Singaporean Child Based on Singapore and UK Norms.	105
4.1.3	Subtests in the WISC-III (Singapore)	107
4.2.2	Standardised test for Executive Function in Children	114
4.2.3.1	Description of the Six Part Test in the BADS-C	115
4.2.3.2.1	Executive Function Processes in Each Subtests (Adapted from BADS-C Manual, pp 24)	116
4.2.3.2.2	Key Denominators of Executive Function in Each of the 6 Subtests	117
4.4.1	Dimensions of Dynamic Assessment Approaches	128
4.4.2	Key Dimensions of a Selection of Dynamic Assessments	130
4.4.2.3	Sample of Probes Used for the Rhyming Task in the S-CPT	132
4.4.3.3	Mediation Processes and Strategies for the CMB Analogies Subtest	135

4.5.3.1.1	Test Items in White et al.'s Study	144
4.5.3.1.2	Scoring Criteria for Strange Stories Based on Graded Justification	150
4.5.3.2	Structure of ToM Battery Used in the Present Thesis	151
4.6	Selected Measures for the Indicators of SEN	152
5.3	Profile of Participants (Study 2)	154
5.4.1.1	Concurrence Between ICF Domains and SEN Indicators	157
5.4.3	Selected Measures for the Indicators Of SEN	161
5.5.1	Sample Testing Format and Average Testing Time Per Child	163
5.6.1.1	Descriptive Statistics	164
5.6.1.2	Correlation Matrix (Pearson's r) between Age and Un-Normed Predictor Variables	165
5.6.2.1	Correlation Matrix (Pearson's r) for ICF SEN Ratings and All Predictor Variables (i.e. Indicators of SEN)	166
5.6.3	Means (and S.D) Scores for All Predictor Variables (by School-Types)	168
5.6.3.1.1	Stepwise Discriminate Function Analyses of Predictor Variable on Children's Group by School Type	169
5.6.3.1.2	No. of Children Classified According to Actual and predicted Groups (Based on FSIQ, ToM & CMB)	170
5.6.3.2.1.	Group Contrasts between Mainstream and Special Schools (Mild	170
5.6.3.2.2	No. of Children Classified According to Actual Group (i.e. School Type) and Predicted Group (Based on ToM Scores)	171
5.6.3.3.1	Group Contrast Between Special Schools : Mild and Moderate	171
5.6.3.3.2	No. of Children Classified According to Actual and Predicted Groups (based on FSIQ)	171
6.3.4	Information About Participants of Focus Group Discussion (FGD) Teams	192
6.3.6.1	An example of the Operational Definition for Codes	185
6.4.2	Summary of Categories Elicited for WISC-III	197
6.4.2.1	Examples of Coded Units (Comments) for the WISC-III	198
6.4.2.2	Operational Definition for 'Improvement to Current Practice'	200
6.4.3	Summary of Categories Elicited for CMB	201
6.4.3.1	Examples of Coded Units (Comments) for the CMB	202
6.4.3.2	Perceptions on the Benefits and Difficulties of Dynamic Assessments	205
6.4.4	Summary of Categories Elicited for the BADS-C	207
6.4.4.1	Examples of Coded Units (Comments) for the BADS-C	208
6.4.5	Summary of Categories Elicited for ToM	211
6.4.5.1	Examples of Coded Units (Comments) for the ToM	212

6.5.2	Comparison of the Treatment validity of the Four Tests	221
7.2	Summary of Hierarchical Regression Analysis Based on Study 3	221
8.1	Summary of Findings	226
8.4	Functional Domains in the ICF Components	237
E1	Outline and Details for Inter-Rater Training (Study 1)	303
H1	Profile and Standard Scores for Individual Participants in BADS-C Field-Test	310
H2	Alpha Values for Overall Scaled Scores (BADS-C Field-Test)	311
H3	Individual Scores and Summary of Children's Executive Function (BADS-C Field-Test)	312
H4	Comparison of Mean Scaled Scores for Children with Autism and Controls (BADS-C Field-Test)	314
I1	Profile of Children in CEFT Field-Test	315
I2	Children's Individual CEFT Scores (CEFT Field-Test)	315
I3	Non-Parametric Correlation Between Language, Nonverbal IQ and CEFT Scores (Spearman's Rho) (CEFT Field-Test)	316
K1	Profile of Children in CMB Field-Test	317
K2	Individual Children's Results by Methods 1 & 2 Scoring Criteria (CMB Field-Test)	318
L1	Items in ToM Battery (Preliminary Study)	320
L2	Participants' Profile (ToM Preliminary Study)	321
L3	Participants' Performance in Individual Items in ToM Battery (Preliminary Study)	322
L4	Percentage of Participants in Each Score Range for Strange Stories (Preliminary Study)	322
L5	Internal Reliability of ToM Battery (Preliminary Study)	323
L6	Inter-Subtask Correlation in ToM Test Battery (Preliminary Study)	323
L7	Correlation Between ToM and Age; Nonverbal; and Verbal IQ (Preliminary Study)	324
L8	Items in ToM Battery Used in Present Thesis	325

## LIST OF FIGURES

### FIGURE

2.3.2	ICF Model of Functioning & Disability	24
2.4.2	Main Locus of Explanations in Autism Theories	47
3.1	ICF Framework for Functioning and Disability	73
3.3.3	Range of Scores for ICF Functioning and Disability by Groups	93
4.3.3.1	Sample Item from CEFT - 'House' Embedded in 'Rocking Horse	124
4.4.3.1	Stimulus Materials and Sample Items for CMB Analogies Subtest	135
4.5.3.1	Stimulus for Interpretive Diversity Test (Cow)	147
6.3.6	Structure for Coding Categories for Comments on Treatment Validity	194
8.4	ICF Model of Functioning & Disability	236

# CHAPTER 1

## INTRODUCTION

### 1.1 CONTEXT

In a recent press release, the Ministry of Education (MOE) of Singapore made the following announcements:

*The Ministry of Education (MOE) has undertaken a review of measures to cater to children with special needs, in both Special Education (SPED) schools and mainstream schools. We will introduce the following initiatives to raise the quality of education for children with special needs:*

*Initiatives for mainstream schools will help students with mild to moderate levels of special needs who can cope with the regular school curriculum to remain in the mainstream schools and do well. Early detection and appropriate intervention and support can help these students achieve successful outcomes in school and in life. The increased resources will enhance current intervention and support programmes in schools.*

*(Enhancing Support For Children With Special Needs, MOE Press Release, 18 Sep 2004)*

A key factor that will affect the success of these initiatives is the efficacy of the process of identifying ('early detection') children with special needs 'who can cope with regular school curriculum'. This is because the provision of effective intervention for children with special educational needs (SEN) depends on the processes and criteria for deciding children's level of needs. Various education systems have used different criteria for evaluating the severity of such needs in children and, consequently, in determining this access to specialised educational provisions (OECD, 2000).

In Singapore, there is a high reliance on IQ scores as the basis for deciding children's access to special educational provisions, as it is often seen as the single most useful indicator of children's learning ability (Lyen, 1997; Quah, 1993). Children with disabilities remain in mainstream if their needs are seen as not severe, and they are perceived to be able to cope with the demands of the mainstream school. On the other hand, if the children face difficulties in coping

and are seen to require intensive specialist support, a referral to special schools would be initiated. Thus, in the Singapore context, indicators of special educational needs have high-stake implications. A reliable and valid indicator of special educational needs is important in ensuring that children get the early intervention that is often recommended (Wahlberg, Obiakor, Burkhardt, & Rotatori, 2001).

The advantages and limitations of IQ assessments for special educational decisions have been the subject of considerable debate. On the one hand, studies have shown that when appropriately used, i.e. when used with reference to appropriate populations and for the purposes that they were intended for, standardised IQ tests have adequate reliability, criterion and concurrent validity (Anastasi, 1988; Jensen, 1980; Sattler, 1992). It has also been argued (Haywood, Brown, & Wingenfeld, 1990) that IQ tests do an adequate job of surveying large groups of children for educational planning (e.g. estimating the number of children requiring special class placements). For children with special educational needs, it has been argued that although educational placement based solely on IQ test scores might not be valid, IQ tests have shown adequate functional utility, i.e. IQ scores can be used as a predictor of future academic achievement (Flanagan, Andrews, & Genshaft, 1997). Support for the predictive validity of IQ tests comes from various studies which suggest that when social factors, such as family background, have been taken into account, IQ remains one important source of variation in school achievement (Mackintosh, 1998; Jenkins & Pany, 1978).

However, critics of IQ tests have pointed out that IQ scores do not appear to be related to instructional practice (Reschly, 1990). As most IQ tests do not assess the cognitive processes and abilities that are most closely associated with academic achievement, they therefore do not yield much information that can be used in the diagnosis and treatment of learning problems (McGrew & Flanagan, 1993; Siegel, 1989). It has also been argued that many intelligence tests were not developed from well researched theoretical frameworks (Mackintosh, 1998) and there is often an overemphasis on product rather than process in traditional IQ tests, i.e. an overemphasis on what has been learnt, rather than how the

information is learnt (Sternberg, 2000). It has also been claimed that IQ tests may be biased against culturally and linguistically diverse populations (Cunningham, 1986). Critics have asserted that because many of the items in IQ tests tap knowledge and information that emanates from western cultures, these tests are biased against children who are not adequately exposed to the same culture (this issue is discussed in greater detail in Section 4.1).

For children with autism, there appear to be particular problems in using IQ tests. For example, it could be argued that correlations linking IQ scores and academic attainment which were established for normal populations, cannot be assumed to be applicable to populations of individuals with autism. Implicit in the predictive correlation is the assumption that the predictor (i.e. IQ) is stable over time. Studies have shown that children with autism showed substantial changes over time in IQ test scores (Lord & Schopler, 1988; Mayes & Calhoun, 2003). In addition, it has been well established that children with autism often experience impairments in language and communication skills (Eisenmajer, Prior, Leekam, Wing, Ong, Gould, & Welham, 1998; Ricks & Wing, 1975) as well as social interaction skills (Cohen, Caparulo, Gold, Waldo, Shaywitz, & Rimland, 1977; Lord & Pickles, 1996). In many of the widely used IQ tests, e.g. the Wechsler Intelligence Scales for Children, 3<sup>rd</sup> Edition (WISC-III: Wechsler, 1991) and the British Ability Scales (BAS: Elliott, 1983), language requirements and the inclusion of tasks that require specific social knowledge make the tests less viable in representing the cognitive abilities of children with autism (Klin, Carter, Volkmar, Cohen, Marans, & Sparrows, 1995).

Given the issues of using IQ tests, it is worthwhile considering alternatives for assessing learning abilities in children with autism, and evaluating if they provide a better gauge of children's special educational needs than IQ scores. Two alternatives are considered, namely cognitive assessments based on the causal theories of autism, and dynamic assessments:

- Cognitive theories of autism have highlighted several aspects of cognition which underlie the patterns of impairments in children with autism (these theories are reviewed in Chapter 2). For example, the theory of mind hypothesis (Baron-Cohen, Leslie, & Frith, 1985) suggests that

mentalising abilities, or the ability to think about thoughts, is the core social deficit in autism. Consequently, it can be argued that for children with autism, the degree of impairment in mentalising abilities can be used as an indicator of special needs. One advantage of this assessment approach is that it is based on a clear theoretical framework on the possible causes of autism. However, a key practical issue is whether such assessments of specific cognitive skills, e.g. measures of theory of mind, provide better indicators of the child's special educational needs than traditional measures of intelligence or general abilities.

- It has been argued that most standardised tests of intelligence measure 'static' functioning, i.e. what has been learned as opposed to how easily or the manner in which something was learned (Sternberg, 2000). The approach used in traditional IQ tests can be contrasted with dynamic assessment methodology, which attempts to assess gains in performance on intellectual/cognitive tasks after the strategies associated with such tasks have been taught. Because children with autism frequently lack the skills necessary for effective adaptive functioning in specific contexts (Schopler & Mesibov, 1995), it may be more accurate to utilize assessment techniques that incorporate assessment of the ability to *gain* from training and not static intellectual / cognitive skills per se.

The primary aim of the present thesis is to identify valid and reliable methods of assessing the special educational needs for children with autism in Singapore. This will be carried out in two phases and involves comparing the evidence for the validity of several different measures or indicators of special educational needs:

Phase 1 : To identify a valid and reliable measure of special educational needs (SEN) for children with autism in Singapore.

Phase 2 : To identify possible indicators that reflect the learning deficits of children with autism (such as IQ scores, performance on specific cognitive tasks related to theories of autism and performance on dynamic assessment tasks); and evaluate their validity.

The outcomes of the thesis could serve as a basis for reviewing and refining the assessment paradigm currently used in Singapore for decisions about special educational needs. Results from the study will provide a useful guide for educational psychologists in their assessment of children with autism, in particular when making high-stake decisions about children's suitability for mainstream settings, and the need for specialist provisions.

## **1.2 OVERVIEW**

The next chapter, i.e. Chapter 2, will address the key concepts in the thesis, namely the concept of special educational needs (SEN), theories of autism and the assessment of children with autism. Research evidence and theories will be discussed and this review will form the basis for identifying the appropriate conceptual framework for SEN, the definition of autism and the assessment approaches that will be used in the present thesis. In addition, the methodologies and criteria for evaluating the validity of indicators of special educational needs will be discussed, with reference to established international standards for psychological and educational testing.

Chapter 3 presents the outcomes of the first phase of the thesis, i.e. the identification of a valid and reliable measure of special educational needs (SEN) of children with autism in Singapore. The findings of the first study, which evaluated the reliability and validity of the International Classification of Functioning Disability and Health (ICF) for children with autism, is reported.

In Chapter 4, the indicators of SEN for children with autism are discussed, namely measures of intelligence, executive function, theory of mind, central coherence and cognitive modifiability. Specific tests for each indicator are identified through a literature review, as well as a series of field-tests, which were conducted to ensure the applicability of the selected tests for the target group, i.e. children with autism.

In Chapter 5, Study 2, which evaluated the criterion validity of the selected indicators of SEN, is reported. In this study, two pieces of criterion-based evidence were sought: firstly, the extent to which these indicators predict

children's level of SEN; and secondly, the accuracy of the indicators in distinguishing children who can cope with mainstream schools, from those that need special schools.

Chapter 6 reports the findings of Study 3, which evaluated the treatment validity of the selected indicators of SEN. The results of this qualitative study will form the basis for identifying the indicators of SEN that demonstrate the strongest utility.

In Chapter 7 a follow-up analysis integrating the data from Studies 2 and 3 is reported.

Finally, Chapter 8 discusses the key findings of the studies in relation to the primary aim of the thesis, i.e. to identify valid measures for assessing the special educational needs of children with autism. The practical application of the key findings of the thesis is considered in relation to the assessment approaches currently used in Singapore for decisions about the special educational needs for children with autism.

## **CHAPTER 2**

### **ASSESSMENT OF SPECIAL EDUCATIONAL NEEDS**

In this chapter, the key concepts in the thesis are reviewed, with a view to identify the appropriate conceptual framework for special educational needs (SEN). In addition, the key criteria for evaluating the indicators of SEN, the definition of autism and the assessment approaches that will be used in the present thesis will be discussed.

#### **2.1 CONCEPTUAL FRAMEWORK FOR SPECIAL EDUCATIONAL NEEDS**

In this section, three different approaches to the concept of special educational needs (SEN) will be reviewed: 1) a focus on individual differences; 2) a focus on environmental demands; and 3) an interactional analysis of SEN. Each is distinguishable by assumptions made about human development; and arising from each are parallel assessment approaches for the identification and evaluation of children with SEN (Frederickson & Cline, 2002). It will be argued that each of the different conceptual frameworks of SEN, despite their limitations, has some advantages. Thus, all three approaches continue to influence how SEN is defined and conceived in practice.

##### **2.1.1. Focus on individual differences**

In this approach, the focus of causation is ‘within the child’. A child with special educational needs is defined as a child who has a disability of body or mind. This approach is clearly reflected in several key classification frameworks of disabilities, e.g. the DSM-IV (APA, 1994) and the ICD-10 (WHO, 1995), where each category of health condition is defined in terms of specific characteristics of the individual person. These characteristics may be biological, behavioural, cognitive or emotional. Assessment approaches that have arisen from the ‘within-child’ perspective of SEN have focused on identifying specific characteristics that distinguish a child from

other 'normal' children. The use of standardised IQ tests is a notable example of such an approach.

The focus on individual differences has shown its practical utility in many ways. For example, the use of normative reference as a criteria for defining levels of functioning is quantifiable; and this gives the approach a sense of objectivity. This is especially important in the assessment of specific domains of functioning, which can only be made by inference and rely heavily on professional judgment, such as the assessment of cognitive and social-emotional functioning. However, a major criticism of the focus on individual differences is the inherent assumption that deficits in a child's functioning reflect either a gap in learning capacity, or faulty development. It can be argued that for this assertion to hold true, assumptions must be made about the appropriateness of learning opportunities and experiences, the effectiveness of teaching, and the extent to which the child's prior learning experiences are comparable to his/her peers. It is also widely recognised that social and educational contexts are important determinants of children's attainment and development (UNESCO, 2002).

### **2.1.2. Focus on environmental demands**

In this approach, children's special educational needs (SEN) are assumed to be the outcome of inadequate or inappropriate opportunities for teaching and learning. Arising from the environmentally focused perspective, assessments of SEN would typically involve detailed analyses of the situational factors that had given rise to and maintained the child's difficulties. For example, in children who are seen to have low cognitive attainments, the focus of assessments would be to identify the extent to which the child had been appropriately taught the basic or pre-requisite skills needed to perform the cognitive tasks.

There are significant limitations to the environmentally focused approach. Environmental factors cannot fully account for the variability of individual differences that are observed among children within the same learning or social

context. For example, in UNESCO's (2002) meta-study of the impact of environmental factors on pupils' achievement in different countries, it was observed that the relative impact of contextual and school related factors vary among developing and developed nations. In the industrial nations of Europe and North America, consistent negative correlations have been found between family size and educational achievement, while similar studies in Kenya and the United Republic of Tanzania have found the correlation to be positive. In contrast, the availability of teaching materials is strongly correlated with student performance in developing countries but not in developed countries. Additional studies dealing with determinants of pupil achievement in developed countries concluded that many school factors, such as class-size, school facilities, per pupil expenditure, instructional time and innovation in instruction methods, did not contribute substantially to gains in scholastic achievement. However, these findings should not be interpreted as indicating that schools do not make a difference, but rather, that above a certain threshold of resources, where perhaps most developed countries are located, other factors may have contributed more to differential learning outcomes, i.e. factors that are relevant to individual variation in pupils beyond differences in school/contextual factors.

### **2.1.3 Interactional analysis of SEN**

As described by Frederickson and Cline (2002), an interactional analysis of SEN encompasses the view that a child's special needs are an interaction between the child's strengths and weaknesses, the level of support available for the child, and the appropriateness of the education provided. As a paradigm for the conceptualisation of special educational needs, the interactional approach has some key advantages. It presents a multi-dimensional analysis of learning and development (namely the biological, cognitive, environmental and behavioural dimensions), and posits the view that each dimension interacts with another. As such, this avoids the partisan/dichotomous view of special educational needs as being either determined by 'within-child' or 'environmental' factors.

Following from the interactionist view, the assessment process becomes an integral part of intervention; and intervention, in turn, informs the assessment process. This interactional feature is characteristic of dynamic assessment approaches, in which the aim is to ascertain the child's learning capabilities when given enhanced instruction (Rothman & Semmel, 1990). This assessment approach can be contrasted with norm referenced tests, which aim to identify the level that a child has achieved without external or additional help. This integration of both quantitative data about children's functioning and qualitative observations of children's responses to teaching offers a possible improvement over the existing approaches for identification and assessment of children with special educational needs (Campione, 1989).

In recent years, several theoretical frameworks for understanding special educational needs that reflect the interactional perspective have emerged. Two examples will be discussed in this section, namely the cognitive framework by Dockrell and McShane (1993), and the International Classification of Functioning, Disabilities and Health (WHO, 2001).

Dockrell and McShane (1993) proposed a cognitive framework for understanding learning difficulties that includes the interaction of learning tasks, the child and the environment. The assessment approach proposed in Dockrell and McShane's model involves the use of both norm and criterion referenced tests, combined in the context of a teaching experiment which "allows the practitioner to incorporate the task, child and environment into the conceptualisation of a learning difficulty" (pp. 38). One implication of Dockrell and McShane's framework is that due to its emphasis on cognitive processes, information gathered about the child has to be interpreted with reference to an implicit model of cognition that is applicable to the various domains of skills. The problem is that there is little agreement with regards to the specific cognitive processes that are invoked during functional skills, such as arithmetic and social-interaction skills. There is even less agreement about the developmental paths of these cognitive processes (Lee, 1997). Using this framework, the analysis of the

task and the child's functioning will be constrained by the particular theory of cognition that the practitioner holds as a reference.

Another example of the interactional framework is the International Classification of Functioning, Disabilities and Health, or ICF (WHO, 2001) which was "...developed in recognition of the inadequacy of a purely medical classification of diseases to provide information for planning purposes, as diagnosis alone does not predict needs or outcomes" (WHO, 2001, pp. 3). The model conceives disability and functioning as outcomes of the interaction between health conditions and contextual factors. This model will be reviewed in detail in Section 2.3. However, it is noteworthy that even in a medical classification, namely the WHO, where traditionally the concept of disability has been led by a 'within-child' paradigm, there is a shift to encapsulate an understanding of functioning that is anchored within an interactional model, or a 'biopsychosocial' model as it is termed in the ICF (WHO, 2001, pp 20).

Most of the assessment frameworks derived from the interactional approach are relatively new and as such, their utility may not be fully realised. Initial apprehension about their use is that they are not easily translated into practice. In addition, since intervention and assessment are viewed as two interlocking processes, the quality and accuracy of assessment becomes contingent on the quality of intervention that the child receives. In one sense, this conceptualisation is circular, as the extent to which the child is identified as having special educational needs may depend on the availability and effectiveness of the special needs provision that has been provided.

However, given the balance of arguments supporting each of the theoretical approaches, it can be concluded that the interactional perspective most closely reflects current knowledge about the factors underlying special education needs, i.e. that it is the effect of both within child and environmental factors. Thus, in the present study, this interactional framework will be used as the basis for defining special educational needs.

Although several assessment frameworks have arisen from the interactional perspective, few have extended their application to the development of assessment instruments. The notable exception is the ICF (WHO, 2001), in which both the theoretical framework and a parallel assessment tool have been developed. In section 2.3 of this chapter, the ICF will be discussed in greater detail. A key issue is whether the ICF assessment tool, i.e. the ICF checklist, is an adequate measure of special educational needs for children with autism.

For the purpose of the present thesis, a sound theoretical framework *per se* is an insufficient basis to draw conclusions about ICF's suitability as an adequate measure of SEN. In addition to strengths in the underlying theoretical framework, measures of SEN must also demonstrate technical qualities that meet acceptable standards for psychological and educational assessment. This is especially important in the contexts of high stakes educational decisions, such as determining children's access to appropriate educational provisions. In the next section, an established international standards for psychological and educational assessments is reviewed, namely the Standards for Educational and Psychological Testing (APA, AERA & NCME, 1999), with a view to identify the appropriate evaluation criteria and methodologies that are relevant for the present thesis.

## **2.2 METHODOLOGY FOR EVALUATING THE VALIDITY AND RELIABILITY OF SEN ASSESSMENTS**

A key issue in the present thesis is the validity of assessments for special educational needs (SEN). In this section, the concept of validity in psychological and educational assessments is discussed, with a view to identify the methodologies and criteria that can be used in the evaluation of indicators of SEN.

### 2.2.1 Definition of Validity

Traditionally, validity has been categorised into three different ‘types’, namely content, criterion related and construct validity (AERA, APA, & NCME, 1966). However, as discussed by Goodwin (1997), the definition of validity in psychological measurements has changed over the last 50 years, and its evolution has led psychometricians and measurement experts to reach general consensus on a definition:

*Validity is an overall evaluative judgement, found on empirical evidence and theoretical rationales, of the adequacy and appropriateness of inferences and actions based on test scores (Messick, 1988, pp33).*

A key element in the recent definition of validity is the shift in emphasis from the static, technical properties of a test, to that of test use. By this definition, it would be meaningless to label a test as being ‘valid’ or ‘invalid’, but rather the process of validation involves the gathering of evidence that would support the interpretations of test score.

Given that the interpretations of test scores are embedded in the context of their uses, evidence for validity is similarly context dependent. Hence decisions regarding the methodologies used by researchers to develop validity arguments for particular measures and the adequacy of the evidence accumulated must be guided by the intended uses or contexts of assessments.

The Standards for Educational and Psychological Testing, or *Standards* (American Psychological Association [APA], American Educational Research Associate [AERA] & National Council on Measurement in Education [NCME], 1999) identified different types of evidence for validity and these are reviewed in the next section.

## **2.2.2 Types of validity evidence as described in the *Standards***

The *Standards* (APA, AERA & NCME, 1999) defined validity as “the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of the test”. In line with the definition used by Messick (1988), which was described earlier, the *Standards* stressed that “it is the interpretations of the test scores by the proposed uses that are validated, not the test itself” (APA, AERA & NCME, 1999, pp 9). As such, when test scores are used or interpreted in more than one way, each intended interpretations must be validated. The *Standards* described five distinct types of validity evidence:

- Evidence based on test content;
- Evidence based on response processes;
- Evidence based on internal structure;
- Evidence based on relations to other variables; and
- Evidence based on consequences of testing.

Table 2.2.2 presents examples of validation activities that may provide the data for each types of validity evidence, as described in the *Standards*.

**Table 2.2.2 Framework for Validity Evidence in the *Standards***

Type of Validity Evidence	Methods	Examples of Validation Activities
Evidence based on test content	Analyses of content validity	<ol style="list-style-type: none"> <li>1. Logical analyses of the extent to which the test content represents the targeted domain.</li> <li>2. Experts' evaluations of the extent to which the items or subparts of a test match the definition of the construct and/or purposes of the test.</li> </ol>
Evidence based on response processes	Analyses of response bias/inconsistencies	<ol style="list-style-type: none"> <li>1. Analyses of individual responses.</li> <li>2. Process studies, examining similarities and differences in responses given by members of distinct subgroups of test-takers.</li> <li>3. Studies on the ways in which judges, observers and interviewers collect, record and interpret data.</li> </ol>
Evidence based on internal structure	Analyses of internal construct.	<ol style="list-style-type: none"> <li>1. Analyses of item inter-relationship.</li> <li>2. Factor analytical studies.</li> </ol>
Evidence based on relations to other variables	Analyses of criterion validity	<ol style="list-style-type: none"> <li>1. Correlational studies on the type and extent of scores and external variables.</li> <li>2. Correlational studies of the extent to which scores forecast or predict criterion performance or scores on measures obtained at a later date.</li> <li>3. Convergent validity studies, investigating the relationship between scores and other test intended to measure similar constructs.</li> <li>4. Divergent validity studies, investigating the relationship between scores and other measures of purportedly different constructs.</li> <li>5. Known-group comparison studies, intended to test hypotheses about expected differences in test scores across specific groups of examinees.</li> <li>6. Studies of the effectiveness of selection, placement and classification decisions.</li> </ol>
Evidence based on consequences of testing	Analyses of consequential validity (or treatment validity)	<ol style="list-style-type: none"> <li>1. Descriptive studies of the extent to which anticipated benefits of testing are realised.</li> <li>2. Descriptive studies of the extent to which unanticipated negative consequences occur.</li> </ol>

#### **2.2.2.1 Evidence based on test content**

Analysis of content validity involves the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain

measured. However, this process goes beyond a mere inspection of the test content, and addresses a deeper issue determining whether the universe of items relevant to the construct of interest is adequately sampled (Kline, 2000; Anastasi, 1988, Anastasi & Urbina 1996). Logical analyses and experts' evaluation of the components of the measure are the key ways content validity evidence is obtained. The strength of the analyses is strongly influenced by the knowledge and experience of the experts, as well as the theoretical framework that is used as a reference for the construct of interest. For example, in Section 2.1, several theoretical models of special educational needs were discussed, namely within-child, environmental and interactional approaches. As discussed, the different models place different emphases on the impact of the quality of intervention (i.e. environmental factors) in determining children's learning. Consequently, in the assessment measures associated with each model, the extent to which items sampling activities/behaviours related to the quality of the teaching/environmental input are included varied significantly.

#### *2.2.2.2 Evidence based on response processes*

This process involves obtaining evidence addressing the question, “To what extent does the type of response in which test-takers engage fit the intended construct”. For example, if developers of a parents' checklist for autism claim that the measure diagnoses autism behaviour in young children, evidence that respondents (i.e. parents of children with autism) are not merely providing socially desirable reports of behaviour would be relevant. For measures that involve direct interview or assessments with children with autism, a key consideration is the extent to which the language and task demands are appropriate, i.e. they accurately reflect the intended construct rather than social and communication characteristics that are extraneous to the intent of the measurement. In addition, given that the focus is children in a specific cultural context, namely Singapore, analyses pertaining to possible cultural bias in children's responses are critical.

### *2.2.2.3 Evidence based on internal structure/construct*

This type of evidence generally answers the question, “To what extent do the relationships among test items and components match the construct as operationally defined?” In the *Standards*, factor analysis is highlighted as the most closely associated activity for this type of validity evidence. However, it could be argued that while confirmatory factor analyses can provide some information on internal structure, it is merely one type of information. Over reliance on factor analyses can lead to a narrow body of empirical support for validity arguments, and as pointed out by critics of this approach, the “internal or factorial validity still needs evidence of a relationship to life events outside the tests themselves if the factors are to have substance,...or educational utility” (Thorndike, 1997, pp159).

It can be argued that, on its own, factor analysis is inadequate for evaluating construct validity. In the *Standards*, evidence for construct validity is not seen in terms of specific psychometric or statistical data, but rather in terms of the strength of *validity arguments* for the interpretations of test-scores:

*A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretations of test scores for specific uses (APA, AERA & NCME, 1999, pp17).*

However, the *Standards* did not provide any guidelines about the specific processes regarding validity argument, and many difficult questions remain, for example:

- How much evidence needs to be accumulated to ensure adequate support for the intended test interpretations?
- To what extent can the findings from one validity study be generalised to other contexts, such as other populations, and purposes of measurements.

#### *2.2.2.4 Evidence based on relations to other variables*

This type of evidence (also known as ‘criterion validity’) involves comparison of the individuals’ performance in a test with performance on another, independent activities. The criterion measure against which the individual’s test scores is compared may be obtained at approximately the same time as the test (i.e. concurrent validation), or after a stated interval (i.e. predictive validity). However, as argued by Anastasi and Urbina (1996), the logical distinction between concurrent and predictive validation is based not on time but on the purpose of testing. Concurrent validation is relevant to tests employed for diagnosis/measuring of existing status, rather than predicting future outcomes. Due to feasibility constraints, this thesis is focused on the assessment of children who have been diagnosed with autism, and for whom decisions about their special educational needs have been made *a priori*. Thus, in this context, the use of concurrent validation criteria would be more appropriate.

A key issue in concurrent validation is the choice of criterion measure. At any one time, there may well be several criterion measures that could be employed, but as a guiding principle commentators agree that at least one criterion should be used for each intended purpose of assessment (Anastasi & Urbina, 1996; Goodwin, 2002). In the context of assessment of special educational needs in Singapore, the purpose of assessment is two-fold: firstly to obtain accurate information on children’s functioning and needs level; and secondly, to obtain information that could be used to decide the appropriate educational placements for the child. Thus, in the present thesis, evidence for criterion validation will be based on these two intended outcomes of SEN assessment.

#### *2.2.2.5 Evidence based on consequences of testing*

This type of evidence is a recent addition to the 1999 version of the *Standards*, and address questions such as, “To what extent are anticipated benefits of the measurement realised?” and “To what extent do the unanticipated effects (negative and positive) occur?”. This notion of ‘consequential validity’ or ‘treatment validity’

refers to the degree to which any assessment procedure contributes to beneficial outcomes for individuals (Cone, 1989). More specifically, Gresham (2002) identified three key components of treatment validity:

1. Treatment utility, i.e. the extent to which there is evidence for test use, particularly as it relates to the social consequences and utility;
2. Cost-benefit analysis, i.e. evaluation of the features or properties of the test that could either facilitate or hinder its use;
3. Incremental validity, i.e. the extent to which an assessment improves existing procedures.

The concept of treatment validity appears to have been more frequently used in the field of clinical and educational assessments, rather than psychometrics. There appears to be an on-going debate and reluctance among psychometricians regarding the use and application of the concept in the context of test evaluation. While some commentators, e.g. Messick (1989), argued strongly that consequential/treatment validity is directly relevant to the evaluation of test use, critics point out that the inclusion would push the study of validation processes beyond the traditional psychometric boundaries into policy issues (Goodwin, 2002). In addressing the aspect of consequential/treatment validity, the *Standards* acknowledge the need to distinguish between issues of validity and social policy. However, no information is provided to guide test users in making that distinction. Instead, the *Standards* stressed the need to identify evidence that supports the claims made regarding the benefits of using the tests, as well as evaluation of its unintended consequences.

Although theoretically, evaluations of the intended and unintended consequences of measurement could include formal empirical studies, such as comparing the stress or anxiety of measured and unmeasured subjects, or the quality of interventions received by children based on different assessment methods, it is often more tenable to use qualitative or descriptive methods for this purpose (See Table 2.2.2).

### **2.2.3 Measurements of reliability**

Reliability is often described as a pre-requisite for validity (Anastasi & Urbina, 1996; Kline, 2000) and is defined as 'the consistency of measurement when the testing procedure is repeated on a population of individuals or groups' (APA, AERA & NCME, 1999; pp25). In the *Standards*, three aspects pertaining to the consistency and stability of test scores are identified: i) stability over time, i.e. test-retest reliability; ii) consistency across different raters/scorers, i.e. inter-rater agreement; and iii) consistency between items within the same measure, i.e. internal consistency.

#### *2.2.3.1 Test-retest reliability*

This is measured by correlating the scores from a set of individuals who take the test on two occasions. The higher the correlation, i.e. the closer the agreement between the sets of scores, the higher would be the reliability, and presumably the better the test. This inference assumes that factors that may have affected the measurement of test-retest reliability have been kept under check, such as clarity and consistency in test administration, scoring and interpretation. Some test commentators suggest that a minimum test-retest correlation of 0.8 is necessary for a test to have any value (Kline, 2000).

However, there are other factors which affect the evaluation of test-retest reliability that need to be considered, namely time and developmental lag. If the time gap between the two occasions of testing is too short, test-takers may remember the items and this familiarity may affect their performance on the second set of testing. This effect has greater implications in test items where novelty is an important element, such as tests of executive function, where the aim is to assess children's planning skills when presented with new or novel problem solving situations. On the other hand, if the time gap between the two sets of test scores is too long, differences may occur due to test-takers' actual maturational or developmental changes. This would reduce the agreement between the sets of scores and reduce the possible correlation. This is especially important where the constructs of interest are known

to show maturational effects in young children, e.g. literacy, social and cognitive skills.

Given that the focus of the present thesis is on the assessment of special educational needs, i.e. assessment of constructs that are relevant for children's development and learning, there is a high possibility that time and developmental lapses may boost or otherwise distort measurements of test-retest reliability; hence the aspect of stability over time will not be emphasised.

#### *2.2.3.2 Internal consistency reliability.*

Internal consistency refers to the extent of correlation between different items measuring the same variable. The most widely used method of measuring internal reliability is Cronbach alpha (Anastasi & Urbina, 1996). A key factor that influences the magnitude of alpha coefficients is the homogeneity of the behaviour domain sampled by the items. Many test constructors, e.g. Guilford (1956), Nunally (1978), argue that high internal reliability is a pre-requisite of high validity. However, others, e.g. Cattell & Kline (1977) assert that many psychological constructs, e.g. intelligence, personality and motivation, involve variables of some breadth and multiple dimensions. Thus, any one item in these tests is often narrower in focus, i.e. more specific, than these variables. If all items are highly consistent, and consequently highly correlated, this may well indicate that the test is too narrow and specific, and may not be valid. For example, some test items, which are mere paraphrases of each other, would have high internal consistency but low validity. While an adequate measure of internal consistency is desirable, e.g. correlation of about 0.6 (Kline, 2000), when the internal reliability reaches beyond 0.9 or higher, some concerns about the presence of redundant items may need to be addressed.

#### *2.2.3.3. Inter-rater agreement* (also known as inter-scorer reliability)

This refers to the extent to which two independent raters or scorers of a test produce similar results for the same test-taker. In standardised administration procedures, and where the criteria for interpreting a test response are unambiguous, the likelihood

that differences in individual's scores are due to differences in the method and criteria of scoring used by testers is negligible. However, inter-rater reliability is crucial in situations where an individual's score in the test is dependent on the tester's judgment of the quality of his/her responses or performance. Many clinical interviews, especially investigator-based interviews, adopt this methodology (a discussion of the investigator-based interview technique is provided in Section 2.5.3).

At the most basic level, inter-rater agreement can be measured by the correlation in the scores given by the two raters/scorers. However, as pointed out by Landis and Koch (1977), one problem with this approach is that it does not take into account the probability that the two raters' agreement can be achieved by mere chance. An alternative method, which overcomes this issue, is the use of the kappa coefficient, where the level of observed agreement between the two raters is compared against the probability of agreement by chance. The difference between the observed and expected agreement in the two scores, i.e. the kappa coefficient value, could range from 0 to 1.0. Kappa values above 0.40 are often seen as reflecting adequate inter-rater or inter-scorer agreement, i.e. significantly above chance levels of agreement (Landis & Koch, 1977).

#### **2.2.4 Summary**

The framework for validity evidence, as described in the *Standards* (APA, AERA & NCME, 1999) will serve as the basis for selecting the evaluation methods and criteria for the present thesis. Following the *Standards*' framework, in the evaluation of the indicators of SEN, the following types of validity evidence will be sought:

- Evidence based on content, which could entail review of the research literature and analyses of the extent to which the test content represents the targeted domain;
- Evidence based on response processes, which could entail analyses of potential biases that may affect participants' responses, as well as interviewers' or scorers' interpretation of data;

- Evidence based on relations to other variables, which could involve identifying the extent to which the indicators of SEN concur with external independent criteria;
- Evidence based on consequences of testing, i.e. analysis of treatment validity.

In the final analysis, through an integration of the different types of validity evidence, the findings of the present thesis will be used as the basis for developing the validity arguments for the use of the most appropriate measures of SEN for children with autism in Singapore.

In the next section, a key measure for the assessment of special educational needs will be reviewed: the International Classification of Functioning, Disability and Health or ICF (WHO, 2001). As discussed in Section 2.1.3, the ICF is notably, one of the few measures, which reflect the interactional perspective of SEN, that has extended its application to the development of an assessment tool, i.e. the ICF checklist. The evaluation methodologies discussed in this section will be relevant in the review of the ICF.

## **2.3 REVIEW OF THE INTERNATIONAL CLASSIFICATION OF FUNCTIONING, DISABILITY AND HEALTH (ICF)**

### **2.3.1. Background and rationale**

The ICF (WHO, 2001) is a recently published tool for capturing data about functioning and disability. Endorsed for use by member states of the United Nations in January 2001, it was developed to complement WHO's ICD-10 classification of disease. Its developers described the key features of the ICF as a tool to 'identify components of health' and 'a standard language for health and health related states' which 'enables users to record individual's functioning, disability and health in various domains in a similar way across cultures' (WHO, 2001, pp 7).

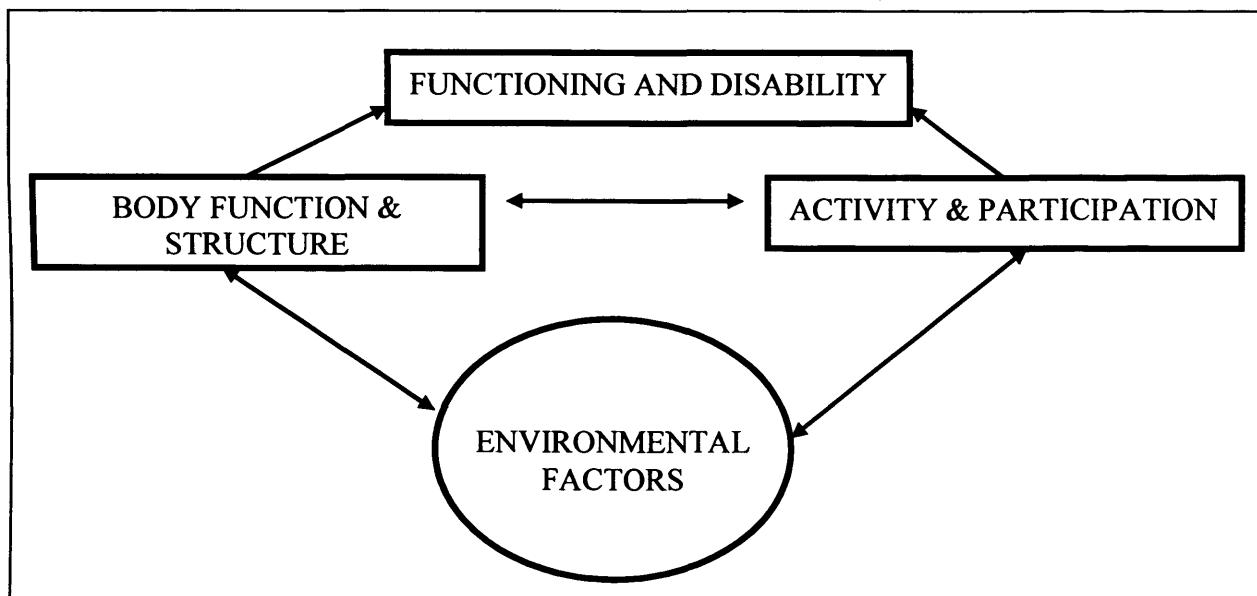
Several rationales were put forward for the development of ICF: firstly, research findings in the health care sector increasingly indicated that diagnosis alone does not predict service needs, length of hospitalisation, level of care, or outcomes; diagnosis per se is not a reliable predictor for receipt of disability benefits, work performance or the likelihood of social integration. Secondly, there was a strong recognition of the fact that a purely medical classification does not provide the information required for evaluating, planning and managing needs. It was believed that the predictive power and understanding of needs are enhanced when diagnostic information is augmented by data on functioning (WHO, 2001).

The development of the ICF was started in 1996 and involved a lengthy process of extensive literature reviews and an analysis of a pool of more than 3000 items of potential classification domains. The initial version, known as ICIDH-2, was field trialed in 15 countries (Ustun, Chatterji, Bickenbach, Kostanjsek, & Schneider, 2002).

### 2.3.2 Description of the ICF

The ICF provides, firstly, a model for understanding an individual's level of functioning and disability (Fig. 2.3.2), and secondly, an evaluation system (including an assessment checklist).

Fig 2.3.2. : ICF Model of Functioning & Disability



In its theoretical orientation, the ICF model (Fig. 2.3.2) reflects the interactional perspective, where functioning and disability are outcomes of an interaction between health conditions and environmental factors. The interaction is complex and bi-directional, as reflected by the multi-directional causal arrows in the diagram. The model does not posit a causal linkage. Instead, at each level, disability occurs within and by means of contextual factors.

The ICF measures two aspects of functioning: 1) Body Function and Structure; and 2) Activity and Participation. Disability is defined as a decrement in functioning in any one or more of these aspects, namely body impairment, activity limitation or participation restriction. The formal definition of the ICF is given in Table 2.3.2.1.

**Table 2.3.2.1: International Classification of Functioning, Disability and Health Definitions**

<b>Body Function</b>	Physiological functions of body systems (including psychological functions).
<b>Body structure</b>	Anatomical parts of the body, such as organs, limbs and their components.
<b>Impairment</b>	Problem in body function or structure, such as significant deviation or loss.
<b>Activity</b>	Execution of a task by an individual.
<b>Participation</b>	Involvement in a life situation.
<b>Activity Limitation</b>	Difficulties an individual may have executing activities.
<b>Participation Restriction</b>	Problems an individual may experience in involvement in life situations.

The aspects of functioning included in the ICF are organised into the following components:

1. **Functioning and Disability**, which includes components of:
  - Body Function and Body Structure; and
  - Activities and Participation.
2. **Environmental Factors**, which evaluate the impact of the environment as:

- Environmental Facilitators and/or Environmental Barriers.

The domains of functioning for each ICF component are shown in Table 2.3.2.2.

**Table 2.3.2.2 : International Classification of Functioning, Disability and Health List of Human Functioning – Functional Domains in Each Component**

<b>BODY</b>	
<b>Function :</b>	<b>Structure :</b>
<ul style="list-style-type: none"> <li>• Mental functions</li> <li>• Sensory Functions and Pain</li> <li>• Voice and Speech Functions</li> <li>• Functions of Cardiovascular, Haematological, Immunological and Respiratory Systems.</li> <li>• Functions of Digestive, Metabolic, Endocrine Systems</li> <li>• Genito-urinary and Reproductive Functions</li> <li>• Neuromusculoskeletal and Movement-Related Functions</li> <li>• Functions of the Skin and Related Structures.</li> </ul>	<ul style="list-style-type: none"> <li>• Structure of the Nervous System</li> <li>• The Eye, Ear and Related Structures</li> <li>• Structures Involved in Voice and Speech</li> <li>• Structure of the Cardiovascular, Immunological and Respiratory Systems</li> <li>• Structure Related to the Digestive, Metabolic and Endocrine Systems</li> <li>• Structure Related to Genito-urinary and Reproductive Systems</li> <li>• Structure Related to Movement.</li> <li>• Skin and Related Structures</li> </ul>
<b>ACTIVITIES &amp; PARTICIPATION</b>	
<ul style="list-style-type: none"> <li>• Learning and Applying Knowledge</li> <li>• General Tasks and Demands</li> <li>• Communication</li> <li>• Mobility</li> <li>• Self-Care</li> <li>• Domestic Life</li> <li>• Interpersonal Interactions and Relationships</li> <li>• Major Life Areas</li> </ul>	
<b>Community, Social and Civic Life</b>	
<b>ENVIRONMENTAL FACTORS (Barriers &amp; Facilitators)</b>	
<ul style="list-style-type: none"> <li>• Products and Technology</li> <li>• Natural Environment and Human-Made Changes to Environment</li> <li>• Support and Relationships</li> <li>• Attitudes</li> <li>• Services, Systems and Policies</li> </ul>	

Since its publication in 2001, the ICF has received positive responses.

Commentators noted that the ICF could be useful in many areas of research on disability and rehabilitation (Stucki, Ewert, & Cieza, 2002; Stucki, Cieza, Ewert, Kostanjsek, Chatterji, & Ustun, 2003), and in providing a universal framework for assessing the impact of a disability. Its main strengths are as follows: firstly, it is one

of the few assessment tools that reflect the interactional perspective. Secondly, as it is intended for use for all health conditions and across all cultures, it is comprehensive in its coverage of functional domains. “ICF is the ruler with which we will take precise measurements of health and disability” (Brundtland, 2002). However, some commentators have highlighted several issues concerning the adequacy and feasibility of the ICF framework and its applicability for special populations.

### **2.3.3 Adequacy of the ICF model**

Ueda and Okawa (2003) argued that the ICF model focused only on the components and domains of the ‘objective world’ and thus excluded the ‘subjective dimension of functioning and disability’, which they defined as:

*...a set of cognitive, emotional and motivational states of mind of any person, ... (which is) a set of active reactions to those things based on his/her personality and such psychic factors as the value system, self image, ideal, belief, the purpose in life and past experience in coping. (Ueda & Okawa, 2003, pp599).*

Ueda and Okawa suggested that the elements of the subjective dimensions could be incorporated into the ICF, by including information that reflects the level of satisfaction that the individual experiences with his/her level of functioning and health. They argued that it is not the level of impairment or environmental factors per se, but the level of satisfaction perceived by the individual that determines his/her well-being and functioning level. However, the assumption behind this argument is that a person’s level of satisfaction is not directly related to the level of impairment that he/she experiences or the level of environmental support received. No evidence was provided by Ueda and Okawa to support this assumption. In contrast, research on the quality of life for individuals with disabilities suggests that life satisfaction corresponds directly to the quality of available support (Oliver & Barnes, 1998). It can be argued that although the ‘subjective dimension’ of functioning was not directly included in the ICF, this information can be inferred from the extent to which the environmental factors are perceived as facilitators or barriers for the individual.

The strongest criticism of the ICF was perhaps made by Pfeiffer (1998), who felt that it would lead to the “medicalisation of disabilities (and thus) a short step to eugenics and a class-based evaluation of people with disabilities” (pp. 503).

Pfeiffer’s criticism of the ICF seems to stem from a social model of disability, where disability is seen as a social creation, perpetuated by the needs of dominant members of society to keep its systems ‘free’ from individuals who are unable to comply with its demands (Tomlinson, 1982). Pfeiffer called for a total abolition of the ICF and proposed instead the use of a ‘minority group paradigm’, where the primary standpoint is that people with disabilities are an oppressed group in society. Rather than giving medical professionals the power to evaluate the quality of an individual’s life and his or her needs, Pfeiffer asserted that quality of life and needs should be evaluated and decided by the individual.

In defence of the ICF, it should be noted that its developers highlighted the importance of including self-evaluations by the patient (or their proxies), in particular for the Activities and Participation component. In addition, the inclusion of environmental factors captures the extent to which factors, such as social norms, policies and systems, act as facilitators or barriers to the individual, and in so doing, reflects the extent to which individuals with disabilities could be marginalised in society. Nonetheless, Pfeiffer is correct in that although the ICF user draws on information from multiple sources, including patients’ self-reports, the final judgment on functioning levels is based on the clinician’s evaluations, and hence the power of judgment does reside in the hands of the health professionals.

However, Pfeiffer’s arguments, while honouring the rights of the individual to ascertain his/her needs, ignores the possibility that some individuals with disabilities may not be able to communicate their needs as well as others, hence the need for health professionals to act as advocates. In addition, if the decisions regarding functioning and needs were made by the incumbent individuals, there could be a danger that resources are allocated to those who are most articulate and fervent in communicating their needs.

### **2.3.4 Feasibility**

As the ICF developers aimed for it to be used across all health conditions and cultural contexts, the ICF includes a comprehensive list of over 200 items. As highlighted by Stucki et al. (2003), this may make it an impractical tool for clinical practice. To address this issue, calls were made to identify core-sets of items for the ICF, i.e. items which are relevant for specific health conditions (Ustun et al., 2002). Currently, as part of the ICF field trials, efforts are being made to identify ICF core-sets for different health conditions, including musculoskeletal, cardiovascular, chronic pain and neurological conditions. The literature searchers conducted as part of this thesis did not find any study that has attempted to identify ICF core-sets for autism.

### **2.3.5 Applicability for specific populations**

Simmeonsson, Leonard, Lollars, Bjorck-Akesson, Hollenweger and Martinuzzi, (2003) highlighted several possible domains that might have to be extended to enable the ICF to be sensitive to functional differences that occur in young children, for example activities such as pre-verbal communication, imitation and caregiver-child interaction. However, it can be argued that these aspects can be seen as fine-grained examples of ICF domains such as quality of non-verbal communication and child-parent relations. Thus, using the available domains, the ICF users may be able to probe for specific examples of functioning that are relevant to young children. Another argument against the expansion of the already comprehensive list of items in the ICF is the issue of feasibility (as discussed in the last section). Unless there is evidence that the exclusion of specific aspects (such as those suggested by Simmeonsson et al., 2003) significantly reduce the sensitivity of the ICF to measure early childhood disabilities, their inclusion in the ICF may be premature.

### **2.3.6 ICF Checklist**

To facilitate the use of the ICF, a checklist was developed (CAS, WHO; 2002) by which professionals could gather information about a person's functioning. The

rating scale (also termed 'Qualifiers' in the ICF) for each of the components is as follows:

Table 2.3.6. : ICF Checklist Structure

Component	Evaluation criteria	Qualifiers (i.e. Rating Scale)	
Body Function & Body Structure	Degree of impairment, i.e. decrement in functioning.	0 = no impairment 1 = mild impairment 2 = moderate impairment 3 = severe impairment 4 = complete impairment	
Activity & participation	Extent of activity limitation and participation restriction.	0 = no difficulty 1 = mild difficulty 2 = moderate difficulty 3 = severe difficulty 4 = complete difficulty	
Environmental Factors	Extent to which the environmental factor acts as a facilitator or a barrier.	0 = no facilitator +1 = mild facilitator +2 = moderate facilitator +3 = strong facilitator +4 = complete facilitator	0 = no barrier 1 = mild barrier 2 = moderate barrier 3 = strong barrier 4 = complete barrier

Apart from the coding sheets, the checklist also allows users to record demographic and other health information. Guidelines for use of the checklist and general questions for interviewing the respondent about functioning problems in activity and participation domains are included (See Appendix A).

### 2.3.7 Reliability and validity of the ICF checklist

As part of the development process of the ICF checklist, field trials across 15 collaborating centers were carried out to establish inter-rater reliability. The study involved live case evaluations conducted by health professionals in each center and were reported to involve inter-rater codings for 1884 cases (CAS, WHO; 2002). However, no information could be obtained regarding the breakdown for these cases in terms of the specific disability conditions and ages of the patients.

Based on these case evaluations, Kappa was calculated for agreement (See Table 2.3.7). It can be seen that the agreement is high for all domains. However, there

were no reported data on the extent to which an individual's ratings for items within a particular component were consistent, i.e. no indices for internal reliability.

**Table 2.3.7 : ICF Field Trial Results – Kappa values (WHO, 2000, slides #43-47)**

<b>Body Functions</b>	<i>Range of Kappa Values<sup>1</sup></i>	
	<i>Min.</i>	<i>Max.</i>
Mental	0.79	0.94
Sensory	0.81	0.90
Voice & speech	0.82	0.82
Cardiovascular, haematological etc	0.80	0.95
Digestive, metabolic etc	0.74	0.87
Genito-urinary and reproductive	0.80	0.86
Neuromusculoskeletal and movement	0.83	0.91
Skin	0.87	1.00
<b>Body Structure</b>		
Nervous system	0.87	1.00
Eye, ear	0.85	-
Voice & speech	0.85	-
Digestive, metabolic etc	0.77	1.00
Genitourinary	0.81	-
Movement	0.66	0.85
Skin	0.72	-
<b>Activities &amp; Participation</b>		
Learning & applying knowledge	0.86	0.92
Communication	0.80	0.95
Movement	0.80	0.88
Moving around	0.79	0.93
Self care	0.83	0.95
Domestic	0.81	0.94
Interpersonal	0.78	0.88
Tasks and major life situations	0.83	0.88
Personal maintenance	0.91	0.97
Mobility	0.84	0.95
Exchange of information	0.88	0.90
Social relationships	0.87	0.95
Home life	0.95	0.98
Education	0.69	0.83
Work and employment	0.79	0.87
Economic life	0.88	0.91
Community, Social and Civic life	0.81	0.94
<b>Environmental Factors</b>		
Products and Technology	0.77	0.83
Natural environment & human made changes	0.60	0.78
Support and Relationship	0.77	-

<sup>1</sup>Where only a minimum value is provided, it indicates that only one item was available for analysis.

Attitudes, values and beliefs	0.83	0.97
Services	0.78	0.85
Systems and Policies	0.77	0.79

Validity studies on the ICF are currently on-going (Stuki et al., 2002) and involve two phases:

Phase 1 aims to identify core sets of ICF domains for 12 different health conditions, including:

- Musculoskeletal conditions (back pain, osteoporosis, rheumatoid arthritis, osteoarthritis),
- Cardiovascular/internal conditions (coronary heart disease, asthma, diabetes mellitus, breast cancer),
- Chronic pain/psychological/neurological conditions (obesity, pain disorders, depressive disorder, stroke).

Preliminary reports indicated that for musculoskeletal conditions, the ICF checklist was able to capture the profile of patients in all four conditions, and good concurrent validity was reported between the checklist and the SF 36 Body Function subscale. However, no detailed information was provided in the interim report to support the conclusions (CAS, WHO, 2002).

Phase One field trials are expected to be completed by 2005, and Phase Two, which aims to test the feasibility, reliability, validity and sensitivity of the condition specific core-sets developed during Phase One, will follow. To accomplish this objective, WHO is conducting a multi-center cohort study with 3000 patients (Ustun, Chatterji, Kostanjek, & Bickenbach, 2003). There are no reported plans for testing the reliability and validity of the ICF for use with children with developmental disabilities, such as autism.

### **2.3.8 Critique of the ICF**

As discussed earlier, the ICF is intended as a measure of functioning for all disability conditions. Its sensitivity for children with a specific disability, namely autism, is unknown. In this section, the ICF will be compared against other measures

of functioning, with a view to identify the most appropriate measure for special educational needs for the present thesis. Two other measures will be discussed, namely the Vineland Adaptive Behaviour Scales (Vineland: Sparrow, Balla & Cicchetti, 1984), which has been developed and used extensively with children across disability categories and the Diagnostic Interview for Social and Communication Disorder (DISCO: Wing, 1999), which was designed specifically for children with autism.

- The Vineland is a standardized normative assessment instrument for documenting children's adaptive functioning. Based on parents' or caregivers' reports, the Vineland provides information in the domains of communication, daily living, socialization and motor skills.
- The Diagnostic Interview for Social and Communication Disorder (DISCO: Wing, 1999) is an investigator-based interview developed to assist clinicians in the diagnoses of autistic spectrum disorders, and includes a comprehensive list of items covering four categories of behaviour, namely behaviour in infancy, onset of recognition of problems, developmental behaviour, and atypical behaviours.

The comparison of the ICF against these two measures is based on the following criteria:

- Content validity, i.e. whether the items included in the measures adequately reflect the domains of functioning that are relevant for the assessment of special educational needs of children with autism;
- Psychometric validity, i.e. whether the reliability and validity indices of the measures are adequate;
- Feasibility, i.e. the practical constraints and benefits of using the measures.

Each of the criteria will be discussed in turn, with reference to the three measures, namely ICF, Vineland and DISCO.

### *Content Validity*

Based on the review of the conceptual framework for SEN, it was concluded that the interactional perspective most closely reflects current knowledge about the factors underlying special educational needs, i.e. that it is the effect of both within child and environmental factors (See section 2.1.3). Hence measures of SEN used in the present thesis should, ideally, be coherent with this perspective, i.e. they should reflect domains of functioning that are impaired in children with autism and relevant environmental factors. Table 2.3.8.1 shows a comparison of the domains that are included in the three measures, namely the ICF, DISCO and Vineland. It can be seen that all three measures capture the aspects of development that are associated with the triad of impairments in autism, namely communication; social-interaction; and imagination (e.g. play activities).

All three measures involve more than 200 items. However, compared with the Vineland the items in both the ICF and the DISCO cover a larger number of functional domains. Notably, the ICF is the only measure that includes environmental factors as one of the aspects which affect a child's developmental functioning. In addition, as the ICF was intended for use across all cultures, efforts were made to ensure that the items reflected as many culturally specific examples as possible (See discussion in Section 3.1.1).

Table 2.3.8.1 Comparison of the Domains Included in the ICF, DISCO and Vineland

<i>ICF</i>	<i>DISCO</i>	<i>Vineland</i>
<b>Method: Investigator-based Interview</b>	<b>Method: Investigator-based Interview</b>	<b>Method: Structured questionnaire or checklist</b>
<b>Administration time :</b> <b>1 – 2 hours</b>	<b>Administration time :</b> <b>2 – 3 hours</b>	<b>Administration time :</b> <b>1 – 2 hours</b>
<b>Body Functions and Structure</b>	<b>Behaviours in infancy</b>	<b>Communication</b>
• Mental	<b>Onset of problems</b>	• Receptive
• Sensory	<b>Developmental behaviours</b>	• Expressive
• Voice & speech	• Gross motor skills	• Written
• Cardiovascular, haematological	• Self-care: toileting, feeding, dressing, hygiene	<b>Daily living skills</b>
• Digestive, metabolic etc	• Domestic skills	• Personal
• Genito-urinary and reproductive	• Social interactions with adults	• Domestic
• Neuromusculoskeletal and movement	• Social interaction with peers	• Community
• Skin	• Social play and leisure	<b>Socialisation</b>
• Nervous system	• Independence	• Interpersonal relations
• Eye, ear	• Imitation	• Play-leisure time
• Voice & speech	• Imagination	• Coping skills
• Digestive, metabolic etc	• Receptive communication	<b>Motor skills</b>
• Genitourinary	• Expressive communication	
• Movement	• Non-verbal communication	
• Skin	• Visual spatial skills	
<b>Activities &amp; Participation</b>	• Skills in understanding pictures, reading & writing, numeracy, etc	
• Learning & applying knowledge	• Cognitive skills	
• Communication	<b>Repetitive Activities</b>	
• Movement	• Motor & vocal responses	
• Moving around	• Movement	
• Self care	• Proximal sensory stimuli	
• Domestic	• Auditory stimuli	
• Interpersonal	• Visual stimuli	
• Tasks and major life situations	• Repetitive routine and change resistance	
• Personal maintenance	• Overall pattern of activities	
• Mobility	<b>Emotions</b>	
• Exchange of information	• Lack of or inappropriate affect	
• Social relationships	<b>Maladaptive behaviours</b>	
• Home life	• Behaviour affecting others	
• Education	• Sleep disturbance	
• Work and employment	• Catatonic features	
• Economic life		
• Community, Social and Civic life		
<b>Environmental Factors</b>		
• Products and Technology		
• Natural & human made changes		
• Support and Relationship		
• Attitudes, values and beliefs		
• Services, Systems and Policies		

### *Reliability*

The reported reliability indices of the ICF, Vineland and DISCO are strong: the Vineland's internal and inter-rater reliability are reported to be high: internal reliability indices (alpha) were between 0.8 to 0.9; and inter-rater correlation was 0.74 (Sparrow et al., 1984). Similarly, the inter-rater agreement for the DISCO and ICF are within good to excellent levels (CAS, WHO, 2002; Wing et al.<sup>2</sup>). However, because they were not intended to provide a composite (or global) index of functioning, neither the DISCO nor the ICF have been evaluated for its internal reliability (i.e. consistency between items of the same component/domain).

### *Validity*

The discriminant validity (i.e. its ability to distinguish children with autism from those without autism) of the DISCO is reported to be good (Leekam et al., 2002; see section 2.5.4.5). Similarly, the Vineland was found to be able to adequately distinguish children with autism, from those with developmental delay, but without autism (Volkmar, Sparrow, Goudreau, Chichetti, Paul & Cohen, 1987). As discussed in section 2.3.7, while the validity of the ICF for use with patients with musculoskeletal conditions have been reported (CAS, WHO, 2002), its validity for children with developmental disability (e.g. autism) is unknown.

Another aspect of validity is the appropriateness of the developmental milestones used to gauge children's functioning. The Vineland is a standardized normative instrument and its validity rests on the assumption that the profile/background of the children tested are similar/comparable to the children used in its normative sample. As the Vineland has not been adapted or normed for the Singapore population, this assumption is problematic for the present study. In contrast, in the ICF and the DISCO, judgments about the appropriateness of children's behaviour are based on developmental milestones which can be adjusted to different population standards (see Section 3.2.2.1 for detailed examples).

---

<sup>2</sup> Discussions on the reliability of the ICF and the DISCO are in Sections 3.14 and 2.5.4 respectively.

### *Feasibility*

As the Vineland includes fewer domains of functioning and it can be completed as a checklist by parents, it entails a shorter administration time, compared with the ICF and the DISCO (see Table 2.3.8.1). In contrast, as the ICF and DISCO use the investigator-based interview method<sup>3</sup>, more time would be involved in the administration of these measures.

However, practical issues, such as the need to minimize administration time, must be balanced with the need to ensure reliability and validity of the information obtained. One of the issues with instruments that rely heavily on parental reports is the inherent assumption that parents have the required understanding of complex social behaviours in order to accurately report its frequency/occurrence (this issue is discussed further in section 2.5.2). Investigator-based interviews, albeit incurring greater time and effort to administer, is one way of overcoming this problem, as they make no such assumptions about parents' understanding. Instead, such interviews rely on the trained interviewer (i.e. the practitioner) to make judgments about the presence and severity of the problem behaviour, based on the parents' description of children's behaviour.

Another aspect of feasibility is the length of training required to use the instruments. For experienced practitioners, both the Vineland and the ICF requires minimal training through self-guided reading and practice. The training in the use of the DISCO however, requires greater time and cost (the specialised training for the DISCO can only be obtained from licensed trainers).

As summarized in Table 2.3.8.2, all three measures have advantages and limitations. However, it can be argued that for the purpose of the present study, the ICF provides the strongest advantages: its content is comprehensive and includes components of environmental factors, which is consistent with the interactionist perspective of SEN

---

<sup>3</sup> A description of the investigator-based interview method and its advantages over parental report/checklist is discussed in detail in Section 2.5.3 .

used in the present thesis; it has demonstrated good inter-rater reliability and validity (for musculoskeletal disorders); and the high 'time cost' (i.e. administration time) in using the measure is balanced with a low training cost.

**Table 2.3.8.2 Summary of the Advantages (✓), Limitations (✗) and Unknown (?) Aspects of the ICF, Vineland and DISCO.**

Criteria	ICF	Vineland	DISCO
Content validity	✓ Comprehensive ✓ Includes environmental factors	✗ Less comprehensive ✗ Omits environmental factors	✓ Comprehensive ✗ Omits environmental factors
Reliability	✓ Good inter-rater reliability ? Internal reliability unknown	✓ Good inter-rater reliability ✓ Good internal reliability.	✓ Good inter-rater reliability ? Internal reliability unknown
Validity	? Discriminant validity for autism unknown ✓ Appropriate developmental norms available.	✓ Good discriminant validity for autism. ✗ Available norms not appropriate.	✓ Good discriminant validity for autism. ✓ Appropriate developmental norms available.
Feasibility	✗ Time cost is high ✓ Training cost is low	✓ Time cost is low ✓ Training cost is low	✗ Time cost is high ✗ Training cost is high.

### 2.3.9. Summary and conclusion

The ICF is potentially a useful measure of functioning and disability. It is based on an interactional perspective, where functioning is seen as the effect of both within child (Body Functions and Structure) and contextual factors (Environmental Factors). Its accompanying assessment tool, i.e. the ICF checklist, equally reflects this interactional perspective. It is comprehensive in its coverage of functional domains and includes aspects that are applicable across different cultures and contexts. Compared with other measures of developmental functioning, namely the Vineland and the DISCO, the ICF offers a number of advantages as a measure of SEN.

Thus, the concept and definition of special educational needs in the present thesis will be based on the ICF framework, as follows:

Children with special educational needs are defined as those who experience activity limitations and participation restrictions, as a result of impairments in body functions and/or environmental barriers.

The initial field trials of the ICF across 15 countries have shown that the ICF has good inter-rater reliability and is able to capture the profile of patients for back pain, osteoporosis, rheumatoid arthritis and osteoarthritis. However, its reliability and validity for use with children with developmental disabilities, such as autism, remains to be seen. In order to be considered as a reliable and valid measure of special educational needs in children with autism, data would be needed to ascertain if the ICF checklist meets the following criteria for reliability and validity:

1. ICF ratings of individual children must show consistency between items within the same component, as well as consistency in the evaluations of the same children by different raters (i.e. internal and inter-rater reliability);
2. ICF measures of children's functioning (namely impairments in body functions/structures, activity limitation and participation restriction) must be able to meaningfully distinguish children with varying levels of special educational needs (i.e. criterion validity); and
3. There must be concurrence between evaluations of children's functioning based on the ICF with other independent measures of functioning, in particular those which have been developed specifically for individuals with autism (i.e. criterion validity).

An evaluation of the ICF against the 3 stated criteria would be undertaken in Study 1 of the present thesis (See Chapter 3).

However, even if the ICF proves to have adequate reliability and validity, it is only one measure of special educational needs. Psychological assessments involve a process whereby the practitioner integrates information from a variety of scores from measures obtained from multiple methods:

*... psychological assessment is concerned with the clinician who takes a variety of test scores, generally obtained from a multiple test methods, and considers the data in the context of history, referral information, and observed behaviour to understand the person being evaluated, to answer the referral questions, and then to communicate findings... (Meyer et al., 2001, pp14).*

In order to fulfil the aim of the present thesis, i.e. to identify methods that can improve the assessment of SEN for children with autism, there is a need to identify other possible indicators that can be used alongside the identified SEN measure (potentially, the ICF). The distinction between a measure and an indicator of SEN can be clarified as follows:

- A measure aims to capture, as comprehensively as possible, all the relevant dimensions of the construct of interest (i.e. SEN), while an indicator focuses only on one dimension;
- A measure focuses on the current level of functioning, while an indicator aims to use information on current performance as a predictor of future performance or functioning;
- A measure, because of its comprehensive coverage, usually requires more time and resources to implement.

In this thesis the ICF is considered as a possible independent **measure** of SEN, while the **indicators** of SEN are identified based on a literature review of the research on the cognitive theories of autism, such as theory of mind, executive function and central coherence (see Section 2.4). Despite its narrower focus, the strength of these cognitive indicators lie in the extent to which they are seen to reflect the core impairments that underlie the learning difficulties of children with autism. The ICF is considered as an ‘independent measure’ because the children’s scores in the ICF are not dependent on any of these cognitive indicators of learning. One of the key aims in the present thesis is to identify which of the cognitive indicators best predict children’s SEN level, as measured by the independent measure, i.e. ICF.

In the next section, the concepts and causal theories of autism will be discussed, with a view to identify the possible cognitive indicators that reflect the core learning impairments in children with autism.

## 2.4 CONCEPTS AND THEORIES OF AUTISM

### 2.4.1. Concept of autism

Clinicians and researchers have achieved some consensus on the validity of autism as a diagnostic category and on the many features central to its definition (Gillberg, 1996). There is broad agreement that autism is a developmental disorder, in which the behavioural manifestations are attributable to underlying dysfunctions. Although the aetiology of these underlying dysfunctions is not yet determined, it is generally believed to be related to deviations in neurological maturation and functioning (Minschew, Sweeney, & Bauman, 1997; Klin & Volkmar, 1997; Hill & Frith, 2003). However, there is considerable debate regarding the subcategories of autism and its related disorders. A central issue concerns the conceptualisation and validity of disorders associated with autism. There are two approaches to the conceptualisation of autism: 1) the categorical approach; and 2) the multi-axial or dimensional approach.

#### 2.4.1.1. *Categorical approach*

In this approach the different disorders associated with autism, e.g. Asperger's Syndrome, Rett disorder, atypical autism and childhood disintegrative disorder, are seen as distinct disorders, although there may be overlaps in terms of the observed behavioural symptoms. The assumption is that they represent qualitatively different conditions, and consequently different aetiology is assumed. This is the approach used in the DSM-IV (APA, 1994) and ICD-10 (WHO, 1995).

There is substantial research support for the 3 diagnostic criteria that are central to the definition used in DSM-IV and ICD-10, namely social abnormalities; impairments in communication; and restricted repetitive and stereotyped patterns of behaviour (see Appendix B for a detailed description of the DSM-IV and ICD-10 diagnostic criteria). There has been significant work in the last three decades, which highlighted social and communication abnormalities as critical aspects in autism (e.g. Cohen, Caparulo, Gold, Waldo, Shaywitz, & Rimland, 1977; Siegal, Anders, Ciaranello, Bienenstock, & Kraemer, 1986; Lord & Pickles, 1996).

Several issues have been raised concerning the DSM-IV/ICD-10 definition of autism. One problem lies in the fact that the categorical approach to autism assumes that a clear, qualitative distinction between autism and other pervasive developmental disorders can be made. For example, in the diagnostic criteria for autism, Rett Disorder and Asperger's Syndrome, impairments in social interaction and communications are equally emphasised (See Table 2.4.1.1).

Table 2.4.1.1 : DSM-IV Diagnostic Criteria for Autism, Rett Disorder and Asperger's Syndrome

	<b>Autism</b>	<b>Rett Disorder</b>	<b>Asperger's Syndrome</b>
Overlapping criteria	<ul style="list-style-type: none"> <li>• Impairment in social interaction;</li> <li>• Impairment in communication;</li> <li>• Restricted, repetitive and stereotyped behaviour.</li> </ul>	<ul style="list-style-type: none"> <li>• Impairment in communication and impaired social interaction;</li> <li>• Severe impairment of expressive and receptive language;</li> <li>• Stereotyped hand movements.</li> </ul>	<ul style="list-style-type: none"> <li>• Impairment in social interaction;</li> <li>• Restricted repetitive and stereotyped patterns of behaviour, interest and activities.</li> </ul>
Distinctive criteria	<ul style="list-style-type: none"> <li>• Onset of delays or abnormal functioning prior to age 3 years.</li> </ul>	<ul style="list-style-type: none"> <li>• An apparently normal prenatal and perinatal period and normal psychomotor development at infancy;</li> <li>• A deceleration of head growth between 5 months and 4 years, and a loss of acquired purposeful hand skills between 5 and 30 months of age.</li> </ul>	<ul style="list-style-type: none"> <li>• No clinically significant general delay in language;</li> <li>• No clinically significant delay in cognitive development or age-appropriate self-help, adaptive behaviour and curiosity about the environment in childhood.</li> </ul>

One of the distinctive criteria for Rett Disorder is a regression in development and a loss of acquired purposeful hand skills before the age of three. However, as pointed out by Wing (1997), these characteristics, which involve the loss of previously acquired skills, is often reported by many parents of children with autism. For Asperger's Syndrome, the distinctive criterion stipulated in the DSM-IV is that of normal development in the first 3 years of life. Wing (1988) made the observation that delays and abnormalities in language and social communication were frequently present in the first 2 years of life of many individuals with Asperger's Syndrome.

Other commentators agree that the criteria for normal early language development makes the category ‘artificially rare’ (Ehlers & Gillberg, 1993). Thus it can be argued that the ‘distinctive criteria’ that are supposed to provide differential diagnoses (See Table 2.4.1.1) are, in fact overlapping.

Other issues with the DSM-IV/ICD-10 definition of autism have also been raised. For example, it has been suggested (Baron-Cohen, 1989) that the sharing of attention and interest (i.e. joint-attention) is a precursor to social reciprocity (i.e. the ability to take on other person’s perspective and interest). This is based on a theory of mind hypothesis (described later in Section 2.4.2) which postulates the absence of mentalising abilities as the fundamental impairment in autism. Some commentators have also suggested that some characteristics are ‘more critical’ than others. For example, Lord and Pickles (1996) suggested that the extent of language delay itself is an index of severity of autism, and language delay accounts for some of the social and communication impairments of children with autism. Hence, they argued that the aspect of language impairment should be given greater emphasis, relative to the other diagnostic criteria. Another dissatisfaction with the DSM-IV/ICD-10 criteria for autism is that they do not take into account the observed fact that characteristics of autism change over-time (Beadle-Brown, Murphy, Wing, Gould, Shah, & Holmes, 2002); and this developmental quality is not adequately reflected in the DSM-IV/ICD-10 criteria.

#### *2.4.1.2. Multi-Axial or Dimensional Approach*

Arising from the dissatisfaction with the categorical approach, several commentators have proposed a multi axial, or dimensional, approach to the definition of autism (Ferari, 1982; Wing, 1997). In an epidemiological study of children with special educational needs in one area of London, Wing and Gould (1979) found that all the features listed in all diagnostic systems clustered together significantly. Based on the observed pattern of clusters, they formulated the ‘triad of impairments’ concept, which encompasses:

1. Impairments in reciprocal social interaction;
2. Impairments in verbal and non-verbal communication; and
3. Impairments in imagination (identified as being closely related to narrow, rigid, repetitive patterns of behaviour).

Using the triad as the defining parameter, children with a variety of clinical conditions were thus seen as having 'autistic spectrum disorder' or ASD. Among the children that fitted the description of having the triad of impairments were those who fitted Kanner's (1943) original description for autism, as well as a small number of children with Asperger's Syndrome. This could be due to the study's focus on children with special educational needs; and thus, possibly missing out the children with Asperger's who are more likely to be in mainstream schools (Ehlers & Gillberg, 1993). At the same time, a large number of children did not fit into any of the named diagnostic categories, but instead had various combinations of features (Wing, 1997). It was also noted that in some of the children, the clinical picture changed with age, and some children with Kanner's Syndrome in early childhood developed behaviours which are more characteristic of Asperger's Syndrome in adolescence.

The concept of Autistic Spectrum Disorder, or ASD, includes all the pervasive developmental disorders as defined in DSM-IV and ICD-10. The spectrum is in fact wider, as it includes even subtle manifestation of the triad.

There have been studies that provide some arguments for broadening the criteria for autism, for example a genetic study by Bolton, MacDonald, Pickles, Rios, Goode, Crowson, Bailey, and Rutter (1994) on families with autism. However, there has not been any reported study which validates the breadth and scope of features that are subsumed as a part of a single autistic continuum, as implied by the concept of the triad of impairment.

#### *2.4.1.3. Summary and conclusion*

In both the categorical and dimensional approaches, assumptions are made about the aetiology of the disorder and its patterns of symptoms; and neither approach is supported by an underlying theoretical basis. Neither approach circumvents the need for external validation: the categorical approach would need to be supported by research indicating that the different categories are clinically different; while the dimensional approach needs to be validated by research indicating that all the conditions subsumed as ASD are, in fact, qualitatively similar.

However, given the balance of evidence, there appears to be a greater consistency between the dimensional concept of ASD and the varied clinical picture that is often seen among individuals with autism. In addition, as it does not exclude children whose developmental history may not fit any of the named diagnostic categories, practitioners would be able to accommodate the changes that often occur in the developmental profile of children with autism. Since its introduction, the concept of ASD has received much acceptance in clinical practice; its widespread use may well be one of the factors behind the upsurge in incidence of children diagnosed with autism (Blaxill, 2002).

In the present thesis, children with autism are defined as those with a pre-existing diagnosis of ASD (Wing & Gould, 1979). Given the breadth of the spectrum and its widespread use, one implication is that there may be large variability in the patterns of behaviour of children with ASD. As such, it would be useful to obtain some measure of the degree to which autism characteristics are seen in the children, especially when group comparisons are made. For this purpose, the Gilliam Autism Rating Scale or GARS checklist (Gilliam, 1995) could be used. The GARS checklist, which will be discussed in Section 2.5.2., is a parent completed checklist that provides a measure of the severity of autism characteristics in children. Information from the GARS will be used to gauge the variability (if any) in the degree of autism characteristics between groups of children in the present thesis.

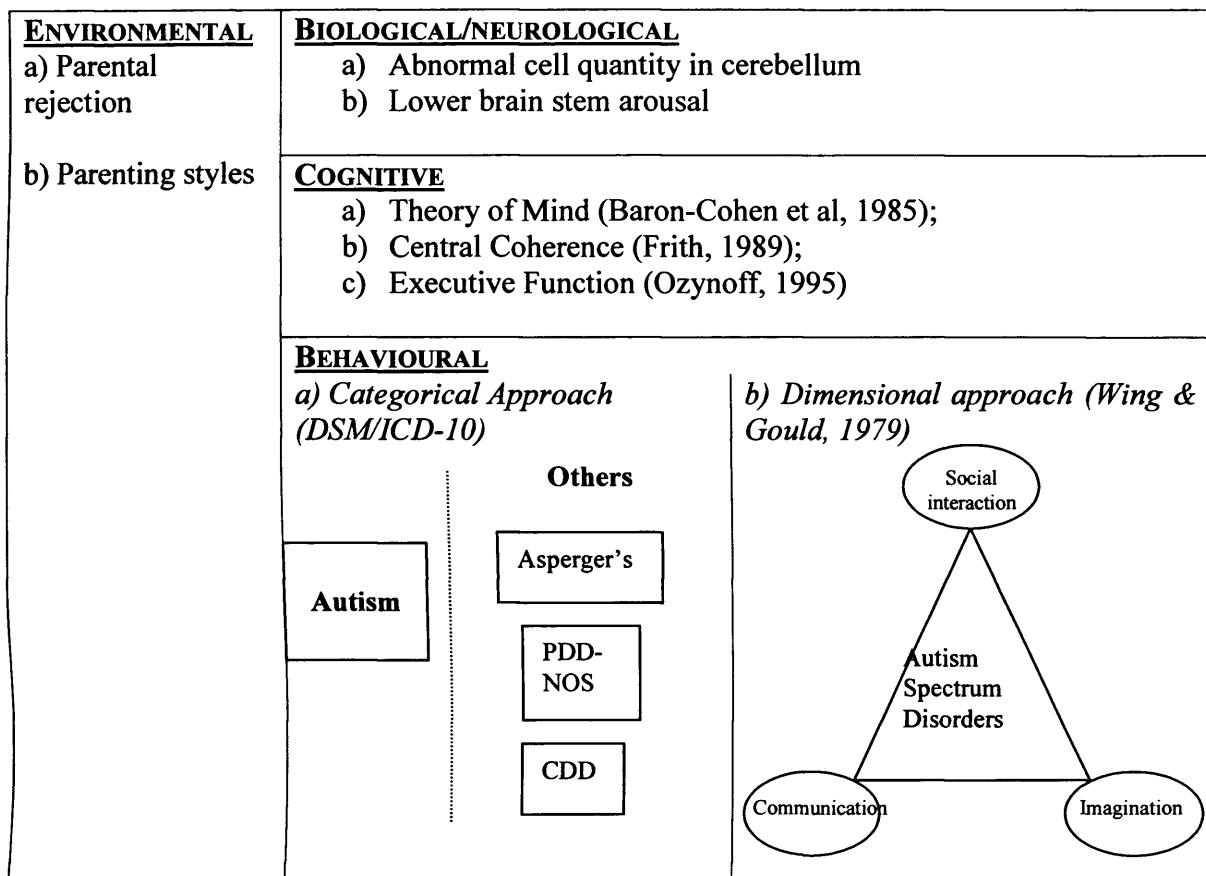
## 2.4.2 Theories of autism

In the previous section, it was highlighted that neither the categorical nor dimensional concept of autism is supported by a clear explanation about the aetiology of the disorder. However, in the last ten years there have been several theories that may offer some insight into the underlying causes of autism. In this section, key theories of autism will be reviewed. Morton and Frith's (1995) causal modeling framework is useful in highlighting how different theories have attempted to explain autism at different causal levels. It is argued that while some theories are potentially complementary, in that they present coherent explanations of the same behaviour at different levels, others are less compatible, i.e. they emphasise contrasting explanations for autism at similar causal levels (See Fig. 2.4.2.).

At the behavioural level, the categorical and dimensional views of autism, in themselves, do not posit any causal explanation, although for each approach, different causal paths might be implicated. It is therefore necessary to consider theories that relate to other levels of explanation so as to identify a definition of autism that provides the most coherent picture linking the observed pattern of behaviours and its hypothesised underlying causes.

The earliest theories of autism attributed environmental factors such as parental rejection and poor maternal communication as the main causes (Battelheim, 1967). However, studies conducted in the 1970's (e.g. Schopler, 1971) found no evidence of parental personality characteristic or impaired parent-child interactions.

Fig. 2.4.2: Main Locus of Explanations in Autism Theories



Reviews of the biology of autism conclude that there is evidence for an organic cause (Coleman & Gillberg, 1985; Schopler & Mesibov, 1987). The evidence relating to an organic link in autism has spurred investigations into the specific neurological patterns that could be associated with the behaviours of individuals with autism. Insights into the neurological basis of autism come from a variety of sources, such as studies on neuropathology (e.g. Bauman & Kemper, 1985) and studies using neuroimaging techniques (e.g. Gillberg, Rosenthal, & Johansson, 1983; Skoff, Mirsky, & Turner, 1980).

The neurological evidence supports some of the behavioural observations of autism. However, it was noted (Bauman & Kemper, 1994) that these studies showed inter-subject inconsistencies. This suggests that studies on the biological basis of autism have not yet fully separated causal and correlative relationships between the

different types of neurobiological pathology and autism. Thus, in the effort to find underlying causes that can be mapped onto the observable characteristics of autism, the intervening cognitive level, linking biological/neural causes and the behavioural observations, is crucial. Currently, there are several prominent cognitive theories of autism, namely Theory of Mind, Central Coherence and Executive Function.

#### *2.4.2.1. Theory of mind*

Baron-Cohen et al. (1985) suggested that the triad of impairments associated with autism (at the behavioural level) stem from an impairment in the ability to 'mind-read'. Their study and use of the term 'theory of mind' follows from a study by Premack and Woodruff (1978) which indicated that normal children, from around the age of 4 years, understand that people have beliefs and desires about the world, and that these mental states (rather than the physical states of the world) determine a person's belief and behaviour. The theory of mind explanation of autism suggests individuals with autism lack this ability to think about thoughts (or 'mentalising' ability) and are therefore specifically impaired in certain social, communication and imaginative skills in which ability for mentalising is required. Findings from various studies involving the use of false-belief tests and tests of mentalising abilities provided some support for this idea (Baron-Cohen et al., 1985; Leslie & Frith, 1988). (These studies and the issues concerning the selection of measures for theory of mind are discussed in Chapter 4).

The absence of theory of mind has been used to explain a wide range of social, language and communicative behaviours that have been associated with autism (Baron-Cohen, Tager-Flusberg, & Cohen, 1995). Taken in its strong form, i.e. the hypothesis that autistic characteristics (behavioural level) are due to the *absence* of mentalising abilities (cognitive level), the theory of mind hypothesis would map well with the categorical approach to defining autism, which assumes that autistic disorder is distinctive from other disabilities involving social and communication disorders.

However, the data from the studies cited by Baron-Cohen et al. (1995) do not fit this ‘strong form’ of theory of mind hypothesis. For example, the theory of mind deficit in autism does not explain other symptoms in autism (e.g. rigidity, repetitive behaviour, sensory hypersensitivity). In addition, in the studies cited to support the theory of mind hypothesis, there were some children with autism who appear to demonstrate adequate mentalising skills.

#### *2.4.2.2. Central coherence*

Frith (1989 & 2003) proposed the idea that a primary deficit in autism is an impairment in central coherence. In normal cognition, there is a propensity to form coherence over as wide a range of stimuli as possible, and to generalize over as wide range of contexts as possible. The normal operation of central coherence compels human being to give priority to understanding meaning, to make ‘sense’ from perceiving connections and meaningful links from meaningless materials. Frith posits that in children with autism, this capacity for coherence is diminished. Thus, their ability to process information is affected in that ideas and thoughts are ‘detached’ from context and lack meaningful connectedness with one another. Frith uses the theory to explain observations that children with autism do particularly well in the block design tasks<sup>4</sup> and embedded figure tasks (Asarnow, Tanguay, Bott, & Freeman, 1987; Shah & Frith, 1983); show good discrimination and categorization abilities, and yet poor generalization of learning (Ungerer & Sigran, 1987); display an attention pattern described as showing a tendency to ‘hyperfocus’ or stimulus over-selectivity (Lovaas, Koegal, & Schreibman, 1979); and tend to show a poor grasp of the pragmatics of language, despite good expressive and receptive vocabulary (Schopler & Mesibov, 1985).

Commentators noted that one limitation of the central coherence theory of autism is a lack of plausible neuro-biological mechanisms that can be linked to the observed cognitive patterns (Hill & Frith, 2003). In addition, like the theory of mind deficit

---

<sup>4</sup> For example, the Block Design subtest in Wechsler Intelligence Scales for Children, which is described in Chapter 4.

discussed earlier, weakness in central coherence is not a universal feature of autism, i.e. there is a small sub-sample of children with autism that does not appear to demonstrate the cognitive profiles associated with weak central coherence.

In terms of its practical applications, there is a lack of studies that evaluate the relationship between strengths in central coherence and the adaptive skills of children with autism. For example, although it is well established that children with autism do better at embedded figures tasks, it is not known if the degree in weaknesses in central coherence is directly related to the severity of social and communication dysfunction that children with autism experience in their daily lives. There are also other aspects of autism behaviours, such as the lack of initiation of new actions, and perseveration (i.e. the tendency to be stuck in a given task set) that may be better explained by reference to other aspects of cognition, such as executive function.

#### *2.4.2.3. Executive function*

Ozonoff (1995) proposed that impairments in executive function underlie the observed deficits and patterns in autism. Executive function is believed to be involved in the following activities:

1. Considering alternatives in planning;
2. Keeping a number of items in working memory;
3. Allocating attention to competing stimuli;
4. Balancing priorities and weighing consequences for alternative courses of action;
5. Considering available resources; and
6. Thinking of possible action before taking action.

What executive function behaviours have in common is the ability to disengage from the immediate environment or external context and guide behaviour instead by mental/internal processes (Shallice, 1982).

There has been some evidence supporting the relationship between autism and deficits in executive function. For example, Prior and Hoffman (1990) noted that children with autism were not able to identify alternative strategies to overcome difficulties, resulting in persistent errors. This was seen in impaired performance in the Wisconsin Card Sorting Test (WCST) and a maze test. Based on a comprehensive test battery, including measures of executive function, theory of mind, emotional perception, verbal memory and spatial abilities, Ozonoff, Pennington and Rogers (1991) found that among children with average ability, impaired performance on tasks of executive function was most widespread in the autistic sample. Ozonoff proposed that the difficulties that children with autism face in theory of mind tasks could be attributed to deficits in executive function. Thus, “what appears to be an inability take another’s perspective may actually be a failure to disengage from the external context and use an internal representation (e.g. someone else’s belief) to solve the problems” (Ozonoff, 1995; pp. 208).

Although the theory of executive dysfunction is able to explain some cognitive aspects associated with autism, some commentators (e.g. Leslie & Roth, 1993) highlighted that children with autism are capable of inhibiting responses or suppressing a salient perceived reality in specific tasks (e.g. when answering questions about photographs or maps). This suggests that the structure of executive function mechanisms may be domain-specific or ‘fractionated’, and perhaps only *some* of the domains in executive function are implicated in autism.

#### *2.4.2.4 Causal theories of autism and indicators of SEN*

For each of the constructs related to the causal theories of autism, namely theory of mind, executive function and central coherence, there are theoretical and empirical bases for the link between impairments in these areas to deficits in the social, communication and cognitive functions of children with autism (as discussed in the previous sections). However, there has not been any reported study that investigated a direct relationship between these constructs and special educational needs.

One of the issues in using these constructs as indicators of SEN is the choice of appropriate measurement/ assessment tools. Most of the empirical work on the theory of mind and central coherence of children with autism is based on experimental data; the validity of using these measures for high stakes psycho-educational testing have yet to be evaluated (The issues on the selection of appropriate measurement tools for the different indicators of SEN are discussed in detail in Chapter 4).

Most of the empirical research on these constructs has been focused on establishing their diagnostic validity, i.e. to establish (criterion) validity in distinguishing children with autism from other clinical populations and normally developing samples (as discussed in the previous section). There appears to be very little exploration of the practical applications of these constructs in the educational assessment of children with autism. For the purpose of the present thesis, it would be important to evaluate the indicators based on the types of validity evidence that are relevant for the intended uses, i.e. assessment of special educational needs. It was argued (Section 2.2.2) that in the context of assessment of special educational needs in Singapore, the evidence for criterion validation would need to be based on the intended outcomes of SEN assessment, namely to obtain accurate information on children's functioning level and decide the appropriate educational placement (i.e. criterion validity); and plan effective intervention strategies, i.e. treatment validity. Studies 2 and 3 in the present thesis aim to obtain these types of validity evidence.

However, it can be argued that these constructs, namely theory of mind, executive function and central coherence are too narrowly focused on within-child factors of learning. As discussed earlier in this chapter (Section 2.1), research on the factors that affect children's learning highlight the importance of the interaction between the strength and weakness within the child, as well as the effects of teaching on children's SEN. Thus, it can be argued that alongside these indicators which are focused on measuring cognitive abilities within the child, it would be also important

to consider indicators that reflect the interactional nature of children's learning. One such indicator is the construct of cognitive modifiability.

#### **2.4.3. Cognitive modifiability in children with autism**

The term 'cognitive modifiability' was first used by Feuerstein (1979) as the central construct in his theory of structural cognitive modifiability (SCM), which views human organisms as open, adaptive and amenable to change. In his original description, Feuerstein refers to 'cognitive modifiability' as the:

*..changes in the state of the organism, brought about by a deliberate program of intervention that will facilitate the generation of continuous growth by rendering the organism receptive and sensitive to internal and external sources of stimulation (Feuerstein, 1979, pp. 9).*

The 'deliberate program of intervention' was described as a specialised interaction between a learner and mediator, involving two forms of interactions: 1) direct learning, i.e. learning through direct interaction between the learner and some environmental learning factor (such as, books, teaching materials); and 2) mediated learning, i.e. changes in direct learning by having the mediator intercede between the learner and environmental factors. The purpose of the mediator is to help the learner interact more productively with learning materials and then interpret the learner's response; and if necessary, modify their responses to increase their understanding.

The concept of 'cognitive modifiability' is central to dynamic assessment tests (Lidz & Elliott, 2000). Underlying this concept is the assumption that an individual's learning potential is not latent or static, but rather can be changed / enhanced with appropriate instruction. Thus, dynamic assessment is based on the link between testing and intervention, and aims to examine the processes as well as products of learning. By embedding learning in evaluation, dynamic assessment approaches assume what is tested is not just previously acquired knowledge, but the capacity to apply and reapply knowledge taught in the dynamic testing situation. The implication is that the level of cognitive modifiability observed in testing conditions is generalisable to other learning contexts.

Feuerstein et al. (1987) described cognitive modifiability in terms of changes in cognition that are applicable across as wide a range of functions as possible. However, other than this broad distinction, there is very little guide and clarity regarding the cognitive behaviours that are hypothesized to be modifiable (Elliott & Lauchlan, 1997; Sternberg & Grigorenko, 2002).

Resulting from the conceptual ambiguity, various researchers have used widely different conceptualisations and approaches to measuring cognitive modifiability (the different approaches are discussed in Chapter 4). In the literature, ‘cognitive modifiability’ is used interchangeably with many other terms, such as ‘learning potential’; ‘processing potential’; ‘dynamic assessments’; and ‘gains in learning’. Consequently, the approaches that are used to measure the construct circumscribe the definition of cognitive modifiability. For example, in approaches which focus on qualitative descriptions of children’s behaviour and the interaction between tester and testee is left open-ended, e.g. the LPAD<sup>5</sup> (Feuerstein et al., 1979), cognitive modifiability is conceptualised in terms of specific cognitive processes indicating improvements (or deficiencies) shown by the child in the respective learning-mediated activities. On the other hand, where the focus is quantitative and the interaction between tester and testee is more structured/controlled, e.g. Tzuriel’s Cognitive Modifiability Battery<sup>3</sup> (CMB: Tzuriel, 1995) or Swanson’s Cognitive Processing Test<sup>3</sup> (S-CPT: Swanson, 1996), cognitive modifiability is operationalised in terms of quantitative gains in performance, or amount of prompts/assistance that were needed for the child to achieve the higher (i.e. improved) performance. As a result, the robustness of the results from studies of cognitive modifiability is often limited, in that the findings from any one study have very limited generalisability and replicability across studies and research groups.

Despite the lack of conceptual clarity, there has been a considerable attempt to develop and use tests of cognitive modifiability with children from diverse

---

<sup>5</sup> These tests are discussed in Section 4.4.2.

populations. Its proponents argue that tests of cognitive modifiability (or dynamic assessments) 'offer (a) useful and even rich alternative to standardized normative tests' (Haywood & Tzuriel, 1992, pp56). These authors claimed that the main advantages of tests of cognitive modifiability are:

- They provide a 'fairer' assessment for children from disadvantaged backgrounds;
- They provide a 'more adequate' assessment of children with special educational needs; and
- They provide information that can 'provide guidance to teachers in drawing up appropriate intervention strategies' (Elliott, 2000, pp716); i.e. better instructional utility.

These claims are discussed in the next sections.

#### *2.4.3.1 A 'fairer' assessment for children from disadvantaged background?*

Proponents of the concept claim that assessment of children's learning which focuses on cognitive modifiability provide a better, i.e. 'fairer'<sup>6</sup>, gauge of children's abilities across a variety of population. However, although some studies have been cited in support of these claims, these tended to be based on case studies and exploratory intervention programmes that aim to promote improvements in a targeted area of functioning. There are very few studies on cognitive modifiability which are based on controlled experimental paradigms, for example:

- Early studies by Tzuriel and Feurstein (1992) with adolescents showed that children from a low socio-economic background achieved a higher level of performance in tests of cognitive modifiability, compared with their performance as indicated by a standardised psychometric assessment (i.e. IQ tests). Other studies (e.g. Tzuriel, 1996), suggested that the children's level of cognitive modifiability were dependent on children's social background. This was explained by the fact that children who had not been exposed to adequate learning opportunities in the past would benefit more from the

---

<sup>6</sup> 'Fairness' was seen in terms of the extent to which children from disadvantaged or minority backgrounds were not unduly penalized due to their lack of familiarity with the test contents.

mediation given within the dynamic assessment procedure than children who had relatively rich learning experiences.

- Hassles (1997) compared the results of IQ scores and cognitive modifiability (using the Learning Potential for Ethnic Minorities test, or LEM) of children from ethnic minority backgrounds (Turkish and Moroccan) in Dutch society, and found that the children with low (i.e. below average) IQ scores often show average or even above average cognitive modifiability (i.e. high LEM score). A stronger correlation was found between children's LEM scores and their progress in literacy and numeracy subjects than their IQ scores, which was taken as a basis to conclude that the measure of cognitive modifiability was '...a more valid means of assessment of minority children and may prevent children from being referred to special education inappropriately' (Hessels, 1997, pp121).

There are some methodological issues that limit the validity of the claims made in the above studies, the most problematic being the issue of practice effects. In the studies cited, children's progress or gains in performance after teaching were interpreted as evidence of modifiability. One issue is whether the changes in performance may be due to the repeated (and intensive) exposure to the intervention materials. Commentators noted that the lack of precautionary measures (e.g. controlling the effects of retesting by using a control population, i.e. unmediated sample) is a common methodological flaw in many of the studies on cognitive modifiability (Grigorenko & Sternberg, 1998).

Despite the methodological flaws, the main utility of measures of cognitive modifiability appears to lie in its interactional perspective, in that it claims to be able to accommodate the effects of inadequate environment on children's learning, and in so doing provide a fairer and more appropriate indicator of children's learning abilities. However, evidence supporting the practical applications of the concept of cognitive modifiability in the assessment of children with special educational needs

have been somewhat limited. As noted by several commentators (Grigorenko & Sternberg, 1998; Elliott, 2003) one of the main reasons for its failure to take root in mainstream practice is related to the lack of evidence for validity.

#### *2.4.3.2 A more 'adequate' assessment for children with SEN?*

It is often reported that when children with disabilities or SEN are tested with dynamic assessments, they achieve higher scores than when tested with standardised tests (Tzuriel & Klein, 1987; Resing, 2000; Tzuriel & Caspi, 1992). For example, in one study (Tzuriel & Caspi, 1992) comparisons were made between the performance of preschool children with hearing impairments and their matched controls on a dynamic assessment, namely the Children's Analogical Thinking Modifiability Test (CATM: Tzuriel & Klein, 1985) and a standardised IQ test, namely the Coloured Progressive Matrices (CPM: Raven, 1957). Both groups showed a higher percentage of correct responses for the CATM, and the post mediation performance of children with hearing impairments in the CATM were not significantly different from the controls. Based on these results, the authors conclude that the dynamic test (CATM) indicated higher 'learning potential' for both groups than the IQ test, and that children with hearing impairments benefited more from mediation than the hearing group.

However, there are several problems with these conclusions. Although the two tests share similar constructs, namely analogical reasoning skills, the CATM and the CPM are different tests with very different formats and difficulty levels.

Comparisons based on children's raw scores (i.e. percentage correct responses) in the tests is dubious; children's improved performance on the CATM may simply reflect lower item difficulty level. It would be more meaningful to compare the children's responses based on their normative scores of the tests, as these would reflect the extent to which the percentage of correct responses deviate from the expected performance of children in the same age group. However, such comparisons are not available as most tests of cognitive modifiability are not standardised and do not have adequate norms.

The higher post mediation scores of the children with hearing impairments in the dynamic assessment were attributed to the use of mediation strategies that were tailored to overcome children's communication difficulties in testing situations. This ability to individualise the testing instructions and methodology to suit children's specific learning needs is a unique advantage of tests for cognitive modifiability. By inference, the needs of children with other forms of communication difficulties, e.g. autism, could also be accommodated using some of the dynamic assessment approaches. In this thesis, a key issue is whether the performance of children with autism in these tests can better predict their ability to cope with the demands of mainstream school than traditional IQ tests. However, in the literature searchers conducted for this thesis, no empirical studies on the cognitive modifiability of children with autism have been located. The criterion validity of tests for cognitive modifiability as indicators of SEN is addressed in Study 2 (See Chapter 5).

Tests of cognitive modifiability have also been shown to distinguish children with SEN, from those with disadvantaged background (i.e. children with poor academic performance and support but without SEN). For example, children without SEN demonstrated high gain scores in the CATM (i.e. the difference between pre and post mediation scores), whereas children with severe SEN showed small insignificant gains (Tzuriel & Klein, 1987). In another study (Swanson, 2000), dynamic assessments of children's working memory (based on the S-CPT) indicated that children with specific learning difficulties (dyslexia) showed lower gains in performance with mediation, compared with children with poor reading (without dyslexia). The test appears to be able to differentiate children whose reading difficulties are attributed to specific deficits in working memory.

A key issue however is the incremental utility of information based on 'gain scores', i.e. do gain scores provide additionally meaningful information about children's functioning? Based on the two studies cited (i.e. Tzuriel & Klein, 1987; Swanson, 2000), the overall pattern of results from dynamic assessments appear to mirror the findings that are based on 'static' tests, i.e. that children with SEN/disabilities

demonstrate poorer performance in cognitive measures than those without SEN. Thus it is debatable whether children's performance after mediation provides unique information about children's cognitive functioning that is not currently available through other modes of assessment, e.g. standardized testing. The incremental utility of tests of cognitive modifiability is evaluated in Study 3 (See Chapter 6).

#### *2.4.3.3 'Better' instructional/treatment utility?*

Commentators noted that the potential value in the assessments of cognitive modifiability lies in the information that can be used to guide intervention (Lidz & Elliott, 2000; Sternberg & Grogorenko, 1998; Elliott & Lauchlan, 1997). There have also been some attempts to clarify the link between the construct of interest, i.e. cognitive modifiability and specific intervention strategies: for example, Lidz's work (1991) which demonstrated some ways in which psychologists can draw upon clinical observations to gain better understanding of the child's difficulties and provide advice on intervention strategies; and Swanson's (1995) work which highlighted how information regarding children's responsiveness to cues and probes during mediation can be used to identify children that would benefit from active learning strategies (e.g. reciprocal teaching), from those that may be more suited for interventions based on drill and practice. However the application of these ideas in mainstream educational practice seems limited. For example, despite the claims of instructional utility, when information on children's cognitive modifiability was given to teachers, it was found the information had little impact on their classroom practice (Lauchlan & Elliott, 1997). The issue of treatment utility of tests for cognitive modifiability is addressed in Study 3 of the present thesis.

#### **2.4.4. Conclusion**

One objective in this thesis is to identify cognitive indicators that can be used in the assessment of SEN for children with autism. A review of the causal theories of autism has highlighted several indicators which may underlie the deficits in the learning of children with autism, namely theory of mind, central coherence and executive function skills. While these indicators are meaningfully linked to possible

causal explanations of autism, they are focused on deficits within the child, and thus do not reflect the interactional perspective of SEN. It was argued that, alongside these indicators, it would be useful to consider the aspect of cognitive modifiability, which focuses on the extent to which children's learning is changed or enhanced with appropriate instruction.

One of the issues that will be investigated in this thesis is the extent to which these indicators can best predict the special educational needs of children with autism. However, prior to this evaluation, consideration needs to be given to the relevant issues in the assessment of children with autism, i.e. the range of methods and assessment approaches that are particularly useful for children with autism. These issues are reviewed in the next section.

## **2.5 ASSESSMENT APPROACHES FOR CHILDREN WITH AUTISM**

A comprehensive assessment of the functioning and difficulties of children with autism should ideally include all aspects of the triad of impairments that characterise autism (i.e. social-interaction, imagination and communication skills), the child's adaptive functioning and other aspects which may affect the child's ability to effectively participate in daily activities, including cognitive skills, maladaptive behaviours, motor coordination, responses to sensory stimuli and educational attainment (Schopler & Mesibov, 1988). In addition, consistent with the interactional perspective of SEN, assessment should also include evaluations of the impact of the environment on the child's functioning. Commentators on the assessment instruments for autism conclude that given the developmental nature of autism and these children's sensitivity to changes, it is important to sample the behaviours over time and across multiple contexts (Klin et al., 1995). Assessment tools that have been specifically designed for individuals with autism can be grouped into two categories: firstly, instruments that rely on direct observation of the child's behaviour (i.e. behavioural observation scales); and secondly, behavioural checklists and investigator-based interviews, which rely on parents' or caregivers' reports.

### **2.5.1 Behaviour observation scales**

The various behavioural observation scales for autism can be grouped into two categories: firstly, structured observation scales, which specify the contexts or conditions in which the observation should be conducted, e.g. the Autism Diagnostic Observation Schedules or ADOS (Lord, Rutter & DiLavore, 1989); and secondly, those which rely on the clinician's unstructured observation of the child, e.g. the Childhood Autism Rating Scale or CARS (Schopler, Reichler, DeVellis, & Daly, 1988). The open-ended, unstructured behavioural scales place less demand on the clinician's ability to set up or create the necessary conditions; this flexibility has increased their utility in practice. The CARS, for example, is reported to be the most widely used measure for autism in the United States (Lord, 1995). However, the lack of contextual control implies that the range of information that could be captured depends on the available opportunity for direct observations. On the other hand, the more structured observation schedules enable information about children's behaviour to be 'collected in a positive but standard context' (Lord, 1995, pp 477). However, the artificial way in which the context is controlled/standardised may limit the range of behavioural repertoire observed. Most of these instruments also tended to focus on deviant behaviour, i.e. atypical characteristics of autism, and hence may be more suitable for diagnostic purposes, rather than as a measure of developmental functioning. Due to these limitations, parental reports appear to be the more favoured method for obtaining information on the child over time and across contexts (Lord, 1995; Klin et al., 1997).

### **2.5.2 Parental reports**

The most common type of instrument used to obtain parental reports are checklists or structured questionnaires. These involve carefully selected questions, pre-determined and standardised, with codings<sup>7</sup> based on whether the respondent reports the presence/frequency of the described behaviours. Many autism

---

<sup>7</sup> Examples of codings are : a) 'yes' or 'no'; or b) a rating scale involving 'never', 'sometimes', or 'frequently'.

checklists/structured questionnaires have been developed in the past 10 years; they are the most widely used method for gathering developmental history and identifying the presence of abnormality or delays in adaptive skills (Lord, 1995). The main advantage of checklists/structured questionnaires is the ease of use: they can be completed by parents in their own time, and much information can be quickly obtained. For example, the widely used Gilliam Autism Rating Scale or GARS (Gilliam, 1995) takes only 10 minutes to complete. In addition, the administration of checklists/structured questionnaires does not require much training. The reliability index (i.e. internal consistency, inter-rater and test-retest reliability) of such checklists are known to be high (Lord, 1995). However, their validity remains a key issue; for example, the GARS checklist tended to under-diagnose children with autism (South, Williams, McMahon, Owley, Filipek, Shernoff, Corsello, Lainhart, Landa, & Ozonoff, 2002), and as such, it may be more suitable as a screening tool to gauge the probability that a child might have autism, rather than as a diagnostic instrument.

The validity of checklists rests on the assumption that all respondents are interpreting the questions in the same way and all have the necessary conceptual understanding to make the distinctions necessary for coding each behaviour. While this may be true for questions about basic self-care, the assumption becomes questionable when dealing with subtle differences in behaviours related to communication and social relationships (which are central to the diagnosis of autism). For example, one of the items in the GARS checklist asked whether the child “uses gestures instead of speech to obtain objects”. To answer the question correctly, the respondents must be clear about three main concepts: firstly hand gestures are deliberate movements of the hands to communicate meanings and hence do not include hand-flapping by the child when he/she is excited or anxious; secondly the gesture is not accompanied by speech, as the focus is on gestures that are used to replace, and not augment speech; and thirdly the gestures are used in making requests for objects. As argued by Le Couteur, Rutter, Lord, Rios, Robertson, Holdgrafer, and McLennan (1989), it is difficult to assume that all

parents have the conceptual knowledge to understand such subtle distinctions in complex social behaviours. This was the main rationale for researchers to adopt and develop *investigator-based interviews* in the diagnosis of autism. This approach will be reviewed in the next section, with particular reference to two examples, namely the *Autism Diagnostic Interview* (Le Couteur et al., 1989), which has been described as the ‘gold-standard’ for the diagnosis of autism (Lord, 1995), and the newly developed *Diagnostic Interview for Social and Communication Disorder* (Wing, 1999), which has been described as a possible ‘alternative’ to the ADI (O’Brien Pearson, Berney, & Bernard, 2000).

### **2.5.3 Investigator-based interviews**

The *Autism Diagnostic Interview or ADI* (Le Couteur et al., 1989) was first developed as a research tool to distinguish between deviance in behaviour that is characteristic of autism, from global developmental delay. However, since its development, its use has expanded to clinical practice. It includes a detailed coverage of three main areas of functioning: 1) social reciprocity; 2) communication; and 3) restrictive repetitive behaviours. In the last 10 years, the ADI has been well used in both research and clinical practice (Lord, 1995).

In contrast to the ADI, the *DISCO, or Diagnostic Interview for Social and Communication Disorder* (Wing, 1999), was first developed to meet the practical needs of clinicians but was also recommended for use in research. It was intended to “assist clinicians in the diagnosis, differential diagnosis and management of autism spectrum disorders affecting social interaction and communication” (Wing, Leekam, Libby, Gould, & Larcombe, 2002, pp 307). It contains a comprehensive list of over 200 items covering four categories of behaviour: 1) behaviour in infancy; 2) onset of recognition of problems; 3) developmental behaviours; and 4) atypical behaviours.

### *2.5.3.1 Features of investigator based interviews*

There are key features of investigator based interviews, common to both the ADI and DISCO, which set them apart from the more conventional autism checklists/questionnaires, namely their coding structure, emphasis on detailed examples, and use of trained interviewers. These features are best illustrated by comparing sample items from an autism checklist (i.e. GARS) with similar items in investigator-based interviews (i.e. ADI & DISCO) (See Table 2.5.3.1).

### *2.5.3.2 Structure of coding*

In contrast to checklists, where the ratings for an individual's behaviour are entirely determined by the respondents/parents, in investigator-based interviews the coding is determined by the investigator/interviewer. The responsibility is placed on the interviewer to obtain all the necessary information in order to make each rating.

There are a variety of specified screening questions or probes, but their purpose is not to obtain yes/no answers from the respondents, but rather to guide the interviewer on the nature of the information to be obtained. Based on the information obtained from parents/caregivers the interviewer makes a series of judgment as to whether the behaviours are present, and if so, in how marked a form.

### *2.5.3.3 Emphasis on detailed examples*

There is also an emphasis on the need to obtain detailed descriptions of actual behaviour; general statements by respondents are not acceptable. For example, if the parent responded 'yes' to the question '***Does the child show other people what she wants by gesture?***'(See Table 2.5.3.1, DISCO item xi2), he/she will be probed to give an account of the child in actual incidents/settings that reflects the target behaviour. It is based on these descriptions of actual behaviours that the interviewer makes a judgment on the coding that would be most applicable.

Table 2.5.3.1 : Comparison of a)GARS, b)ADI & c)DISCO items on the use of gestures

<p><b>A) GARS – item 27 – Communication (non-verbal)</b>  <b>Uses gestures instead of speech to obtain objects.</b>  <u>Coding (by parent/caregiver)</u>  0 = Never observed; 1 = Seldom observed; 2 = Sometimes observed; 3 = Frequently Observed.</p>
<p><b>B) ADI – Item 31 – Conventional / Instrumental Gestures</b>  <i>Does (child) wave good bye? When does this happen? Does he/she ever use other common gestures, such as blowing a kiss, clapping for a job well done, putting a finger to his/her lips to mean 'be quiet', or shaking a finger for 'bad'? Does he/she ever use gestures other than pointing to let you know what she/he wants? Does he/she use gestures when he/she is trying to get help or to get your attention (for example, beckoning to someone or putting out a hand with his/her palm extended to ask that you give him/her something)?</i>  <u>Coding (by interviewer)</u> <ol style="list-style-type: none"> <li>0. Appropriate and spontaneous use of a variety of conventional or instrumental gestures</li> <li>1. Spontaneous use conventional or instrumental gestures, but limited in range and/or contexts</li> <li>2. Inconsistent spontaneous use of a variety of conventional or instrumental gestures</li> <li>3. No use of conventional or instrumental gestures</li> </ol> </p>
<p><b>C) DISCO – Items xi2 to xi4 – Communication (Nonverbal) – Use of gestures</b></p> <p><b>xi2 Does the child show other people what s/he wants by gesture (imperative gestures)?</b>  <u>Coding (by interviewer)</u> <ol style="list-style-type: none"> <li>0. No imperative gestures</li> <li>1. Takes people by the hand or arm to designate an object he/she wants</li> <li>2. Reaches for and touches to a desired object (Typically develops by age 9 mths)</li> <li>3. Points to a desired object (Typically develops by age 1 yr)</li> </ol> </p> <p><b>xi3 Does child try to affect other people's behaviours by gesture (instrumental gestures)?</b>  <u>Coding (by interviewer)</u> <ol style="list-style-type: none"> <li>0. No instrumental gestures</li> <li>1. One or two gestures indicating 'come here', 'go away' (Typically develops by age 2 yrs)</li> <li>2. Uses a range of such gestures fluently, such as wagging a finger to show disapproval.</li> </ol> </p> <p><b>xi4 Does child uses gestures to describe objects or actions, e.g. indicating size (descriptive gestures)?</b>  <u>Coding (by interviewer)</u> <ol style="list-style-type: none"> <li>0. No descriptive gestures</li> <li>1. Some very simple descriptive gestures just beginning (Typically develops by age 2 yrs)</li> <li>2. Can indicate simple ideas, eg could mime 'drinking' if thirsty (Typically develops by age 3-4 yrs)</li> <li>3. Can indicate properties of object in detail, eg size, shape (Typically develops by age 7 yrs)</li> </ol> </p>

#### 2.5.3.4 Trained interviewers

This interview style is heavily reliant on skilled interviewers who have been trained, in addition to general interviewing skills, in the conceptual distinctions involved in

each and every one of the ratings in the interview schedule. In both the DISCO and the ADI, the types and method of probing are left open to the judgment and skills of the interviewers.

#### **2.5.4 Comparison of the ADI and the DISCO**

Although both instruments share the same methodology, the DISCO differs from the ADI in several ways. While there are superficial or ‘surface’ differences in the way questions are structured in the two instruments, there are two fundamental differences between the ADI and the DISCO:

##### *2.5.4.1 Single VS Multiple Purpose*

The DISCO aims to go beyond autism and diagnose other “developmental disorders on the borderline of the spectrum (of autism), a psychiatric disorder or any combination of these conditions. It is also intended as a guide for recommendations concerning education, occupation, leisure and care” (Wing et al, 2002, pp. 310). Consequently, items in the DISCO cover a wider range of developmental behaviours than the ADI. The developmental domains in the DISCO which are excluded in the ADI are: self-care; independence; visual-spatial and other cognitive skills (e.g. memory). The DISCO also includes greater details about atypical responses to sensory stimuli, motor stereotypies, repetitive routines and emotional disturbances, and has sections on catatonia, psychiatric disorders, difficulties in sexual behaviours and forensic problems. The ADI, in contrast, focuses primarily on the diagnosis of autism and excludes developmental indicators that are not directly relevant for the diagnosis.

##### *2.5.4.2 Categorical VS Dimensional Concept of autism*

The autism algorithm in the ADI is based strictly on the diagnostic criteria set out in the DSM-IV (APA, 1994). As discussed in the Section 2.4.1.2, the DSM-IV classification framework is categorical, in which autism is defined as a disorder that is distinct from other pervasive developmental disorders. In contrast, the DISCO is based on the dimensional framework of autistic spectrum disorder (or ASD), which

includes all categories of pervasive developmental disorders as defined in DSM-IV. This difference in the conceptual framework for autism between the DISCO and the ADI could result in different diagnostic outcomes for the same child. For example, based on the DISCO, children who would otherwise be classified as 'not-autistic' according to DSM-IV criteria could be diagnosed as having 'autism spectrum disorders', such as children with Aspergers' Syndrome and Rett Disorder.

#### *2.5.4.3 Assumptions*

There are several assumptions underlying the use of the investigator-based interviews. While its designers acknowledge and indeed stress that the reliability of the instruments is dependent on trained and highly skilled interviewers, other assumptions are less explicit.

For example, both the ADI and the DISCO rely on the assumption that all parents/caregivers are equally able to provide detailed descriptions of the child's relevant behaviours. The greater the specificity and details included in the description, the greater is the accuracy of the codings for the target behaviours. However, research has shown that parents of children who have received a clear diagnosis are more able to report and recall details of the child's behaviour compared to parents of children who did not have a diagnosis (Schopler & Reichler, 1972). Thus whether a child has had a clear diagnosis might affect the reliability of information obtained through the interviews.

The ADI and the DISCO also makes similar assumptions about the context of the interviews. In practice, there are often high-stakes outcomes of a clinical diagnosis. As highlighted by Zoppi and Epstein (1999), during interviews in clinical or medical settings, it cannot be assumed that the interviewer's (i.e. clinicians') and interviewee's (i.e. patients') agenda are similar. For example, in cases where the child's parents are keen to confirm a diagnosis in order to secure maximal additional provisions for the child, they may accentuate reports of the child's behaviours that they believe to be characteristics of autism. This would affect the validity of the

responses obtained from the interviews, or more specifically, the validity related to response processes (as discussed in Section 2.2.2.2)

#### *2.5.4.4 Reliability issues*

In the development of both the ADI and DISCO, careful attention was given to establishing inter-rater reliability, i.e. the degree to which different raters concur when using the same instrument. The method for establishing inter-rater reliability were similar for both the ADI (Le Couteur et al., 1989) and DISCO (Wing et al, 2002), where second raters coded the target behaviours based on video-taped interviews conducted by the first rater. Overall, the inter-rater agreement in both the ADI and the DISCO were high. However, a higher proportion (80%) of items in the DISCO reached the ‘excellent agreement’ levels of (kappa values  $>0.75$ ), compared with only 40% of ADI items. The inter-rater index for the DISCO appears to have exceeded that of the ADI and it can be argued that the greater specificity in the DISCO items contributed to the higher inter-rater reliability.

#### *2.5.4.5 Validity Issues*

One aspect of validity has been studied in both the ADI and the DISCO is that of discriminant validity, i.e. the extent to which the instrument is able to distinguish groups of individuals with autism from others with different clinical conditions. Most commonly, this is achieved by comparing the child’s diagnosis based on the use of the instrument (i.e. algorithm diagnosis) with pre-existing clinical diagnosis of the child. The assumption in this approach is that the pre-existing clinical diagnosis is a ‘true’ indicator of the child’s condition.

In the ADI validity study (Le Couteur et al., 1989) children with autism were compared with children with mental handicap (without autism). Results showed that although some of the children with mental handicap met one or two criteria for autism, none of the children from this group met all the 3 criteria for autism. It should be noted that in the ADI algorithm, diagnosis of autism is made only when

the child meets all the 3 criteria. The ADI appears to be able to clearly distinguish children with autism and mental handicap, with no overlaps in the two groups.

However, there is a problem in the design of the ADI study. The comparison group, i.e. children with mental handicap, differed in many other aspects that are not related to autism, most significantly in the level of intellectual functioning. Any differences between the two groups could also be due to other variables, which may not be related to autism. It is therefore important to compare the results of the group with autism with samples of control children with comparable general ability levels. One way to achieve this comparison would be to match the sample according to IQ levels. Alternatively, the comparison could be based on sub-groups of the sample of children with autism according to IQ levels. This latter method was employed in the DISCO validity study (Leekam et al., 2002).

In the DISCO study (Leekam et al., 2002), comparisons were made between three sub-groups of children: 1) children with autism with high nonverbal IQ; 2) children with autism with low nonverbal IQ; and 3) children with learning disability and language disorder. In addition, adjustments were made for differences in verbal and non-verbal ability in the groups by the use of logistic regression analysis. Overall, the results of the DISCO validity study were favourable, i.e. a significantly higher proportion of children with autism (both in the high and low functioning sub-groups) meet the algorithm for autism, compared with children who have learning disabilities and language disorders.

#### *2.5.4.6 Lack of construct equivalence*

As the ADI and DISCO were based on different conceptualisations of autism, there is a lack of construct equivalence in the two instruments, i.e. they are not measuring the exact same trait/characteristics. While the ADI is based on the more narrow categorical definition (as used in DSM-IV), the DISCO is based on the dimensional concept of ASD. It could be argued that, contrary to claims by O'Brien et al. (2000) it may not be appropriate to use the DISCO as an “alternative” to the ADI. Instead,

users need to base their selection on the appropriateness of the instrument to the purpose of assessment and the definition or construct of autism that is used.

### **2.5.5 Summary on investigator-based interviews**

Investigator-based interviews, such as the ADI and the DISCO, were developed and used in the diagnoses of autism as a way of improving the reliability and validity of data obtained from parental reports of children's behaviour and developmental functioning. Provided that the assumptions behind their use hold true, investigator-based interviews can potentially be effective in capturing data over time and across contexts. Compared with the more conventional measures of obtaining parental reports, such as the use of checklists, investigator-based interviews have the advantage that they do not assume that parents have the conceptual understanding to distinguish complex social behaviours. Reliability studies on the ADI and the DISCO demonstrated that overall, the inter-rater reliability of investigator-based interviews is high. Comparatively, the DISCO's overall inter-rater agreement showed a marked improvement over the ADI. Discriminant validity studies of the ADI and the DISCO used different comparison groups. The design for the DISCO validity study enabled a more fine-grained differentiation between children with autism with high and low ability, and children with other types of learning disabilities.

In the present thesis, the objective of Study 1 (reported in the next chapter) is to obtain data to ascertain whether the ICF, which was developed as a generic tool for all health condition, meets the stated criteria of reliability and validity. One of the main aims is to discover whether measurements of children's functioning based on the ICF concur with independent measures of functioning developed specifically for children with autism. Given that the focus of assessment in Study 1 is the measurement of children's levels of functioning and not diagnosis, the DISCO is arguably a more appropriate tool. Comparison between measures of children's functioning based on the ICF and the DISCO will be used to guide decisions about

whether the ICF has the sensitivity to capture differences in functioning in children with a specific disability, i.e. autism.

## **2.6. Summary and conclusion for Chapter 2**

The aim of the present thesis is to identify valid and reliable methods of assessing special educational needs (SEN) for children with autism. It was argued that the interactional perspective, in which both within-child and environmental elements are seen as contributory factors to children's functioning and disabilities, provides a useful framework for defining and conceptualising special educational needs. The ICF (WHO, 2001) is one of the few available resources that provide both a conceptual framework and an assessment tool which reflects this interactional perspective. However, as it has been developed as a generic tool for all health conditions and validated mainly for use with individuals with physical disabilities, its applicability for children with autism needs to be ascertained. This validation is the focus of Study 1.

The ICF relies on information obtained from one source/method, i.e. clinical interview. As the process of psychological assessment implies the use of multiple sources of information, obtained from a variety of different methods, it is necessary to identify possible indicators of SEN that can be used alongside the ICF. These indicators focus on specific dimensions in children's functioning that reflect the core impairments associated with autism. A review of the causal theories of autism highlighted aspects of cognitive function which are possible indicators of the learning impairments of children with autism, namely theory of mind, central coherence and executive function. While there is theoretical and empirical support for these constructs as reflecting the core impairments in children with autism, their use as indicators of SEN has not been evaluated.

In addition, the difficulties face by children with autism in applying or generalising their learning across contexts suggest that a key impairment may stem from the children's difficulties in benefiting from training, rather than a deficit in a specific

cognitive area. This aspect of the transfer of learning, known as cognitive modifiability, is considered as another possible indicator of SEN in children with autism.

Using the *Standards* (APA, AERA & NCME, 1999) framework, the validity evidence on these indicators will be compared with the SEN indicator that is currently used in Singapore, namely measures of intelligence (the issues of using measures of intelligence, or IQ scores, as an indicator of SEN is discussed in Chapter 4).

The next chapter reports the findings of the first study, which aimed to obtain evidence of reliability and validity of the ICF as an independent measure of SEN for children with autism.

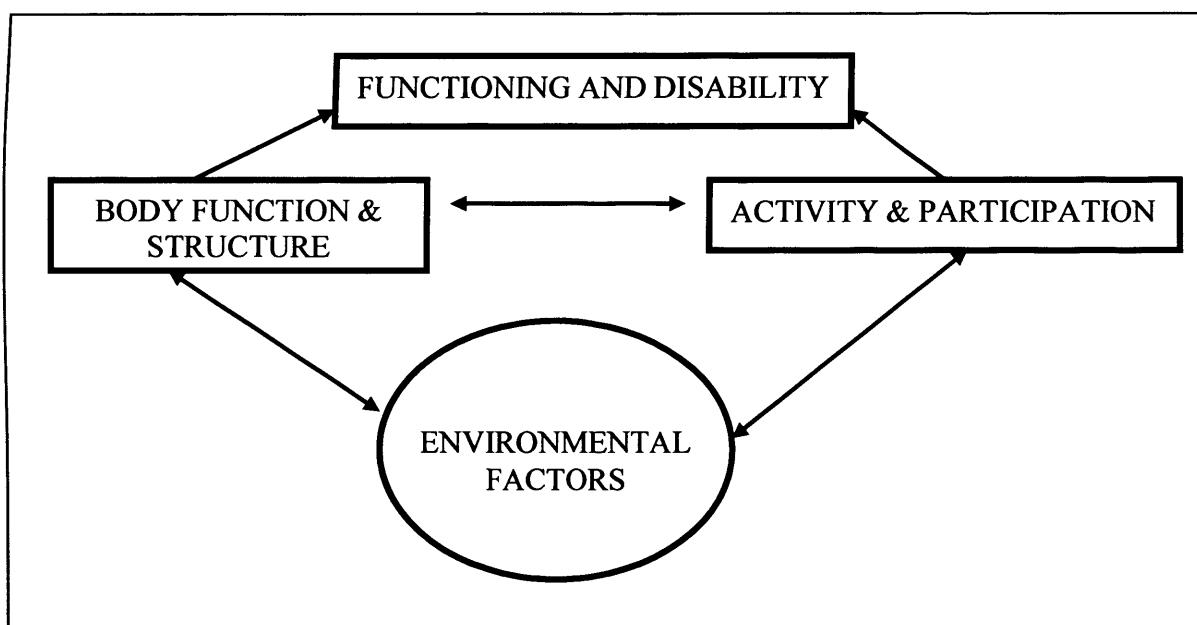
## CHAPTER 3

### STUDY 1 : RELIABILITY AND VALIDITY OF THE INTERNATIONAL CLASSIFICATION OF FUNCTIONING, DISABILITY AND HEALTH (ICF) FOR CHILDREN WITH AUTISM

#### 3.1 INTRODUCTION

The focus of the thesis is on the assessment of special educational needs (SEN) of children with autism. In the previous chapter, concepts of SEN were reviewed. It was argued that the interactional perspective best reflects current knowledge on the nature of children's special educational needs, i.e. that it is influenced by both within child and environmental factors. Among the various frameworks that reflect this interactional perspective, the ICF (WHO, 2001) model of disability and functioning is, arguably the most comprehensive. It is also one of the few of such frameworks that are accompanied by a parallel assessment tool, i.e. the ICF checklist.

Fig 3.1. : ICF Framework for Functioning and Disability



Based on the ICF framework (See Fig. 3.1), the concept of SEN used in the present study is defined as follows:

Special educational needs refers to the limitations in activity and participation, as a result of impairments in body functions and structures, or environmental factors, where:

- Body Function refers to physiological functions of body systems (including psychological functions);
- Body structure refers to anatomical parts of the body such as organs, limbs and their components;
- Impairment refers to problem in body function or structure such as significant deviation or loss;
- Activity refers to the execution of a task by an individual;
- Participation refers to involvement in a life situation;
- Limitation refers to difficulties an individual may have executing activities or problems an individual may experience in involvement in life situations;
- Participation restrictions are problems an individual may experience in involvement in life situations;
- Environmental factors make up the physical, social and attitudinal environment in which people live and conduct their lives.

However, before the ICF can be considered as an adequate measure of SEN, it must demonstrate properties that meet acceptable standards of reliability and validity. The Standards for Psychological and Educational Tests and Assessments, or *Standards* (APA, AERA & NCME, 1999), was reviewed in the previous chapter, and based on the discussions, the following aspects of evaluation were identified:

- Evidence based on content;
- Evidence based on response processes or test bias;
- Evidence based on external criteria;
- Evidence based on treatment validity; and
- Evidence of reliability.

### 3.1.1 Evidence based on content

As discussed in section 2.3.7, one of the strengths of the ICF is its comprehensive coverage of behaviours related to 23 domains of functioning and disability (see Table 3.1.1).

Table 3.1.1 : The ICF Functional Domains

<b>BODY</b>	
<b>Function :</b>	<b>Structure :</b>
Mental functions	Structure of the Nervous System
Sensory Functions and Pain	The Eye, Ear and Related Structures
Voice and Speech Functions	Structures Involved in Voice and Speech
Functions of Cardiovascular, Haematological, Immunological and Respiratory Systems.	Structure of the Cardiovascular, Immunological and Respiratory Systems
Functions of Digestive, Metabolic, Endocrine Systems	Structure Related to the Digestive, Metabolic and Endocrine Systems
Genitourinary and Reproductive Functions	Structure Related to Genitourinary and Reproductive Systems
Neuromusculoskeletal and Movement-Related Functions	Structure Related to Movement.
Functions of the Skin and Related Structures.	Skin and Related Structures
<b>ACTIVITIES &amp; PARTICIPATION</b>	
Learning and Applying Knowledge	
General Tasks and Demands	
Communication	
Mobility	
Self-Care	
Domestic Life	
Interpersonal Interactions and Relationships	
Major Life Areas	
Community, Social and Civic Life	
<b>ENVIRONMENTAL FACTORS</b>	
Products and Technology	
Natural Environment and Human-Made Changes to Environment	
Support and Relationships	
Attitudes	
Services, Systems and Policies	

The content validity of the ICF can be traced to the extensive consultation process undertaken by its developers over five years; a total of 38 experts from 15 countries provided inputs to the item development of the ICF. Several refinements of the original ICIDH were made before the ICF was finalised in 2001 (Ustun et al., 2002).

The rigour in the process of content validation was motivated by the intention for the ICF to be used as a tool for all disability conditions ‘across (all) cultures’ (WHO,

2001, pp7). This is reflected in the way items are defined to include as many culturally specific examples as possible. For example, under the domain ‘Self-Care’, ‘Dressing’ is defined in the ICF as follows:

*Carrying out the coordinated actions and tasks of putting on and taking off clothes and footwear in sequence and in keeping with climatic and social conditions, such as putting on, adjusting and removing shirts, skirts, blouses, pants, undergarments, saris, kimono, tights, hats, gloves, coats, shoes, boots, sandals and slippers. (WHO, 2001, pp151).*

For the purpose of the present thesis, a crucial aspect of the content of the ICF is that it is one of the few assessment tools that incorporate the elements of both within-child and environment factors in its item sampling. It can be noted that the environmental factors sampled in the ICF not only pertain to physical environments (e.g. products and technology, natural and human-made changes to the environment), but include the social environment as well (e.g. support, relationships and attitudes of significant individuals as well as the society/community towards the individual with disability). In addition, ‘macro’ aspects of the environment, such as services, systems and policies, are also included. Thus, it can be argued that the items sampled in the ICF are not only comprehensive; it is also coherent with the theoretical framework of SEN that is used in the present study, i.e. the interactional perspective.

### **3.1.2 Evidence based on response processes**

Analyses of response bias involve obtaining evidence that firstly, the types of responses used fit the intended construct, i.e. special educational needs of children with autism; and secondly, there is consistency in the way judges/interviewers collect and interpret the data (see discussion in Section 2.2.2.2). The developers of the ICF checklist recommended “the use of multiple data sources such as self-report, medical examinations, clinical records, report from family members etc.” (CAS, WHO, 2002, pp 3). However, there were no guidelines on how information from different sources should be combined. Thus, if there were a mismatch between parent’s reports and direct observation of the child’s behaviour, the ICF provides no guidelines for resolving the conflicting information.

As discussed in Section 2.5, in the assessment of functioning for children with autism it is important to obtain information of the child across time and multiple contexts. For this purpose it was argued that parental reports of the child behaviour would be more appropriate and feasible than direct observations. Thus for the present study, only one source of information will be used, i.e. parent's reports.

Based on the review of autism instruments (section 2.5), it was argued that the most effective method for obtaining parental reports of the behaviour of children with autism is investigator-based interviews; unlike conventional interview or checklists, they do not assume that parents have the conceptual understanding to distinguish complex social behaviours. The aim of investigator-based interviews is not merely to obtain 'yes' or 'no' answers from respondents, but to obtain information or reports on behaviours, through questioning or probing, that will enable the interviewer to make a judgment about the existence of an impairment, and if so, to what level of severity. This method is also recommended by the developers of the ICF:

*In most cases, an interview schedule will be applied when filling the checklist. The first set of questions should be asked in the same way, followed by a series of probes which the clinician is free to use or make up his/her own probes. The idea is to match the report given by the respondent with the examiners' glossary definition provided in the ICF manual, and then the examiner makes a judgment taking multiple sources of information into account. (CAS, WHO, 2002, pp 3).*

However, the problem of leaving the clinician 'free to use and make up his/her own probes' may reduce the consistency in the quality of questions used during the interview. A standardised interview protocol would be able to overcome this problem. For the purpose of the present thesis, an interview schedule, or protocol, for the investigator-based interview was developed (See Appendix C). This entailed developing the specific questions and relevant probes that can be used to elicit parents' description of the target behaviours defined in the ICF manual.

The ICF is not a normative test, and as such there are no norms for the checklist. However, in deciding the ICF ratings for individual behaviours, in particular, impairments in developmental skills, such as language, literacy and mathematics, the clinician is required to make judgements on the extent of a child's delay. It is assumed that the user of the ICF has access to (or knowledge of) developmental milestones for the relevant population. The ICF developers recognised this problem and acknowledged that 'as ICF based population norms and cross mapping with assessment instruments are not yet available, the ICF checklist user has to use clinical judgment while using the scale' (CAS, WHO, 2002, pp 4). However a problem in leaving evaluation criteria to 'clinical judgments' is that consistency of judgments made across domains and across cases may be compromised. One-way to overcome this problem is to specify, *a priori*, the relevant developmental norms that can be used as the basis for interviewers' judgement.

### **3.1.4. Evidence of reliability**

As discussed in Section 2.3.7, inter-rater reliability of the ICF was evaluated based on live case evaluations conducted by health professionals in 15 collaborating centres involving 1884 patients with physical health conditions, such as musculoskeletal disorders (CAS, WHO; 2002). However, no information could be obtained regarding the breakdown for these cases in terms of the specific disability conditions and ages of the patients. Although the study concluded that high agreement for all ICF domains were found, this cannot be generalised for children with autism. It has not been possible to find reports of studies which investigated the internal reliability of the ICF checklist. This may be due to the fact that the second phase of the ICF field trials is currently ongoing (Ustun et al., 2002). Thus, evidence of internal and inter-rater reliability for the use of the ICF for children with autism would need to be sought.

### **3.1.5 Evidence based on external criteria**

As discussed in Section 2.3.7, preliminary reports indicated that for musculoskeletal conditions, the ICF checklist was able to capture the profile of patients in all four

conditions, and good concurrent validity was reported between the checklist and the SF 36 Body Function subscale (CAS, WHO; 2002). However, there are no reported plans for testing the reliability and validity of the ICF for use with children with developmental disabilities, such as autism. Thus, validity evidence specific to children with autism will need to be obtained.

### **3.1.6. Evidence based on treatment validity**

As the ICF is very new and it has not been used in practice (most of the reported use of the ICF has been for research purposes) (Ustun et al., 2002), it would be premature to evaluate the treatment validity of the ICF. However, commentators have been very positive about its clinical potential (McLeod & Bleile, 2004; Simmeonsson et al., 2003; Brundtland et al., 2002).

### **3.1.7 Objectives of the present study**

In summary, analyses of content validity of the ICF indicated that items sampled in the checklist are comprehensive and adequately reflect the interactional framework. To minimise response bias, a structured interview protocol will be developed to elicit parents' reports of children's behaviour. This protocol will be in the format of an investigator-based interview, and will include the relevant developmental milestones for children in Singapore. For the purpose of the present thesis, information obtained from reported studies on the reliability and criterion validity of the ICF are inadequate. In order to be considered as an adequate measure of SEN in children with autism, the ICF must be able to meet the following criteria:

1. ICF ratings of individual children must show consistency between items within the same component, as well as consistency in the evaluations of the same children by different raters (i.e. internal consistency and inter-rater reliability).
2. ICF measures of children's functioning (i.e. impairments in body function/structures, activity limitation and participation restriction) must be able to meaningfully distinguish children with varying levels of special educational needs (i.e. criterion validity); and

3. There must be concurrence between evaluations of children's functioning based on the ICF, with other independent measures of functioning, in particular, those which have been developed specifically for the children with autism (i.e. criterion validity).

For the first criterion, two aspects of reliability will be tested, i.e. inter-rater reliability and internal consistency reliability. Internal consistency reliability, which focuses on the consistency of ratings for individual children within each ICF component, will be gauged by measuring the correlation between individual item ratings and children's overall scores. Inter-rater reliability will be measured by estimating the correlation in the rankings of individual children across different raters, as well as agreement of different raters on individual ICF items.

The second criterion focuses on the convergence of evaluations based on the ICF with a pre-existing, independent index of special educational needs. For the purpose of the present study, one possible pre-existing measure of special educational needs is the child's current educational placement, i.e., whether the child is in a mainstream or autism-specific special school. As discussed earlier in Section 1.1, in the Singapore context, children with autism remain in mainstream if their difficulties are not severe, while children are referred to special schools if they show severe difficulties of functioning. Thus, one indicator that reflects a child's level of special education needs in the Singapore context is whether he/she is placed in a special school. Special schools in Singapore are characterised by a higher level of individual support and specialised provisions that are tailored/adapted to meet the child's special needs, such as purpose built facilities, modified curriculum, and access to therapists and trained teachers. To investigate the second criterion, two groups were chosen for the study, i.e. children with autism from mainstream and special schools. These children represent two groups of children, which by this criterion, clearly differ in their functioning and needs levels. It is predicted that based on their scores on the ICF checklist, children with autism in special schools will show greater

impairments in functioning, greater limitations in activity and participation, as well as higher levels of environmental support.

For the third criterion, the aim was to see the extent of convergence in children's scores based on the ICF and their scores obtained from another instrument which measures a similar domain. Based on the review of established assessment instruments for autism (Section 2.5) the DISCO was considered to be the most appropriate tool for this purposes. There is one domain common to both the ICF and the DISCO, namely the measure of children's developmental functioning and disabilities. In the ICF, this is reflected in the scores of Functioning and Disability (See Fig. 3.1), while in the DISCO, this is reflected in the ratings for items measuring current developmental functioning and atypical characteristics. If the ICF has the sensitivity to reflect functioning levels of children with autism, children's total ratings on Functioning and Disability would correlate significantly with the scores on the DISCO.

### **3.2. METHODOLOGY**

#### **3.2.1 Participants**

A total of 40 children from Singapore, aged 5:0 years to 11:11 years with pre-existing diagnosis of autism, and who were in either mainstream schools (Group 1, n=21) or special schools (Group 2, n=19) participated in the study. Participants were recruited through their respective schools and the Autism Resource Centre (Singapore). As shown in Table 3.2.1, the children in the two groups are comparable in terms of their mean ages, sex ratio and language background. Overall, the total sample showed an over-representation of boys.

All the participants had a pre-existing diagnosis of autism or ASD. The diagnoses were given by clinical/educational psychologists from health or education ministries, or psychologists in private practice. There may be variability in the diagnostic criteria used in the pre-existing diagnosis. To gauge this variability, the GARS checklist (described in Section 2.3) was given via post to all participants prior to the

interview. The standard scores derived from the GARS, i.e. the Autism Quotient, represent the extent to which the child shows behaviours that are indicative of autism. An Autism Quotient of  $\geq 80$  is regarded as showing significant characteristics corresponding to the DSM-IV (APA, 1995) diagnostic criteria of autism.

As shown in Table 3.2.1 (section heading 'GARS Autism Quotient'), the two groups had about equal number of children who showed at least average probability of meeting DSM-IV criteria for autism, i.e.  $AQ \geq 80$ . However, it could be noted that a sizable number (18) fell below the GARS cut-off criteria for autism characteristics. This should not be taken as indicating that these children are not autistic, but rather, reflects the increasing use among practitioners of the broader dimensional criteria for autistic spectrum disorder (ASD) as defined by Wing & Gould (1979) (Blaxill, 2002).

Table 3.2.1.: Profile of Participants

Sex and Age	Mainstream	Special Sch	Total
Number	21	19	40
Boys	19	16	35
Girls	2	3	5
Mean age	8.4 yrs	8.1 yrs	8.2 yrs
Socio-economic status*	Mainstream	Special Sch	Total
High	10	7	17
Middle	6	4	10
Low	5	8	13
Home languages	Mainstream	Special Sch	Total
English	21	19	40
Mandarin	10	9	19
Malay	1	2	3
Tamil	0	1	1
GARS Autism Quotient	Mainstream	Special Sch	Total
$AQ \geq 80$	11	11	22
$AQ \leq 79$	10	8	18

\*SES level is proxied by father's education level, as this has been reported to be the best predictor of SES level in the Singapore population (Singapore Statistic Department, 2000).

### **3.2.2. Measures**

#### *3.2.2.1 ICF checklist*

The ICF checklist used for the present study was based on the full version of the ICF. For the ICF interviews, 204 items were selected, grouped under the following 3 components:

1. Impairments in Body Function and Structures (94 items);
2. Activity limitation & Participation restriction (72 items); and
3. Environmental Facilitator/Barriers (38 items)

From the total pool of items in the ICF, the following groups of items were excluded:

- Items that were not relevant to the age group (e.g. items pertaining to marriage, child-rearing, intimate relationships, employment);
- Items that were not relevant to the cultural and geographical contexts (e.g. items about the use of domesticated animals for work, changes in seasonal weather conditions); and
- Items related to specific physical/sensory disabilities (e.g. items on the use of Braille, sign language).

As discussed earlier, an investigator-based interview protocol was developed for the present study (See Appendix C). For each item in the ICF, in addition to recording the presence of an impairment, judgment was also needed to qualify the magnitude of the functioning problem (impairment, or decrement in performance) and the extent to which an environmental factor is a facilitator or barrier. This was done using the 5 point rating scale (or ‘qualifiers’) ranging from 0 (no impairment/facilitator/barrier) to 4 (complete impairment/facilitator/barrier). For all items, in addition to the 5-point rating scale, options for the rating of ‘not specified’ (code 8) or ‘not applicable/relevant’ (code 9) were available. These were considered as indicating ‘absence of a reported problem’ (WHO, 2001, pp 222). Thus in the analysis, codes 8 and 9 were replaced by a ‘0’.

The ICF recommends that each point in the rating scale be calibrated in different domains to population standards as percentiles (See Table 3.2.2.1). However, for items measuring the extent of impairments in young children, it can be argued that it would be more appropriate to calibrate the ratings based on the discrepancy between the child's current developmental attainments, and his/her chronological age (see Table 3.2.2.1, items in B). The developmental milestones used the present study are based on available normative data from the following sources:

- Denver Developmental Screening Test – Singapore Version (Lim, Chan, & Yoong, 1994);
- Developmental Continuum for Oracy, Literacy and Numeracy Skills (Ministry of Education, Republic of Singapore, 2000a);
- Singapore Math Achievement Test (Ministry of Education, Republic of Singapore, 2000b); and
- Adaptive Behaviour Checklist – Singapore (Ministry of Education, Republic of Singapore, 2000c).

### *3.2.2.2. Diagnostic Interview for Social and Communication Disorder or DISCO*

The DISCO (Wing, 1999) was used as a concurrent measure of functioning and disability levels. As discussed in section 2.5.3, the DISCO contains a comprehensive list of over 200 items covering 4 categories of behaviour: 1) behaviour in infancy; 2) onset of recognition of problems; 3) developmental behaviours and 4) atypical behaviours. As the focus of the present study is the children's current level of functioning, only items for developmental and atypical behaviours were used. For the purpose of the present study, a three-point rating scale was used for all items, which corresponded with the coding system of the DISCO: 0 for 'no concern'; 1 for 'mild concern'; and 2 for 'marked concern'. For items measuring developmental functioning, ratings were based on the extent of developmental delay, i.e. ratings of 'mild concern' and 'marked concern' were given to children showing delays between 1:0 to 2:11 years and delays of 3 years or more, respectively.

**Table 3.2.2.1 : Calibration of Anchor Points for ICF Ratings Scales**

<b>A. For items measuring frequency of impairments in Body Function and Activity &amp; Participation</b>	
0	No problem
1	Mild problems, i.e. present less than 25% of the time, with an intensity that a person can tolerate and which happens occasionally over the last 30 days.
2	Moderate problems, i.e. present less than 50% of the time, with an intensity, which is interfering in the persons day to day life and which happens frequently over the last 30 days.
3	Severe problems, i.e. present more than 50% of the time, with an intensity, which is partially disrupting to the persons day to day life and which happens frequently over the last 30 days.
4	Complete problems, i.e. present more than 95% of the time, with an intensity, which is totally disrupting the person's day-to-day life and which happens every day over the last 30 days.
(based on guidelines suggested by CAS/WHO, 2002)	
<b>B. For items measuring extent of impairments in Body Function and Activity &amp; Participation</b>	
0	No impairment means the child achieving age-appropriate skill level or the highest level of developmental stage for the particular skill.
1	Mild impairment means a discrepancy of less than two years between the child's developmental attainment and his/her chronological age.
2	Moderate impairment means a discrepancy of two to four years between the child's developmental attainment and his/her chronological age.
3	Severe impairment means a discrepancy of more than four years between the child's developmental attainment and his/her chronological age.
4	Complete impairment means a discrepancy of more than four years and in addition, the child does not have the pre-requisite for the development of the skill.
<b>C. For items measuring impact of Environmental Facilitator</b>	
0	No facilitator.
1	Mild facilitator means the child using/relying/encountering the environmental factor less than 25% of the time.
2	Moderate facilitator means the child using/relying/encountering the environmental factor less than 50% of the time.
3	Strong facilitator means the child using/relying/encountering the environmental factor more than 50% of the time.
4	Complete facilitator means the child using/relying/encountering the environmental factor more than 95% of the time.
<b>D. For impact measuring impact of Environmental Barrier</b>	
0	No barrier.
1	Mild barrier means that the environmental factor has inhibited/disrupted child's functioning and participation less than 25% of the time.
2	Moderate barrier means that the environmental factor has inhibited/disrupted child's functioning and participation less than 50% of the time.
3	Severe barrier means that the environmental factor has inhibited/disrupted child's functioning and participation more than 50% of the time.
4	Complete barrier means that the environmental factor has inhibited/disrupted child's functioning and participation more than 95% of the time.

### 3.2.3 Procedure<sup>8</sup>

Invitations for the participants were sent through their schools and the Autism Resource Center<sup>9</sup>. From the initial 200 'open invitation letters' (See Appendix D), 60 parents responded (30%). The poor rate of response may have been affected by the SARS<sup>10</sup> outbreak in Singapore during the data collection period, as travel within the city was significantly restricted and schools were closed for an extended period of time. Two groups of respondents were not included in the present study, i.e. children less than 5 years old and those whose educational placement was unconfirmed (e.g. children who were considering transfer to/from special schools). Upon receiving informed written parental consent (See Appendix D), interviews were carried out by the researcher with the parents of all the children. Nine interviews were recorded and these were used in the inter-rater reliability study, where a second rater independently completed the ICF checklists based on video/audio recordings of the interviews. Both the researcher (first rater) and the second rater are qualified educational psychologists who are experienced in conducting interviews with parents of children with disabilities and familiar with the language used in the interview, i.e. Singapore English. Prior to the second rating, a two-day training session, which included joint-rating sessions by the two raters were carried out (See Appendix E for details of the inter-rater training session).

A sub-sample of 10 children from each group was selected for a second interview using the DISCO. Interviews using the DISCO and the ICF were carried out by the same researcher who is a trained certified user of the DISCO. The order of the ICF and DISCO interviews were counterbalanced, with an interval of between 3 to 5 weeks.

---

<sup>8</sup> Procedure for the study was approved by UCL ethics committee for non-NHS human research (Project ID number 0038/001)

<sup>9</sup> Autism Resource Centre is a non-profit organization that provides support and therapy services for children with autism in Singapore.

<sup>10</sup> Severe Acute Respiratory Syndrome

### **3.3 RESULTS**

The results will be discussed in relation to the three criteria that the ICF has to meet in order to be considered as a reliable and valid measure of special educational needs of children with autism in Singapore.

#### **3.3.1. Results for the first criterion**

The first criteria was: ICF ratings of individual children must show consistency between items within the same component, as well as consistency in the evaluations of the same children by different raters.

For the purpose of analysing internal consistency reliability, children's ratings for the individual items were totalled to obtain the following components scores:

1. Body Function & Structures;
2. Activity & Participation;
3. Environmental Barriers; and
4. Environmental Facilitators.

In addition, two composite scores were obtained:

5. Functioning & Disability (composite of components scores 1 and 2 above);
6. ICF SEN score (composite of components 1, 2, and 3 above).

The composition of items for the Functioning and Disability composite scores is based on the ICF framework (see Fig. 3.1), while the ICF SEN score is based on the definition of children with SEN used in the present thesis:

*Children with special educational needs are defined as those who experience activity limitations and participation restrictions, as a result of impairments in body functions/structure and environmental barriers (c.f. Section 2.3.8).*

##### ***3.3.1.1 Internal consistency reliability***

Reliability analyses were carried out for each of the component and composite scores, using Cronbach Alpha. It should be noted that the alpha values were calculated based on items with variance greater than zero, and consequently 70

(26.92%) of the interview items were excluded from the reliability analysis. The reasons for zero variance for these items are summarized in Table 3.3.1.1.1.

Table 3.3.1.1.1: Items Excluded from Reliability Analysis

<i>ICF Component</i>	<i>Items Excluded</i>	<i>Reasons for zero variance</i>
Body Function & Structure	<ol style="list-style-type: none"> <li>1. All items under 'Body Structure'.</li> <li>2. All items under 'Functions of cardiovascular, haematological, immunological and respiratory systems'.</li> <li>3. All items under 'Digestive, metabolic and endocrine systems'.</li> <li>4. All items under 'Neuromusculoskeletal and movement-related functions'.</li> <li>5. All items under 'Functions of the skin and related structures'.</li> </ol>	All children in the sample were reported to have no impairment in any of these areas, i.e. rating = 0.
Environmental Factors	<ol style="list-style-type: none"> <li>1. Items for assistive products and technology used for daily living and mobility.</li> <li>2. All items under 'Natural environment and human-made changes to the environment', e.g. climate, physical geography, air quality.</li> <li>3. Items on amount and quality of physical and emotional support provided by strangers.</li> </ol>	<p>All children in the sample were reported not to need such products &amp; technology, i.e. rating = 9 (irrelevant).</p> <p>For all children in the sample, these were reported to have neither a facilitative nor a barrier effect, i.e. rating = 0.</p> <p>None of the children in the sample relied on direct support by strangers, i.e. rating = 9 (irrelevant).</p>

As shown in Table 3.3.1.1.2 the reliability coefficients (alpha values) for Body Function and Structures, Activity and Participation, and Functioning and Disability (composite) exceeded the 0.8 level for 'high reliability', while for the components with lesser number of items, i.e. Environmental Facilitators and Barriers, alpha values exceeded the 0.6 criteria for 'adequate reliability' (Field, 2000).

**Table 3.3.1.1.2: Internal Reliability (Cronbach Alpha) of ICF Component & Composite Scores**

Component score	No. of items included in reliability analyses <sup>+</sup> (%)	Alpha
1. Body Function & Structures	67 (76.12%)	0.88**
2. Activity & Participation	74 (100.00%)	0.98**
3. Environmental Barriers	23 (54.76%)	0.61*
4. Environmental Facilitators	26 (61.90%)	0.74*
<b>Composite Scores</b>		
<b>Functioning &amp; Disability (Composite from 1 &amp; 2)</b>	141 (80.11%)	0.97**
<b>ICF SEN score (Composite from 1,2 and 3)</b>	164	0.96**

<sup>+</sup>Only items with variance greater than zero were included in the analyses.

\*Exceeds the 0.6 levels for 'adequate reliability'.

\*\*Exceeds the 0.8 levels for 'high reliability'.

### 3.3.1.2. *Inter-rater agreement*

Concurrence in total ratings was calculated based on the correlation of the total ICF component scores given by the first and second raters. Due to the small sample (N =9) non-parametric analysis was used (Spearman's rho). Results indicated that the correlations between the total ratings were high for all components of the ICF (see Table 3.3.1.2.1):

**Table 3.3.1.2.1 : Correlation (Spearman r) Between Ratings of First and Second Raters**

ICF Components		First Rater
Second rater		
	Body Function & Structures	0.82**
	Activity & Participation	0.78*
	Environmental Facilitators	0.93**
	Environmental Barriers	0.80**

\* $p \leq 0.05$ . \*\*

$p \leq 0.01$

Kappa coefficients were calculated for all items with variance greater than zero.

Following Landis and Koch (1977) the kappa values were interpreted as follows:  $\geq 0.75$  indicating 'excellent reliability';  $\geq 0.65$  indicating 'good reliability';  $\geq 0.40$  indicating 'adequate reliability' and  $<0.40$  indicating agreement that are at chance levels or less. The number and percentage of items in each of ICF components

achieving the various levels of reliability is shown in Table 3.3.1.2.2. For all components of the ICF, a majority of items (44.34%) reached ‘excellent agreement’ levels. Overall, 95.63% of the items reached adequate to high agreement levels (See Appendix F for Kappa values for individual ICF items).

**Table 3.3.1.2.2 : Number and Percentages of ICF Items with Adequate to excellent Agreement**

ICF Component	$k \geq 0.75$ Excellent agreement	$k \geq 0.60$ Good agreement	$k \geq 0.40$ Adequate agreement	$k < 0.40$ Chance levels	Total
<b>Body Function &amp; Structures</b>	28 52%	13 24.07%	12 22.22%	1 1.85%	54
<b>Activity&amp; Participation</b>	25 43.86%	15 26.32%	14 24.56%	3 5.26%	57
<b>Environmental Facilitators</b>	2 14.29%	6 42.86%	5 35.71%	1 7.14%	14
<b>Environmental Barriers</b>	3 25.00%	3 25.00%	5 41.67%	1 8.33%	12
<b>Total</b>	58 42.34%	37 27.01%	36 26.28%	6 4.38%	137

### 3.3.2 Results for the second criterion

The second criterion was: ICF measures of children’s functioning must be able to meaningfully distinguish children with varying levels of special educational needs. To evaluate the extent to which the ICF was able to distinguish between children in mainstream and special schools, ICF scores of children in mainstream (Grp 1) and special schools (Grp 2) were compared (Table 3.3.2). Preliminary analyses using the Kolmogorov-Smirnov goodness-of-fit test to a normal distribution was carried out. Results indicated that all ICF variables were normally distributed, except for Grp 1 scores on Environmental Barriers (Kolmogorov-Smirnov  $Z = 0.184$ ,  $p \leq 0.01$ ). Thus in the comparison of means for ICF scores, both parametric (independent-sample t-test) and non-parametric analyses (Mann-Whitney U test) were carried out (See Table 3.3.2).

**Table 3.3.2 : ICF Scores – Group Means**

ICF Components Scores	Mainstream			Special School			Mean	Rank	t value	z value
	Mean	SD	Std Error	Mean	SD	Std Error				
Body Function & Structures	32.95	10.58	2.31	68.16	17.79	4.08	11.64	30.29	-7.69**	-5.04**
Activity & Participation	58.14	18.87	4.12	152.05	41.29	9.47	11.05	30.95	-9.40**	-5.38**
Environmental Barriers	6.62	5.34	1.17	10.00	4.31	0.99	16.60	24.82	-2.15*	-2.23*
Environmental Facilitators	25.81	8.29	1.81	39.31	6.33	1.45	12.95	26.84	-5.74**	-4.30**
Functioning & Disability (Composite)	91.09	24.81	5.41	220.21	53.72	12.32	11.10	30.89	-9.92**	-5.35**
ICF SEN score (Composite)	97.33	27.10	5.91	228.37	52.42	12.03	11.05	30.95	-9.78**	-5.38**

\* $p \leq 0.05$ , \*\* $p \leq 0.01$

As predicted, the Functioning & Disability scores of children in special schools were significantly higher than the Functioning & Disability scores of children in mainstream schools,  $t(38) = -9.916, p \leq .01$ ; indicating that the children in special schools have more severe impairments in functioning and greater limitations in activity and participation. Similarly, the scores on Environmental Facilitator for Group 2 were significantly higher than Group 1, indicating that children in special schools received greater levels of support,  $t(38) = -3.82, p \leq .001$ . The composite ICF SEN score, which is merely reflecting the scores obtained from Functioning and Disability and Environmental Barriers, also indicates similar results, i.e. children in special schools (Group1) show higher scores than those in mainstream school (Group 2),  $t(38) = -9.78, p \leq .01$ .

### **3.3.3 Results for the third criterion**

The third criterion was: There must be concurrence between evaluations of children functioning based on the ICF, with other independent measures of functioning which have been developed specifically for children with autism.

To evaluate the degree of concurrence between levels of functioning and disability measured by the ICF and the DISCO, children's ratings for individual DISCO items were collated to obtain a Total DISCO score, which reflects the extent to which the children show delays in developmental functioning and severity of autism characteristics.

Correlation analyses (Pearson's r) were carried out for the total sample as well as for the 2 groups, i.e. mainstream and special school. Comparisons between the Total DISCO and ICF Functioning and Disability scores for all 19 children indicated a strong correlation;  $r = 0.87, n = 19, p \leq 0.01$ .

However, as shown in Table 3.3.3.1, when correlations within the mainstream and special school groups were examined separately, the correlation was stronger for children in special schools, while for children in mainstream schools, the correlation was not significant.

**Table 3.3.3.1: Correlation Matrix (Pearson's  $r$ ) Between ICF Functioning & Disability and Total DISCO Scores by Groups**

		ICF Functioning & Disability	
		Total N=19	Mainstream Grp1; n=9
Total DISCO	Total N=19	**0.87	0.396
	Special Sch Grp2; n=10		**0.76

\*\* $p \leq 0.01$

The difference in the correlations between the two groups prompted a closer analysis of their ICF scores. It was found that the range of scores for mainstream children for Functioning and Disability is limited, compared with the range of scores for children in special schools (See Fig. 3.3.3). To investigate this further, the variance of ICF scores of the two groups were compared using the Levene's test for equality of variance. It was found that the variance of the two groups were significantly different for the composite Functioning and Disability scores, as well as its sub-components, i.e. Body Functioning, and Activity and Participation (See Table 3.3.3.2). This suggests that the lack of significant correlation between ICF and the DISCO scores for children in mainstream schools may be due to the very limited range of scores for ICF scores obtained by children in this group.

**Fig. 3.3.3. Range of Scores for ICF Functioning and Disability by Groups**

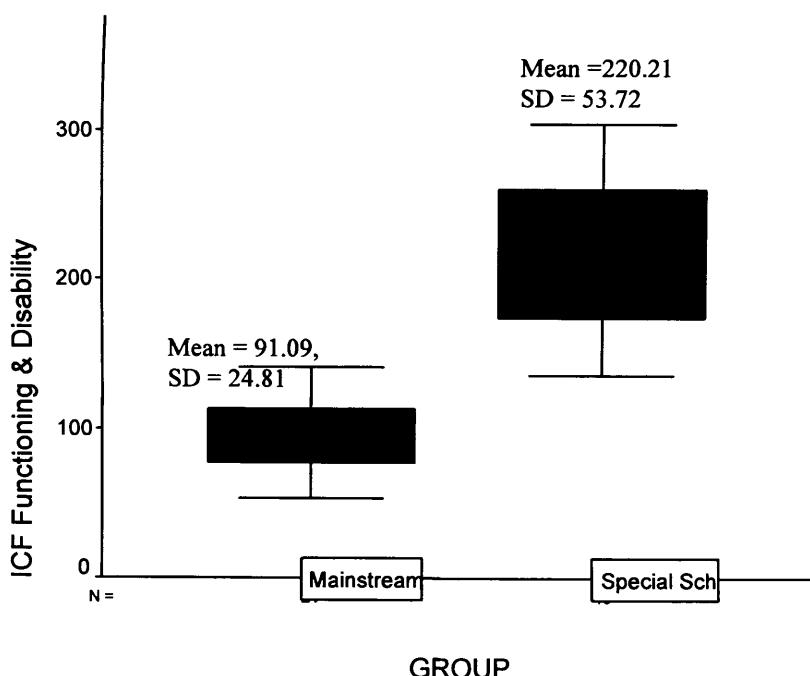


Table 3.3.3.2 : Levene's Test for Equality of Variance

ICF Component Scores	Mean Score		F value
	Mainstream	Special Sch	
1. Body Function & Structures	32.95	68.16	3.85*
2. Activity & Participation	58.14	152.05	11.51**
3. Environmental Barriers	6.62	10.00	0.76
4. Environmental Facilitators	25.81	39.31	0.97
Functioning & Disability (Composite of 1 & 2)	91.09	220.21	8.88**
ICF SEN score (Composite of 1, 2 & 3)	97.33	228.37	9.61*

\*\* $p \leq 0.01$ ; \* $p \leq 0.05$

Nb.  $p$  values indicates the probability that the variance of scores in one group is significantly smaller/greater.

### 3.3.4 Summary of results

Results on internal reliability indicated that the internal consistency between children's individual ratings and their scores on the ICF components ranged between adequate to high levels, with items for Environmental Facilitators and Barriers having slighter lower alpha values than Disability & Functioning and ICF SEN scores. This was somewhat expected as the Environmental components have fewer items. Inter-rater reliability was satisfactory, with 95% of the items indicating inter-rater agreement between adequate to excellent levels.

Results on validity indicated that the scores of children in special schools were significantly higher than those in mainstream for all the ICF components. There was also a strong correlation between children's Functioning & Disability scores with total DISCO scores, although the correlation was significant for children in special schools only. Further analysis of the two groups indicated that the range of values for Functioning and Disability for the mainstream group was significantly restricted.

### **3.4. Discussions**

The findings will be discussed in terms of the objectives of Study 1, and the limitations of the study which will impact the generalisability of the present findings. In the last section, the implications of these findings for the next phase of the thesis will be discussed.

#### **3.4.1 Evaluation of the first criterion**

*The first criterion was: ICF ratings of individual children must show consistency between items within the same as component, as well as consistency in the evaluations of the same children by different raters.*

Internal reliability index (alpha) for all ICF components were high. This suggests that there is a coherent link between individual item ratings for each child, and his/her total ratings. The present results are consistent with results of the ICF field-trials involving patients with muscular-skeletal disorders (WHO, 2000), where inter-rater reliability was reported to be high, and agreement levels for the items on Environmental factors appear to be slightly lower than the other ICF components (See Section 2.3.7). The greater diversity in the items in the Environmental component may also account for the reduced inter-rater agreement observed.

Based on the independent ratings of the two raters, there was a significant correlation in the children's total ICF ratings. At the same time, there were adequate to high inter-rater agreement in the ratings of the two raters for more than 95% of the items in the ICF. This suggests that evaluations of children based on the ICF were consistent across raters. However, it should be noted that the raters underwent an intensive two-day training session, which involved joint-rating sessions. In addition, the second rater was given prior information on the types of probes used in the interviews. Thus, the reliability of inter-rater agreement achieved reflects not only the properties of the ICF checklist; it is also reflective of the effects of the intensive training.

### **3.4.2 Evaluation of the second criterion**

*The second criterion was: ICF measures of children's functioning (i.e. impairments in body functions/structures, activity limitation and participation restriction) must be able to meaningfully distinguish children with varying levels of special educational needs.*

As discussed, in the Singapore context, children's placement in mainstream or special school is a pre-existing independent index of children's level of special educational needs. Children are referred to special schools following professionals' assessments and judgments that the child's impairment is severe, such that his/her needs cannot be met in mainstream schools, i.e. the child requires higher levels of specialised support which, in the Singapore context, are only available in special school settings. In contrast, children remain in mainstream if their functioning difficulties were assessed to be mild (i.e. not severe), and they do not require long-term specialised support. In ICF terms, children in special schools in Singapore would be expected to show high levels of impairments in Body Function, Activity and Participation, and high levels of Environmental Facilitator. Results of the present study supported this prediction. At the same time, children in special schools also had greater scores for Environmental Barriers, which suggests that although they were receiving additional specialised support and provisions, these were inadequate to completely overcome the difficulties arising from the impairments in Body Functions, and limitations in Activity and Participation.

Previous studies have shown that the ICF checklist was able to distinguish the profiles of patients with musculoskeletal, cardiovascular and chronic neurological conditions (CAS, WHO, 2002). The present findings provided some evidence for the criterion validity of the ICF with patients with a developmental disorder, namely autism.

### **3.4.3 Evaluation of the third criterion**

*The third criterion was: There must be concurrence between evaluations of children's functioning based on the ICF, with other independent measures of functioning which have been developed specifically for the autistic population.*

The DISCO is an autism specific instrument that includes measures of children's developmental functioning and atypical characteristics. Findings from the present study indicated that children's scores on ICF Functioning and Disability correlated significantly with their scores on developmental functioning and atypical characteristics in the DISCO. This suggests that there is good concurrence between the ICF, which has been developed as a generic tool for all disability conditions, with measures of functioning derived from an autism specific instrument, i.e. the DISCO.

It can be concluded that the ICF met all the three criteria and could be considered as an adequate measure of special educational needs of children with autism. However, a number of limitations and methodological issues in the present study can be identified.

### **3.4.4 Methodological issues**

#### *3.4.4.1 Interviewer's prior knowledge*

In the present study, the researcher conducted all the interviews (for both the ICF and DISCO). While this reduced the variability that may be due to interviewer's style and skills, a possible limitation is that the interviewer was not blind to the background of the children. Knowledge on whether they were from mainstream or special school, might have affected the quality of probing during the interviews. However, it can be argued that given that the interviews were based on a structured protocol (See Appendix C), where probes and guidelines for interviewer judgments were determined *a priori*, the quality of questioning and probing across subjects were adequately controlled.

#### *3.4.4.2. Inter-interviewer reliability*

While inter-rater agreement was evaluated in the present study, *inter-interviewer* reliability, i.e. whether the quality of interviewing had an effect on reliability, was not investigated. Admittedly, it may be difficult to completely control for this factor, as interviewing the same person twice would result in the respondents' tendency to provide similar responses. However, if more than one interviewer is used in the inter-rater reliability study, differences in quality of probing / questioning by different interviewers could be controlled by counterbalancing the role of interviewers and raters, i.e. one

interviewer acting as the second rater for interviews conducted by the other interviewer, and vice-versa. For the present study, this approach was not feasible.

#### *3.4.4.3 Sample selection*

The generalisability of the present findings could also be limited to the particular sample selected for the study. For example, the sample is limited in age range (5 to 11:11 years) and includes an over-representation of boys from English speaking homes, with middle to high SES. As highlighted by Simmeonsson et al. (2003), there might be aspects of behaviour of very young children which have been excluded in the ICF. In addition, as all the participants used English as one of the dominant home-languages, all interviews were conducted in Singapore English.

Participants in the study volunteered for the interview. It may be that parents who were interested to articulate their views on autism or special educational needs may have been more motivated to participate in the study, while parents who were not conversant in English, or not confident to articulate their views, did not respond to the invitation letters. This could partly account for the relatively high proportion of parents from high SES in the sample. Thus, the reliability and validity of the ICF that is established from the present study cannot be generalized to other groups, in particular younger aged children (under 5 years old) and those from homes where Singapore English is not used.

#### *3.4.4.4 Group differences*

A further analysis indicated that the correlation between the DISCO and ICF scores for the mainstream group was not significant, due to the small range (limited variance) in the children's ICF Functioning and Disability scores. This finding can be interpreted in several ways: firstly it may be that the instrument (i.e. ICF checklist) was not adequately sensitive to reflect marginal differences in Functioning and Disability that may occur among mainstream children, who were mostly children with high-functioning autism, with little difficulty coping with the demands of mainstream schools. On the other hand,

it may be that ASD children in mainstream schools in Singapore have a very uniform functioning profile.

The data from the present study was not able to identify which is a more plausible explanation. To fully explore this issue, a larger sample would be needed, in particular children in mainstream schools whose Functioning and Disability levels are within the moderate to severe ranges, i.e. those whose ICF scores overlaps with children in special schools. This might be achieved by including mainstream children who are being considered for transfer to special schools. However, the inclusion of this group will reduce the clarity in the distinction between the two groups of children selected for the present study, i.e. children with autism with low SEN levels (in the mainstream schools) and those with higher SEN levels in the special schools.

### **3.4.5 Conclusion**

For the purposes of the present thesis, results from Study 1 indicated that the ICF has adequate validity and reliability as a measure of special educational needs of children with autism, and thus could be considered for use in the phase two of the study, which aims to identify indicators which best reflect children's special educational needs. The next chapters focus on the selection of the measures for the indicators of SEN that can be used alongside the ICF in Study 2 of present thesis.

## CHAPTER 4

### SELECTION OF MEASURES FOR THE INDICATORS OF SPECIAL EDUCATIONAL NEEDS

As discussed earlier, in Chapter 2, the indicators of SEN that have been identified as relevant for children with autism are: theory of mind, executive function, central coherence, and cognitive modifiability. One of the main aims of the present thesis is to evaluate the extent to which any of the indicators (or their combination) could provide a more valid assessment of SEN for children with autism than measures of intelligence (which is currently the main indicator of SEN used in Singapore). In this chapter, the operational definitions, measurement issues and available instruments for each of the identified indicators are discussed to identify the most appropriate measures for use in the present thesis. The following sections in this chapter focus on each of the indicators of SEN:

- 4.1 Measure of intelligence;
- 4.2 Measure of executive function;
- 4.3 Measure of central coherence;
- 4.4 Measure of cognitive modifiability; and
- 4.5 Measure of theory of mind.

The selection of measures for use in the present thesis is based on the following criteria:

- Analysis of its psychometric properties indicates that the measure has adequate reliability; and validity (in relation to the intended constructs);
- The task demands are appropriate for the intended participants in Study 2, i.e. children with autism, aged 8 to 12 years.

#### 4.1 MEASURE OF INTELLIGENCE

This discussion focuses specifically on the use of measures of intelligence for special education. In the literature, other notions of intelligence have emerged, such as the concepts of multiple intelligence (Gardner, 1993) and emotional intelligence (Goleman, 1995). These are excluded from the present review, as they are not directly related to issues concerning the assessment of children with special educational needs.

The use of children's scores in intelligence tests as reflecting their need for special education dates back to 1897, when Alfred Binet began work on tests of individual differences with a group of "subnormal" children in Paris schools. This later resulted in the Simon-Binet test of intelligence, which first appeared in 1905 (Wasserman & Tulsky, 2005). Since then, there have been several other attempts at developing measures which are intended to identify children who appear to have difficulties learning, such as the Wechsler Scales (Wechsler, 1991, 1993 & 2004), the British Ability Scales (Elliott, Murray & Pearson, 1983) and the Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983 & 2004).

#### **4.1.1 What is intelligence; how is it measured?**

There has been a long-standing debate about what intelligence tests measure. Commentators noted that a major problem lies in the fact that the debate is circular; the notion of IQ carries different meaning in different contexts (Sternberg & Detterman, 1986). Consequently, it appears that the concept of intelligence cannot be reliably and consistently defined, independent of the particular tool/instruments that claim to measure 'intelligence' (Howe, 1997; Richardson, 1996; Sternberg, 1984 & 2000). It is unsurprising that the most frequently quoted 'definition' is 'intelligence is what intelligence tests test' (Boring, 1923). In selecting a measure of intelligence for the present thesis, particular attention has to be paid to its context of use (i.e. as an indicator of SEN), and its intended population, namely children with autism in Singapore.

As discussed in the introductory chapter of this thesis, on the one hand studies have shown that when appropriately used, i.e. when used with reference to the appropriate populations and for the purposes that they were intended, intelligence tests have adequate reliability and predictive validity. On the other hand, the lack of treatment validity and the existence of bias have frequently been raised as the primary limitations for the use of IQ tests in SEN assessments. The issues of criterion and treatment validity of IQ tests for SEN assessments of children with autism are addressed in Studies 2 and 3 of the present thesis. In the next sections, two measurement issues which affect the selection of IQ tests

for use in the present thesis are discussed, namely issues relating to the construct (or model) of intelligence and test bias.

#### **4.1.2 Measurement issues**

##### *4.1.2.1 Construct or model of intelligence*

Commentators tend to identify two major theoretical models of intelligence testing, each focusing on slightly different aspects of the construct, namely the general ability (or *g* factor) and multi-factor models of intelligence.

The concept of the general factor in intelligence was first asserted by Spearman (1904), who proposed that ‘all branches of intellectual activity have in common one fundamental function’ (p.284), which is described as the ‘general mental energy’. This ‘*g*’ factor is a mathematically derived general factor, which stems from the variance shared by a series of cognitive/intelligence tasks. Summarising the literature, Jensen (1998) argued that the *g* factor correlated with scholastic performance, reaction time, training and job success, occupational status and earned income. The validity of the *g* factor has been challenged by others, and these challenges are discussed in the later sections.

The multi-factor model of intelligence was put forward by Thurstone (1938), who developed the technique of multiple-factor analysis. This enabled the analysis of correlation matrices and the extraction of separate ability factors which seemed unrelated to each other. Challenging the notion of a *g* factor, he obtained scores from a sample of university students on 56 tests, and extracted seven primary factors. Thurstone initially recommended that an individual should be described in terms of a profile of mental abilities, instead of a single index of intelligence; but later, following the development of higher order factor analysis techniques, he conceded the possible existence of an overarching factor at a higher-order level, which is akin to Spearman’s *g*.

Many of the widely used intelligence tests for children, such as the Wechsler Intelligence Scales for Children (Wechsler, 1991, 1993 & 2004) and the British Ability Scales (Elliott et al., 1983), have incorporated the concept of a hierarchical model of intelligence, where

at one level intelligence is described in terms of specific cognitive processes/factors; and at another level a general or global index of ability is obtained (see Table 4.1.1.2).

Table 4.1.1.2 Measures of Intelligence for Children

Name of Tests <sup>11</sup>	Construct	Structure	Ability indices provided
Wechsler Intelligence Scales for Children (3 <sup>rd</sup> editions) or WISC (Wechsler, 1991)	General intelligence	1. General Intelligence 'g' factor, comprising of: • Verbal; and • Non-verbal abilities.	1. Full Scale IQ (Global index) 2. Verbal IQ 3. Performance IQ
British Ability Scales or BAS (Elliott et al., 1983)	Interrelated cognitive processes and abilities	1. Domain specific abilities 2. General ability	1. Diagnostic and Achievement scales 2. Verbal ability; Non-verbal reasoning and Spatial abilities scales 3. General Conceptual Ability index (Global index).
Kaufman Assessment Battery for Children, or KABC (Kaufman & Kaufman, 1983)	Non-verbal cognitive abilities	1. Learning Ability 2. Sequential processing 3. Simultaneous processing 4. Planning ability	1. 'Learning scale' 2. 'Memory' scale 3. 'Simultaneous' scale 4. 'Planning' scale 5. 'Knowledge' scale 6. Mental Processing Index (Global index)

Compared to the BAS and the KABC, the construct and structure of the WISC-III is closest to the notion of the 'g' factor. This is reflected in Wechsler's (1939) definition of intelligence:

*Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment. It is global because it characterizes the individual's behaviour as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable (p.3).*

The BAS and the KABC were initially developed to provide a profile of specific cognitive abilities rather than a summative (or global) IQ index (Elliott et al., 1983; Kaufman & Kaufman, 1983). However, in both tests, global indices of ability which are based on composite subtest scores were included (See Table 4.1.1.2). It can be argued that in all the 3 tests reviewed, the notion of the 'g' factor is implied.

<sup>11</sup> Although there are recent versions of these tests, namely WISC-IV (Wechsler, 2004); BAS-II (Elliott, 1996) and KABC-II (Kaufman & Kaufman, 2004), the review focuses on the previous versions of the tests, as most of the research evidence for validity are based on these earlier versions. In addition, although some of the items in the later versions are new, the underlying structure and constructs of these tests remains unchanged (Kaufman et al., 2005; Zhu & Weiss, 2005; Hill, 2004).

The evidence for  $g$  is derived statistically, i.e. from the use of factor analyses. As discussed in Section 2.2, one problem with this approach in establishing construct validity is that the statistical or factorial validity still needs evidence of a relationship to life events outside the test itself if the factors are to have explanatory utility (Thorndike, 1997). As yet, there are no neurological correlates that support the existence of a  $g$  factor in intelligence. This lack of theoretical basis has been one of the prevailing criticisms of standardised measures of intelligence.

For all these tests, strong psychometric properties are reported in the manuals. Internal reliabilities are high and confirmatory factor analyses provide support for the assumed internal structure of the tests. Criterion validity has been established by examining the correlations between global indices of the tests with measures of scholastic abilities (e.g. reading tests) and with other IQ tests, e.g. the criterion validity of the KABC and the BAS were established based on the correlations of these tests with the WISC-III Full Scale Index (Wechsler, 1991; Elliot, 1983; Kaufman & Kaufman, 1983).

#### *4.1.2.2. Test bias*

One of the strongest criticisms against the use of standardised intelligence tests is that they are biased against culturally and linguistically diverse populations. Critics have asserted that because many of the items in these tests tap knowledge and information that emanates from western cultures, they are biased against children who are not adequately exposed to such cultures. The evidence cited to support this assertion is the fact that individuals from ethnic minority cultures typically register a mean IQ that is significantly lower than the norms (e.g. Hilliard, 1979).

However, some commentators (e.g. Ortiz & Dynda, 2005) argue that a distinction must be made between culture loading and culture bias (extent to which the evidence of validity is significantly different for different groups). All tests have cultural content (loading), so individuals from different cultures may not perform on a given test in a manner that would be expected, given their age; as their culture may not share the same knowledge or skills being sampled by the tests. However, this does not imply that the

reliability and validity of the test for children from different cultures are automatically different (test bias). Studies that have investigated the extent to which validity of the Wechsler Scales differ for children from different cultural groups report that there is little evidence of test bias in terms of test reliability (measurement of error or accuracy); test structure (item difficulty); factor structure (theoretical structure, cluster or composite scores) and prediction (academic success or achievement) (Cummins, 1984; Jensen, 1980; Valdes & Figueiroa, 1994).

One approach to overcoming the issue of test bias that has been recommended is to focus on the measurement of non-verbal abilities (e.g. McCallum & Bracken, 2005; Naglieri, 1997). However, there are several problems with this approach. Firstly, by excluding test items which rely on verbal skills (e.g. receptive and expressive vocabulary), the content validity of these tests as measures of intelligence may be compromised. Secondly, The assumption that non-verbal tasks are 'culture-free' (Naglieri, 1997) is highly questionable. For example, Table 4.1.2.2 shows the scores of a Singaporean child (aged 11:8 years), whose home languages are English and Mandarin.

**Table 4.1.2.2: Comparison of the WISC-III Scores of a Singaporean Child based on Singapore and UK Norms.**

WISC Subtest	Raw Score	Scaled Score (UK Norms)	Scaled Score (Singapore Norms)
<i>Verbal Scale</i>			
Information	25	18	15
Similarities	26	15	14
Arithmetic	30*	19	16
Vocabulary	38	12	14
Total Scaled Score (pro-rata)	-	80	71
<b>Verbal IQ Index</b>		<b>137</b>	<b>125</b>
<i>Performance Scale</i>			
Picture completion	27	16	16
Coding	79	19	14
Block Design	69*	19	16
Symbol Search	39	18	14
Total Scaled Score (pro-rata)	-	90	72
<b>Performance IQ Index</b>		<b>152</b>	<b>131</b>
<b>Full Scale IQ Index</b>		<b>148</b>	<b>131</b>

(\*Ceiling or maximum raw score)

As illustrated, when different cultural norms were used (i.e. UK versus Singapore) on the same set of raw scores, there is a marked difference in the child's corresponding standard scores. For both the Verbal and Performance IQ indices, the difference is more than one standard deviation (SD=15). This suggests that both verbal and non-verbal tasks are susceptible to variations that occur across different cultures and populations.

A key assumption in the use of standardised intelligence tests for children from diverse populations is that of comparability. Ortiz & Dynda (2005) used the term 'acculturation' to describe the extent to which individuals' general acquisition and learning of the cultural elements of the society in which he or she is being raised. The assumption of comparability in standardised intelligence tests is that the child's level of acculturation is similar or comparable to that of the children who made up the normative sample of the tests. In the present thesis, to ensure that this assumption is met, the use of local Singapore norms is important.

For the present thesis, it can be argued that the WISC-III (Singapore) (Wechsler, 1996) is the most appropriate measure of intelligence: it is the only measure of global intelligence that has been adapted and normed for Singapore children<sup>12</sup>.

#### **4.1.3 Description and review of the WISC-III (Singapore)**

##### *4.1.3.1 Analysis of content*

The contents of the WISC-III Singapore (Wechsler, 1996) were kept as similar as possible to the original test, i.e. WISC-III (Wechsler, 1991). It comprises 13 subtests (see Table 4.1.3) which are categorised into two components – Performance and Verbal Subtests.

---

<sup>12</sup> The WISC-III, which is the only standardized intelligence test that has been adapted for the local population, is also the most widely used IQ test among psychologists in the Ministry of Education.

**Table 4.1.3 : Subtests in the WISC-III (Singapore)**

<i>Subtest Name</i>	<i>Description</i>
<b>Verbal Subtests</b>	
1. Information	A series of orally presented questions that tap the child's knowledge about common events, objects, places and people.
2. Similarities	A series of orally presented pairs of words for which the child explains the similarity of the everyday objects or concepts they represent.
3. Arithmetic	A series of arithmetic problems which the child solves mentally and responds to orally.
4. Vocabulary	A series of words presented orally which the child defines.
5. Comprehension	A series of orally presented questions that require the child to solve everyday problems or to show understanding of social rules and concepts.
6. Digit Span	A series of orally presented number of sequences which the child repeats verbatim, and in reverse order.
<b>Performance Subtests</b>	
7. Coding	A series of simple shapes or numbers each paired with a simple symbol. The child draws the symbol in its corresponding shape/number.
8. Picture Completion	A set of colourful pictures of common objects and scenes each of which is missing an important part which the child identifies.
9. Picture Arrangement	A set of colourful pictures, presented in a mixed up order, which the child rearranges into a logical story sequence.
10. Block Design	A set of modelled or printed two-dimensional geometric patterns which the child replicates using two-colour cubes.
11. Object Assembly	A set of jig-saw puzzles of common objects, each presented in a standardised configuration, which the child assembles to form a meaningful whole.
12. Symbol Search	A series of paired group of symbols, each pair consisting of a target group and a search group. The child scans the two groups and indicates whether or not a target symbol appears in the search group.
13. Mazes	A set of increasingly difficult mazes, printed in a response booklet, which the child solves with a pencil.

Individual children's scores in the test are described in terms of the following:

- Scaled score Standardised score for individual subtests.
- Performance IQ index Standardised composite score based on scaled scores for subtests in the Performance Subtests\*.
- Verbal IQ index Standardised composite score based on scaled scores for subtests in the Verbal Subtests\*.
- Full-scale IQ index Standardised composite score based on the aggregate of Scaled Scores in both the Performance and Verbal subtests
- Factor based index Standardised composite score based on scales derived from factor analyses of the WISC-III subtests, namely 1) verbal comprehension; 2) perceptual organisation; 3) freedom from distractibility; and 4) processing speed.

*\*Note: A minimum of 4 subtest scores are required for the computation of the Performance or Verbal IQ indices.*

Modifications in the WISC-III (Singapore) were as follows:

- Changes in items that were culture specific, e.g. in the Information subtest, the original question “What are the four seasons of the year” was changed to “What are the four points of the compass”. The rationale was that four seasons of the year are not experienced in Singapore.
- Changes in language structure, e.g. in the Comprehension subtest, the question “What is the thing to do when you cut your finger” was changed to “What should you do when you cut your finger?” which is more consistent with the language structure of Singapore English (Foley, 1998).
- Changes in the ordering of items according to difficulty level, e.g. in the Arithmetic subtest, the question “If I cut an apple in half, how many pieces will I have?” was changed from the 6<sup>th</sup> position in the original test, to the 9<sup>th</sup> position in the WISC-III (Singapore).

In the test manuals for the WISC-III, it was reported that the items were developed and reviewed by expert panels to ensure that they reflect their intended constructs (Wechsler, 1991). However, details regarding the selection and composition of the panels are lacking.

#### *4.1.3.2 Analysis of reliability and internal structure*

The reliability and validity indices (confirmatory factor analysis) for the WISC-III (Singapore) were comparable to the original test. Confirmatory and exploratory factor analyses reported in the test manuals provided strong statistical support for the overarching *g* factor and the four-factor model claimed by the test developers. Stability and reliability indices also provide strong evidence of internal consistency.

At the same time, research on other comparable versions of the Wechsler Scales, such as the Wechsler Adult Intelligence Scales (WAIS), is cited (Wechsler, 1991) as evidence that the subtests and factor-based indices involve specific processes; however, most of the cited evidence is inferential. It relies on subjective identification of common item content and inferences about the psychological processes that would explain the

correlation between the items. Factor loadings are explained by invoking common psychological processes, without direct evidence that examinees actually used those processes during testing. For example, the factor index 'Freedom from Distractibility' is based on the loadings of scores on the Arithmetic and Digit Span subtests. However, there were no attempts to examine the extent to which children's responses in these subtests were affected by their ability to inhibit the effects of competing verbal/auditory stimuli.

#### *4.1.3.3 Analyses of validity*

Correlations among a variety of measures are provided in the test manuals; these generally support the notion that the test measures general intelligence, predicts academic achievement, is related to clinical diagnostic categories, and is consistent with previous versions of the test (Wechsler, 1991).

In the administration manual, the test developers claim that the WISC is useful for the purposes of 'educational planning', 'resource provision' and 'placement decisions' (Wechsler, 1991, pp7). However, the technical manual (p.101) addresses this issue in a single paragraph, arguing that it is the sole responsibility of the test user to supply evidence regarding test consequences and its utility. Although extensive data from clinical groups are provided (implying support for claims of diagnostic utility), this data is not provided at the individual level and is therefore difficult to evaluate. The extensive description in the technical manual of how users may use test results to identify cognitive strengths and weaknesses would appear to support claims made for its value in planning clinical and educational interventions. However, no evidence is cited or provided in direct support of these claims.

As discussed Chapter 1, contrary to the claims made by the test developers, critics have pointed out that IQ scores do not yield information that can be used in the diagnosis and treatment of learning problems (Reschly, 1997; Mc Grew, 1993; Siegal, 1992).

#### *4.1.3.4 Validity for children with autism*

In the technical manual (Wechsler, 1993 & 2004), data based on children with autism is provided; implying the test's usability for this clinical population. However, the language requirements and inclusion of tasks that require social knowledge cast doubt on the viability of the WISC-III in representing the learning abilities of children with autism. For example, difficulties in the development of social understanding is one of the core impairments in autism (Wing & Gould, 1979) and this might be expected to severely impact on a child's ability to 'show understanding of social rules and concepts', as is required in the Comprehension subtest of the WISC-III (See Table 4.1.3).

The use of global IQ scores in children with autism has long been criticized. Due to the varied and uneven profile of strength and weakness, the global estimates were felt to be an invalid reflection of the child's overall functioning (Frith, 1989, 1991 & 2003; Attwood 1998). Several studies have been conducted in the last three decades with a view to identify cognitive or IQ profiles of children with autism, as reflected in standardized IQ tests (e.g. Freeman, Forness & Ritvo, 1985; Lincoln, Courchesne, Kilman, Elmasin, & Allen, 1988; Lockyer & Rutter, 1970). In their review of these studies (which were based on Wechsler Intelligence Scales for Children) Lincoln, Allen & Kilman (1995) highlighted that:

- i. While most of studies reported good predictability and stability for global intelligent quotients for children with autism between preschool and school ages, exceptions to the findings of test stability were also found. For example, children with autism showed substantial changes over time in IQ scores and large increases were found between early estimates of IQ (using the Bayley Scale) and subsequent IQ scores. While the variability in IQ scores over time is true for all groups of children, it appears that the levels of instability or variability of IQ scores over time is higher for children with autism (Mayes & Calhoun, 2003).
- ii. Attempts to identify cognitive profiles based on Performance or Verbal IQ index have highlighted high inter individual variations. For example, while some studies show that children with autism have a higher Performance IQ, others reported the reverse pattern, i.e. higher Verbal IQ.

iii. More consistency is seen when cognitive profiles are based on individual subtests scores. A consistent finding is that children with autism show relatively better performance than controls in the Block Design subtests of the Wechsler.

From the above analysis, it appears that not all the claims made by the test developers regarding the intended purposes of the scale are supported by evidence. While evidence based internal structure seems to be fairly strong, evidence based on test content and relation to other variables is inconsistent. Despite the claims of the test's usefulness for educational planning, resource provision and placement decisions, the evidence for treatment validity (or instructional utility) is negligible. This thesis will address the issues of criterion and treatment validity of the WISC-III for children with autism (Studies 2 and 3).

## 4.2 MEASURE OF EXECUTIVE FUNCTION

### 4.2.1 What is executive function; how is it measured?

As discussed in Section 2.4.2.3, executive function is the ability to consider alternatives in planning. It does not refer to a single cognitive process, instead executive function is an umbrella term for all of the complex set of processes that underlie flexible, goal-directed responses to novel or new situations, in particular novel situations that involve:

- Planning and decision making;
- Error correction or troubleshooting;
- Initiation of novel sequences of actions;
- Danger or technical difficulty; or
- The need to overcome a strong habitual response.

(Shallice & Burgess, 1991; Shallice, 1982)

Early research in executive function used experimental tasks which were carefully designed to isolate specific behavioural responses associated with executive function, such as the Tower Of Hanoi test (Shallice 1982), the Wisconsin Card Sorting test (Heaton, 1993) and the classic Stroop test (Stroop, 1935).

- In the Tower Of Hanoi test, individuals must move disks from a prearranged sequence on three different pegs to match a goal determined by the examiner.

However, conditions are imposed, e.g. the individual must complete the task in as few moves as possible. This test focuses on one main executive function, namely planning.

- Another typical executive function test is the Wisconsin Card Sorting test, which is seen as mainly a test of mental flexibility (or set-shifting). In this test, an individual must first sort out cards on one of three possible dimensions (colour, number, shape) according to a non-spoken rule and is then required to shift rules to sort cards along a different dimension. The experimenter tells the participant whether she/he has placed the cards correctly (i.e. followed the correct rule), but does not give the participant the rule explicitly.
- The Stroop test is a well-known test of inhibition. In this test, a participant must first read a list of colour names written in black ink and then read a list of colour names written in coloured ink, where the ink colour is congruent with the colour word (e.g. 'blue' written in blue ink). Finally the individual must name the colour ink that words are written in (e.g. say 'red' to the word blue written in red ink). In this test, the individuals' ability to overcome the interference of one input (i.e. the word) on their performance of another (naming the colour) is seen as reflecting inhibition.

While these tests have been shown to be useful in isolating specific process associated with executive function among adults (Pennington & Ozonoff, 1996) their use as a measure of overall executive function in children is somewhat limited. This is due to the lack of standardized administration/scoring procedures and developmental norms. In addition, there are several methodological issues when these tests are used with children with language and communication disorders, e.g. autism.

#### **4.2.2 Measurement Issues**

Hughes & Graham (2002) highlighted three ways in which limitations in language among children may affect their responses in executive function tasks:

1. Children with limited language may find that the verbal instructions in these tasks may tax their verbal comprehension. This places additional demands on mental processes, which may in turn, influence the children's overall performance.
2. Many of the 'classic' executive function tasks depend on over-learned (i.e. automatic) literacy skills, e.g. the Stroop test. As fluent literacy emerges late in development, especially for children with language and communication disorders, the Stroop task may not be appropriate.
3. Language abilities play a direct role in the control of actions. For example, children's vocabulary may limit their ability to store phonological information in working memory and to engage in mnemonic strategies such as verbal rehearsal. These strategies are seen as important facilitators in executive function (Shallice, 1983).

In addition to the limitations highlighted by Hughes and Graham (2002), another source of difficulty in using the test with children concerns the issue of task demands. Most of the executive function tasks include non-executive skills, such as recall, psychomotor dexterity (e.g. tasks that require physical manipulation of small apparatus), writing skills, visual discrimination skills and spatial skills. Tasks for children must be designed such that the additional demands placed on these non-executive skills are kept to a minimum.

In recent years, several standardised tests for measuring executive function in children have been developed (See Table 4.2.2). Most of these tests were originally developed for the adult population, and the adaptations include simplifying the language demands and using test-formats which are more suitable for children, e.g. game-format, computer presented stimuli, colour coded and illustrated materials.

However, most of the standardised tests on executive function for children focus only on one aspect/dimension of executive function. For example, the Test of Everyday Attention for Children, or TEA-Ch (Manly, Nimmo-Smith, Watson, Anderson, Turner & Robertson, 2001) focuses on attention; while the Maudley Attention and Response Suppression, or MARS (Rubia, Taylor, Smith, Oksanen, Overmayer, & Newman, 2001)

focuses on impulsivity. A notable exception is the Behavioural Assessment for Dysexecutive Syndrome for Children, or BADS-C (Emslie, Wilson, Burden, Nimmo-Smith & Wilson, 2003) which captures a wider range of core difficulties associated with impairments in executive functions, such as mental inflexibility, novel problem solving, impulsivity and planning (See Table 4.2.2).

**Table 4.2.2 : Standardised Test for Executive Function in Children**

Test	Earlier versions	Task features	Focus	Psychometric data
Test of Everyday Attention for Children (TEA-Ch)	TEA (adults version)	8 manual tasks in game-like format.	3 aspects of attention: <ul style="list-style-type: none"><li>• Sustained attention</li><li>• Selective attention</li><li>• Attentional control</li></ul>	Developmental norms for 6 to 16 years.
Maudley Attention and Response Suppression (MARS)	-	Computerized tasks	3 aspects of impulsivity: <ul style="list-style-type: none"><li>• In response to inhibition;</li><li>• In sensori-motor coordination;</li><li>• In time estimation.</li></ul>	No published norms
Behavioural Assessment for Dysexecutive Syndrome for Children (BADS-C)	BADS (adults version)	6 tests in a variety of format, e.g. pen and pencil, manipulation of apparatus.	7 executive function processes: <ul style="list-style-type: none"><li>• Mental flexibility;</li><li>• Novel problems solving</li><li>• Sequencing</li><li>• Using feedback</li><li>• Planning</li><li>• Impulsivity</li><li>• Following instructions</li></ul>	Developmental norms for 8 to 16 years, adjusted by IQ.

In the present thesis, the scope of SEN is defined in terms of broad and comprehensive domains of functioning. Thus, it would be important to consider measures that include as wide as possible, the repertoire of abilities associated with executive function. In addition, executive function is often seen as multi-componential or fractionated (Shallice & Burgess, 1991); and (as discussed in Chapter 2), it is still unclear which aspects of executive function are impaired in children with autism. From the three tests identified, the BADS-C appears to be better suited for this thesis, as it covers the widest range of executive function skills.

#### 4.2.3 Review of the BADS-C

##### 4.2.3.1. Test description

Published in October 2003, the BADS-C is an individually administered, standardised test, which has been adapted from the original version for adults (Wilson et al., 1996) It comprises of 6 subtests, namely Playing Card Test, Water Test, Key Search Test, Zoo Map Test 1, Zoo Map Test 2 and Six Part Test. Key features of the 6 subtests in the BADS-C are described in Appendix G; Table 4.2.3.1 is the description for one of the subtest, the Six Part Test.

Table 4.2.3.1 : Description of the Six Part Test in the BADS-C

<b>Tasks involved</b>	<b>Information recorded</b>	<b>Scoring criteria</b>
<b>Six Part test</b>		
<p>The aim of this test is to see the children's ability to plan, schedule task and self-monitor performance.</p> <p>Children are given 3 colour coded tasks to do : green tasks (simple arithmetic); blue tasks (picture naming) and red tasks (sorting). Each of the tasks has 2 parts, i.e. part 1 &amp; parts 2, so in total, there are six parts to the test.</p> <p>The children's task is to attempt each of the six parts within a time limit of 5 minutes. They are not required to complete all of the items (impossible to do), but rules/conditions are given, i.e. the children cannot move from one part of one colour to another part of the same colour consecutively.</p>	<p>Order of parts attempted by child.</p> <p>Time spent on each of the six parts.</p> <p>Number of times rules/conditions are broken.</p> <p>Total time for completion of tasks.</p> <p>Whether or not the child is aware that s/he has broken the rules</p>	<p>Scores awarded for:</p> <ul style="list-style-type: none"> <li>• Each part attempted</li> <li>• Use of clear strategy</li> </ul> <p>Deduction of scores for:</p> <ul style="list-style-type: none"> <li>• Each task in which the rule was broken</li> <li>• If child returns to any part of the test 3 times or more (i.e. perseverating).</li> </ul>

##### 4.2.3.2 Analysis of content validity

A key question concerning the validity of the BADS-C is whether the test adequately taps the skills and processes associated with executive function. In the BADS-C, executive function is defined as:

*...a set of behavioural competencies which include the ability to plan, sequence behaviours, sustain attention, resist interference, utilize feedback, coordinate simultaneous activity, change set and generally deal with novel situations (BADS-C Manual, pg. 5).*

However, although the 6 subtests tap a variety of executive functioning, these cannot be seen as exhaustive or comprehensive. For example, inhibition (which is the ability to

suppress an over-learnt, i.e. automatic, response) was not included. The classic inhibitory test is Stroop test, which relies on fluent literacy skills and thus perhaps not suitable for children. However, no alternative test for inhibition was included in the BADS-C, even though it is an important element of executive function (Shallice & Burgess, 1991).

The developers of the BADS-C identified seven specific areas of executive functioning that are examined in each of the 6 subtests (see Table 4.2.3.2.1 below).

Table 4.2.3.2.1 : Executive Function Processes in Each Subtests (adapted from BADS-C Manual, pp 24)

	Playing Card	Water test	Key Search	Zoo Map test 1	Zoo Map test 2	Six Part
Executive functions identified by BADS-C developers						
1. Flexibility and perseveration	✓	✓			✓	✓
2. Novel problem solving		✓	✓	✓		
3. Sequencing		✓		✓		✓
4. Using feedback		✓	✓	✓	✓	✓
5. Planning			✓	✓		✓
6. Impulsivity			✓	✓	✓	
7. Following instructions.				✓	✓	✓

The identification of the executive function processes in the 6 subtests was provided to ‘assist clinicians in interpreting results’ (BADS-C Manual, pp 23). However, there are some difficulties in using these seven areas of executive functions in analysing the contents of each subtest. The operational definitions of the seven processes were not provided and there was significant overlap in the seven executive functions included in the 6 subtests. For example (See Table 4.2.3.2.1);

- Only the Zoo Map and Six Part tests were highlighted as involving the ability to ‘follow instructions’, although arguably, all the tests involve the ability to understand and adhere to verbal instructions.
- The function ‘using feedback’ was identified in all but one subtest (namely, Playing cards), while ‘flexibility and perseveration’ was identified in 4 out of 6 subtests.

Thus, a child's relatively low scores in any one subtest may be due to impairments in more than one of the identified executive functions. It would be difficult for users to make any conclusive interpretation based on subtest scores due to the lack of clear distinction between the different tasks and the underlying executive functions.

One alternative way of gauging the content validity of the BADS-C subtests is to compare the 6 subtests against key elements of executive functions that have been established in the literature. Reviewers of executive function research have highlighted three elements which can be regarded as the key denominators underlying all executive function tasks:

- The tasks require the ability to disengage from immediate environment/context to guide actions (Shallice, 1991);
- The tasks involve the execution of a novel (vs familiar) sequence of actions (Hughes & Graham, 2002); and
- The tasks involve making a choice between alternative responses (vs the execution of a single action responses) (Hughes & Graham, 2002).

Based on analysis of the 6 subtests according to the 3 key denominators above it can be seen that all the 6 subtests in the BADS-C meet the criteria as tasks that involve executive function processes (see Table 4.2.3.2.2). However, the *extent* to which specific executive functions are involved in each subtest is less clear.

Table 4.2.3.2.2 : Key Denominators of Executive Function in Each of the 6 Subtests

	Playing Card	Water test	Key Search	Zoo Map test 1	Zoo Map test 2	Six Part
Key denominators of executive function processes (Shallice, 1991; Hughes & Graham, 2002)						
1. Ability to disengaged from immediate contexts.	✓	✓	✓	✓	✓	✓
2. Execution of novel sequence of actions.	✓	✓	✓	✓	✓	✓
3. Making a choice between alternative responses	✓	✓	✓	✓	✓	✓

#### 4.2.3.2. Analysis of response bias

The format of tasks in the BADS-C is varied (e.g. the use of pen and paper tests, manipulation of apparatus). Although some of the subtests, such as the Six Part test, involves writing down answers for simple sums and names of objects, children's written responses in these tests were not scored, and hence literacy abilities are not a prerequisite. However, it can be argued that for children with poor writing and numeracy skills, these tasks may place additional burden on the children's working memory, which is an important component of executive function skills (Shallice, 1991). Consequently, children with less fluent literacy or numeracy skills will be disadvantaged in the Six Part test.

Although the instructions for the tasks have been simplified, these still entail some level of verbal complexity. The children must be able to understand and retain a series of verbal instructions in working memory in order to perform the tasks. For example, the following are the instructions for Key Search, which has comparatively, simpler instructions than other subtests:

*I want you to imagine that this square is a large field. Somewhere in this field you have lost your keys. You don't know exactly where you have lost them because you have been all over the field, all you know is that they are somewhere in the field. Starting from this dot I want you to draw a line with the pen to show me where you would walk to search the field to make absolutely certain that you would find your keys. Although I will be timing you, there is no time limit. Take as long as you like to search the field but make sure you search carefully so that you are certain to find the keys no matter where they are. Here's the pen, start from the dot and draw a line to show where you are walking to search the field. (BADS-C Manual, pp 11).*

No demonstration or 'teaching items' are included, and from the above instructions the children are expected to produce a systematic and implementable plan of search for the keys. Due to the high language demands, the basal age limit for the BADS-C is 8:0 years. However, as children with autism commonly experience significant delays in language, it is important to establish the minimum language abilities required to perform the tasks adequately, independent of the child's chronological age.

#### *4.2.3.3. Analysis of norms*

In the BADS-C, children's raw scores are converted into scaled scores (range 1 to 19, mean = 10, SD = 3), which are adjusted for the child's age and estimated IQ. Age bands are given in 12 months' intervals, while estimated IQ is divided into 3 bands (70-89; 90-110; and >111). The rationale provided for the adjustment by age and estimated IQ is that the 'performance on the test in general varied strongly with age and varied moderately with estimated IQ' (BADS-C manual, pg 14). The adjustment of standard scores by age and IQ is useful when making group comparisons, in particular where children's IQ and ages vary widely.

However, the age bands provided in the BADS-C norms tables are in intervals of 12 months. It is arguable whether such large age bands can adequately reflect developmental differences. In addition, the basal for the lower IQ band is 70 IQ points, implying that for children with IQ less than 70, no exact comparable norms are available.

The IQ estimates in the normative sample are based on children's scores on a reading test, namely the Wechsler Objective Reading Dimensions (Rust et al., 1993), or WORD. The UK version of the WORD and the WISC-III (UK) were co-normed and a strong relationship in children's scores in the two tests was established. This correlation was the basis for using children's WORD score to predict IQ estimates<sup>13</sup>. There are several issues with this method of IQ estimate: firstly, based on the published normative data (Rust et al., 1993) children's WORD scores were most strongly correlated with Verbal IQ rather than overall or Full Scale IQ. It would be more accurate to use Verbal IQ scores as the basis of interpretation for children's score on the BADS-C. Secondly, the correlation between reading (WORD) and IQ (WISC-III) scores were established based on the responses of normally developing children. It cannot be assumed that the same relationship holds true for children with specific impairments in language, such as autism. It could be argued that for some children with autism, the conversion of raw scores into BADS-C scores, based on IQ estimates derived from reading scores, may not be valid.

---

<sup>13</sup> This information is based on the researcher's personal communication with Mr Paul McKeown, the director of Harcourt Assessments (Europe); email correspondence dated 23 Apr 2004.

#### *4.2.3.4. Analysis of reliability*

In the BADS-C manual, inter-rater reliability for the six subtests was reported to be high; the lowest inter-rater reliability index was for the scores for perseveration ( $r=0.53$ ). Although significant, given that this is a standardized test with scripted administration and scoring procedures, 0.53 is a relatively weak correlation. Perseveration is defined as occurrences when the child repeats a ‘previous or inefficient response 3 or more times’ (BADS-C manual, pg 11). However, there was no definition for an ‘inefficient’ response, which makes it open to subjective interpretations by the tester/scorer. For research purposes, ‘inefficient response’ may need to be clarified or defined a priori.

There were no indices for internal reliability for the BADS-C, and no explanation for its absence. The use of an overall score for BADS-C (aggregated from the total subtests scores) assumes that each of the 6 individual subtests is part of an overall construct. Some degree of consistency between children’s scores in individual subtests and overall scores is therefore assumed. In the present thesis, where the aim is to use the BADS-C as a overall index of executive function, rather than focus on children’s profile in specific subtests, it is important to obtain a gauge of the internal reliability of the measure.

#### *4.2.3.5 Summary and conclusion*

Several features of the BADS-C make it suitable for use in the present thesis:

- It includes a wide range of tasks which (taken as a whole) examine a fairly broad repertoire of executive functions skills. Compared with other standardized tests for children, the BADS-C captures a wide range of processes associated with executive function. In addition, all the tasks in the 6 subtests involve the key elements/denominators of executive function established in the literature, namely the ability to disengage from immediate environment/context, and the execution of a novel sequence of actions and a choice between alternative responses. This provides some evidence for content validity.
- Inter-rater reliability of the test was reported to be high.

- The test scores were adjusted for age and IQ. This is an advantage for the present thesis, as the participants are expected to have different ages (8 to 12 yrs) and ability levels. However, the basis of estimation of IQ in the BADS-C rests on an assumed link between IQ and reading skills, which may not be applicable for children with autism.

In addition, several issues were highlighted:

- Although the test protocol were adapted and simplified for children, it cannot be assumed that the adaptations would be adequate for children with language and communication disorders, such as autism.
- No internal reliability measure was reported and the consistency between children scores in individual subtests and overall scaled scores is unknown. It would be important to establish the internal consistency reliability for the purpose of the present thesis, as the aim would be to obtain a global (i.e. overall) measure of executive function in children with autism.

To address these issues, a field-test was conducted with a small sample (N=10) of children with autism aged 8 to 12 years (See Appendix H). The findings of the field-test indicated that the internal reliability of the Overall BADS-C scores was adequate (alpha greater than 0.6), and that there was a marked difference in the scores of children with autism and those reported for the standardisation sample. The average scale score for children with autism were 14 points lower than the normative sample. The results of the field test provided additional support for the use of BADS-C as a measure of executive function of children with autism, aged 8 to 12 years.

#### **4.3 MEASURES OF CENTRAL COHERENCE**

##### **4.3.1. What is central coherence; and how is it measured?**

As discussed in Chapter 2, central coherence refers to the propensity, hypothesised to be in-built, to form coherence over as wide a range of stimuli, and to generalize over as wide a range of context as possible (Frith, 1989). Studies have indicated that in children with

autism this capacity for coherence is diminished; ideas and thoughts are ‘detached’ from context and lack meaningful connectedness from one another.

The concept of strong and weak central coherence maps well with the notion of ‘field-dependence-independence’ in cognition, which was developed by Gestalt psychologists Witkins, Oltman, Raskin & Karp (1971). Field independence refers to a lack of influence of context both in visual perception and social interaction. In a field independent mode of perceiving, parts of the field are experienced as discrete from the organized ground. Consequently, individuals with strong field independence have a tendency to disregard context. In contrast, individuals with strong field dependence are dominated by the overall contextual organisation of the surrounding field, and consequently parts of the field and ground are experienced as ‘fused’, i.e. inseparable (Witkins et al. 1971, pp. 4).

The initial measures for central coherence were based on a test that were developed for assessing field-dependence/independence, namely the embedded figures test (Witkins et al., 1971). In a landmark study by Shah and Frith (1983), the *Children’s Embedded Figure Test (CEFT)* was used, which involves detecting a hidden figure (e.g. a house) among larger meaningful drawing (e.g. a rocking horse). Shah and Frith found that children with autism were significantly superior in the task, which requires field-independent perceptual skills. The strong field independence is seen to reflect a diminished central coherence. This is the key implication of the central coherence theory, i.e. that individuals with autism have a unique profile of perceptual and cognitive abilities in which superiority in processing local, detail level information is contrasted with inferiority in processing global and contextual information.

Other paradigms for testing field-dependence-independence have been used as measures of central coherence, e.g.:

- *Silhouette identification task* (Humphreys & Riddoch, 1987). This task involves black and white drawings of real objects (e.g. boat) and other non-objects made up of two real objects (e.g. a pistol with a trumpet as the barrel) that were recognizable either from their internal features or from their outline. Each

drawing had a line drawing version (i.e. where details are present), and silhouette version. Participants were measured in terms of their reaction times to determine whether each figure was a real object or not. Findings showed that children with autism showed no difference in reaction times between the two versions of the stimuli, indicating that unlike normally developing individuals, the responses of children with autism did not improve when provided with contextual details.

- *Hierarchical letter task* (Lamb, Robertson & Knight, 1990). This is a computer generated task involving white and black global patterns of H, S, A and E formed from smaller local patterns of the same letters, all presented in black background. Thus, the stimulus consisted of ‘an H made up of A’s’ or ‘H’s forming an A’. Participants’ reaction time was recorded each time they identified the presentation of the target letters H & S (A & E were distractors). Differences in the response time and accuracy between individuals with autism and normally developing participants suggested that individuals with autism were focusing on details, rather than the larger configuration of the letters.
- *Semantic processing tasks*. Evidence for weak central coherence among individuals with autism has also been examined using verbal/reading tasks. For example, using tasks involving memory of related and unrelated list of words, Tager-Flusberg (1991) found that children with autism are less likely to use related category information to aid recall. Using Snowling and Frith’s (1986) homograph reading task, Happe (1997) found that children with autism, even those with high verbal IQ, failed to use sentence context to disambiguate homograph pronunciation (e.g. in distinguishing between: ‘There is a big tear in her eye’, and ‘In her dress, there was a big tear’).

#### 4.3.2 Measurement issues

One of the issues in testing central coherence in children is the use of tasks that require fluency in specific skills, such as literacy (letter/number recognition) and visual discrimination. Many of the experimental tasks for central coherence involving sentence reading and letter detection may not be developmentally appropriate for children.

As the different tasks involve different demands, the different test paradigms are not equally sensitive in detecting weaknesses in central coherence. A recent study by Mottron, Burack, Iarocci, Belleville & Enn (2003) compared the responses of high functioning adolescents (15 years old) with autism with IQ matched controls on several tasks, including a traditional hierarchical letter task, a configural processing task (similar to the Silhouette task) and a disembedding task (similar to the embedded figures task). It was found that the disembedding task showed the strongest difference between the two groups.

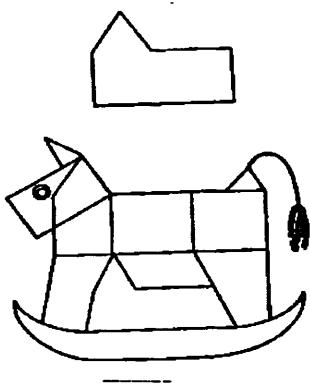
Another difficulty in using experimental tasks is the lack of standardization, where variations in stimuli design, administration procedures and outcome measures used (e.g. reaction time, response accuracy or error rate) would affect the variance observed in the data. The Children's Embedded Figure Test (CEFT) is the only standardised test which has been shown to be able to distinguish children with strong/weak central coherence (Shah & Frith, 1983).

#### **4.3.3. Review of the Children's Embedded Figures Test (CEFT)**

##### *4.3.3.1 Description of the CEFT*

The CEFT was constructed as an adaptation for children of the 'embedded figures test' (Witkins et al., 1971). It includes 24 test items, in which a simple standard shape (a house) had to be detected within a complex, meaningful, embedded drawing, e.g. a rocking horse (See Fig. 4.3.3.1):

Fig. 4.3.3.1 : Sample item from CEFT - 'house' embedded in 'rocking horse'



The CEFT is presented following standardized administration procedure. The children are first given a set of discrimination items, so as to establish their perceptual ability to recognize and distinguish the target shape (i.e. the house). During this discrimination exercise, the tester emphasises the need for accuracy in terms of the size, shape and orientation of the target shape. If the child fails the pre-requisite visual discrimination tasks, he/she will be considered unsuitable for testing. The tester then gives practice items to familiarise the child with the test demands. During these practice items, teaching of the correct response is allowed. Children's responses are scored in terms of accuracy in detecting the target shape.

#### *4.3.3.2 Analysis of response bias*

The CEFT does not involve expressive language skills; thus, children with poor communicating skills (such as those with autism) would still be able to respond adequately to the task demands. In addition, the items contain objects/scenes which are commonly depicted in children's books that are available in Singapore, e.g. houses, rocking horse and tent. For the participants in the present study, difficulties that might arise from the lack of cultural familiarity with the objects/pictures in the CEFT would be minimal.

#### *4.3.3.3 Analysis of norms*

The CEFT manual provides a table of norms which enables children's raw scores to be converted into age-equivalent scores. The norms are in intervals of 24 months, which may be too large to reflect developmental differences in young children. For research purposes, the use of total raw scores as the outcome measure would be more appropriate.

#### *4.3.3.4 Analysis of reliability and validity*

Internal reliability estimates of the CEFT were high; ranging from 0.83 to 0.90. Validity estimates were reported based on strong correlations between CEFT scores and its precursor test, the Embedded Figures Test (EFT) (Wilkins et al., 1971). However, these findings were based on children aged 9 to 12 years, and as such, do not establish the

validity of the test at a younger age, except by inference. Moreover, as the CEFT is designed as an extension of the EFT, a high correlation between the two tests would be expected.

Criterion related validity of the CEFT was also reported based on significant correlations between the CEFT (Witkins et al., 1971) and subtests of the Wechsler Intelligence Scales for Children, or WISC-III (Wechsler, 1993). CEFT scores correlated significantly ( $r = 0.49$ ) with scores on the WISC-III subtests which involved perceptual discrimination skills and figure/ground distinction, namely the Block Design, Object Assembly and Picture Completion subtests. On the other hand, no significant correlation was found with WISC-III verbal comprehension scores, suggesting a distinction between cognitive abilities associated with central coherence, and children's language abilities.

#### *4.3.3.6. Conclusion*

Based on the review, CEFT appears to be a suitable measure for use in the present thesis. Although it focuses on one dimension, namely field-dependence in a visual processing task, it has been reported to show adequate reliability and criterion validity for use with children. Although in terms of its content the CEFT taps only one ability domain, namely perceptual field-dependence, studies have shown that compared with other central coherence tasks, the embedded figures demonstrated the strongest evidence for discriminant validity.

As the CEFT was normed on a sample of normally developing children, it was suggested that when used with children 'with varying degrees of intellectual impairment, pilot testing (or field-test) could be used to determine the suitability of the CEFT for particular groups to be tested (Witkin et al., 1971, pp 25).

The findings from the field-test for the present thesis, which is reported in Appendix I, provided additional support for the use of CEFT as a measure of central coherence in children with autism: 1) the children in the sample were able to cope with the task and language demands of the test; and 2) there were no ceiling or floor effects. The CEFT

total score appears to be a more appropriate outcome measure than the CEFT age-equivalent score which did not provide adequate discrimination of the children's performance in the test.

## **4.4 MEASURE OF COGNITIVE MODIFIABILITY**

### **4.4.1 What is cognitive modifiability; how is it measured?**

Cognitive modifiability, or the extent to which a child's learning can be changed with teaching/instruction is measured through the process of dynamic assessments (see discussions in Chapter 2). This approach has the following characteristics (Lidz, 1987):

1. The tests involve interaction between the tester and the child;
2. There is a focus on the learner's metacognitive processes and responsiveness to intervention; and
3. Assessment procedures, which follow a pretest-intervention-posttest format.

There is no one package of materials that typifies dynamic assessment; rather it is an approach that has many interpretations, degrees of standardization and applications to a wide variety of content. Campione and Brown (1987) suggested three dimensions along which dynamic assessment approaches can be differentiated: 1) focus; 2) interaction; and 3) target. In addition to the 3 dimensions suggested by Campione and Brown, an important fourth dimension is the intended (i.e. principal) use of the test. This is because the utility of tests is maximised when they are used for the population and purposes that they were intended to serve (APA, AERA & NCME, 1999).

Because of its unique characteristics, some commentators have argued that the standards and criteria used in the evaluation of 'static' tests, such as psychometric analyses of reliability or validity, cannot be directly applied to dynamic assessments (Grigorenko & Sternberg, 1998; Carlson & Wiedl, 2000). However, if dynamic assessments were intended to fulfil the same functions as other psychological assessments, it can be argued that the same standards of evaluation should also be applied.

**Table 4.4.1: Dimensions of Dynamic Assessment Approaches**

<u>Dimension</u>	<u>Continuum of Variation</u>	
<b><u>1. Focus</u></b> - method of measuring change potential	<u>Test-teach-test procedures.</u> Change potential is defined in terms amount of improvement observed in the child's performance.	<u>Diagnostic procedures.</u> Change potential is defined in terms of tester's descriptions of the child's use of specific processes or deficient cognitive functions.
<b><u>2. Interaction</u></b> - type of tester-testee interaction	<u>Structured &amp; standardized</u> , e.g. a standardized interview, counting the number of prompts given to the child.	<u>Clinical</u> , i.e. the examiner adapts his or her administration of the test to the child's responses. Thus the interaction is highly individualized and unstructured.
<b><u>3. Target</u></b> - types of assessed skills/processes.	<u>Content specific</u> skill areas, e.g. reading, maths, specific curriculum related skills.	<u>Generalised across content areas</u> e.g. categorization, analogies, and other abstract reasoning skills.
<b><u>4. Principal Use</u></b>	<u>Clinical population</u> , e.g. children or adults with specific learning disabilities.	<u>General population</u> , e.g. adults, preschoolers, school-aged children.

In the next section, dynamic assessments are reviewed based on the four dimensions, as well as criteria from the Standards for Psychological and Educational Testing (APA, AERA, NCME, 1999), which were described in Chapter 2. The aim of the review is to identify a dynamic assessment that can be used in the present thesis, i.e. to compare the validity of measures for cognitive modifiability against other indicators of special educational needs for children with autism.

#### **4.4.2 Review of dynamic assessments**

Table 4.4.2 lists several key dynamic assessment tests for children that have been published in English<sup>14</sup> and describes the tests along four dimensions: focus; interaction; target skills; and principal use (or intended population).

<sup>14</sup> Dynamic assessment tests for children in other languages are excluded in this review, e.g. the Leipzig Learning Test, in German (Guthke & Beckmann, 2000); the Learning Potential for Inductive Reasoning, in Dutch (Resing, 2000).

#### *4.4.2.1 Focus*

The focus in some dynamic assessment tests, e.g. the Learning Potential Assessment Device, or LPAD (Feuerstein, Rand & Hoffman, 1979), is to observe and record qualitative information about the child's behaviour / responses during learning, such as the emergence of new strategies used by the child, or elimination of deficient/ineffective cognitive functions. The outcome of the assessment is qualitatively rich data about the child's learning, as well as the strategies that are needed to bring about change potential (i.e. cognitive modifiability). In contrast, other dynamic assessment tests focus on obtaining quantitative measures reflecting changes or improvements in the child's performance, e.g. Cognitive Modifiability Battery, or CMB (Tzuriel, 1995), and Swanson Cognitive Processing Test, or S-CPT (Swanson, 1996). As the aim of the present thesis is to compare the validity of the measure of cognitive modifiability with other indicators of SEN, the use of dynamic assessment tests that focus on the test-teach-test procedure is arguably more appropriate, as these would yield quantitative data that can be used as a basis for comparison.

Table 4.4.2 : Key Dimensions of a Selection of Dynamic Assessments

Name of test	Author	Focus	Interaction	Target Skill	Intended Population & Principal Uses
<b>Learning Potential Assessment Device (LPAD)</b>	Feuerstein, et al. (1979)	The test focuses on identifying deficient cognitive functions and relates these to a 'cognitive map' that covers task parameters.	Unstandardised approach is used in which mediation is provided in whatever fashion the clinician feels is appropriate for maximizing the testee's performance.	A variety of cognitive operations in different domains.	Low achieving children. To ascertain the child's modifiability, and to guide intervention aimed at mediating the child deficient cognitive functions.
<b>Application of Cognitive Functions Scale (ACFS)</b>	Lidz (2000)	The test yields both description of the cognitive processes shown by the child, and the extent of developmental improvements made by the child in the respective learning activities.	Semi-standardized system of administration.	A variety of cognitive learning processes and learning strategies associated with typical early childhood learning activities.	Children aged 3 – 5. To provide evidence of the child's development of cognitive functions related to typical demands of preschool curricula.
<b>Cognitive Modifiability Battery (CMB)</b>	Tzuriel (1995)	The <u>clinical version</u> yields qualitative data, e.g. observations about the child's behaviour and responses to responses to mediation.  The <u>measurement version</u> yields quantitative data, i.e. child's scores on Pre-teaching, Post-teaching and Gain scores.	In the <u>clinical version</u> , mediation is unstructured and on-going, i.e. tester adapts administration of the test according to child's responses.  The <u>measurement version</u> involves structured, Pre-Teaching, Teaching and Post-Teaching phases. Interaction in the Pre & Post Teaching Phase is standardized.	Nonverbal reasoning skills (e.g. analogy, memory, seriation), which are seen as transferable across, content areas.	Kindergarten to Grade 3, and older children with learning difficulties. <u>Clinical version</u> provides description of child's cognitive skills and non-intellective factors. <u>Measurement version</u> yields gain scores that are intended to provide an indication of cognitive modifiability.
<b>Swanson – Cognitive Processing Test (S-CPT)</b>	Swanson (1996)	The test yields 3 standardized scores: <ul style="list-style-type: none"> <li>Initial score, i.e. highest level of unassisted performance;</li> <li>Gain score, i.e. highest score obtained under probing condition;</li> <li>Probe score, i.e. number of prompts or hints necessary to achieve the higher score under probing condition.</li> </ul>	Standardized testing and set of prompts that emphasize sequential processing strategies.	Single cognitive process, i.e. working memory.	Specific learning disabilities, 5yrs old to adult.  Provides an index of processing potential that can be used in the identification of children with learning disabilities / SpLD.

#### 4.4.2.3. *Interaction*

Dynamic assessments differ in the extent to which the interactions between tester and testee are structured. For some tests, e.g. the LPAD, the instructions and mediation given by the tester is individualised for each testee, and the practitioner is free to use any technique that he/she feels appropriate for maximising the child's performance. The problem with this approach is that the conditions in which the children's performances are measured differ significantly. While the individualised instruction may have provided the optimal environment for diagnostic purposes, the data obtained have limited value for comparison purposes.

To overcome this issue, some dynamic assessments have attempted to provide some degree of standardisation in the administration procedures. For example, in the Application of Cognitive Functions Scale, or ACFS (Lidz, 2000), a semi-structured system of administration is used, where the tester is provided with guidelines on the type of the activities that the child could be engaged in, and the intervention/ mediation that could be given for the different responses. However, the tester needs to exercise his/her judgment to decide the frequency and choice of mediation strategies that are optimal for each child. Although the test provides a quantitative gauge of change potential (in terms of the number of developmental tasks that the child had mastered without mediation), the extent to which the scores from different children can be compared is limited.

As the aim of present thesis is to systematically compare children with different levels of special educational needs across a variety of indicators of SEN, it is important to ensure that the data is obtained from uniform/standardised procedures. In the CMB, standardised administration is given in the pre and post teaching phases of the test, while in the teaching phase, the type and nature of assistance provided is unstandardised and no quantitative scores are obtained. Differences between the pre and post-teaching scores can provide a quantitative index of the child's cognitive modifiability.

In the S-CPT, the interaction between tester and testee, including the nature and type of prompts used in the mediation is highly structured and controlled. During testing, hints (or probes) are presented to the examinee following an erroneous response. (See Table 4.4.2.3).

Table 4.4.2.3 : Sample of probes used for the Rhyming Task in the S-CPT

<u>Test Description</u>	
<i>Test sequence</i>	<i>Examiner's questions and probes</i>
Test item	Stimulus: 'car-star-bar-far' Question: Can you tell me all the words in order?
First error	Probe : The last word in the sequence was 'far' Question : Now can you tell me all the words in order?
Second error	Probe : The first word in the sequence was 'car' Question : Now can you tell me all the words in order?
Third error	Probe : The middle word in the sequence was 'star' and 'bar'. Question : Now can you tell me all the words in order?
Fourth error	Probe : All the words in the sequence were 'car-star-bar-far' Question : Now can you tell me all the words in order?

The probes and hints used (See Table 4.4.2.3) are mainly based on verbal cues/priming. Swanson's (2000) data suggested that this mode of priming is appropriate to produce significant improvements in recall, as the post-test scores were significantly higher than pre-test scores. However, as the post-test items were identical to the stimulus used in the pre-test, practice effects may have contributed to the significant increase in post-test scores.

There are also difficulties with the nature of the mediations (or hints) used. A key aim of any mediation is to adapt and tailor interventions to individual children's needs/responses. However, the hints used were mainly changes/reframing of the stimulus item (See Table 4.4.2.3), given in a controlled (and somewhat contrived) way. These cast doubt on the efficacy of the hints as a mediation strategy. Moreover, verbal cueing may not be the most optimal strategy for improving recall of children with autism, who mostly have significant impairments in communication. It can be argued that children with complex learning difficulties (such as autism) may need to rely on intervention or mediational strategies that are more visual, and focusing on non-intellective factors such as motivation and attention (Tzuriel, Samuels & Feuerstein, 1988).

#### *4.4.2.4. Target skill*

Dynamic assessments differ in their target skills: some tests focus on specific domain/skill areas, e.g. the ACFS which focuses on learning domains related to the preschool curriculum; others aim to measure changes in cognitive skills that are transferable across different content areas, e.g. the CMB which focuses on non-verbal problem solving skills. For the present thesis, the use of tests which target skills that are transferable across a variety of content areas is more appropriate, given that the scope of special educational needs encompass multiple domains of functioning (See definition of SEN, Section 2.1.3).

#### *4.4.2.5 Principal uses*

The earlier dynamic assessments, e.g. LPAD, are intended for low achieving and severely disadvantaged children. This can be linked to the socio-political background for the emergence of such tests. For example, in his psychological interventions with immigrants to Israel, Feuerstein came to the conclusion that the use of IQ tests resulted in many children improperly labelled as intellectually inferior and placed in special education. The LPAD was the culmination of his efforts to devise measures that could assess the children's capability to develop and improve their deficient functioning (Feuerstein et al., 1979). Consequently, the intended use of the LPAD was clinical and diagnostic, and the primary target population was school children with significantly low achievement. Similarly, the ACFS scale was intended for low achieving children in preschool, i.e. ages 3 to 5 years (Lidz, 2000).

While the developers of CMB and S-CPT share Feuerstein's concern that children's learning potential is often underestimated, in contrast they emphasise the importance of standardised administration and test reliability. The principal aim of these tests is not merely to obtain diagnostic information, but to measure change. In the CMB, the index of change is seen in terms of the gains shown by the child before and after mediation (Tzuriel, 1996); while in the S-CPT change is seen in terms of the difference in children's scores with and without prompts (Swanson, 1996). Given that the present research involves a systematic comparison of children with special educational needs, the use of a more standardised, reliable measures of change is more appropriate.

#### *4.4.2.6 Conclusion*

Based on the analyses of the focus, interaction, target skills and principal purposes, the Cognitive Modifiability Battery (measurement version) (Tzuriel, 1995) appears to be the most appropriate for use in the present thesis. The CMB focuses on quantifying change potential that is observed with and without mediation/assistance; the interaction between tester and testee are structured and standardised; and the CMB targets cognitive skills that are seen as generalisable across different content areas.

#### **4.4.3 Review of the Cognitive Modifiability Battery (CMB)**

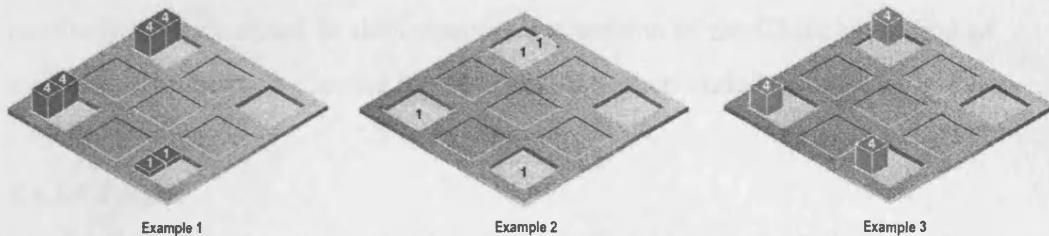
The CMB is described as “a dynamic assessment (DA) measure of learning potential as well as an intervention instrument for the development of cognitive operations, problem solving strategies, and deficient cognitive functions” (Tzuriel, 2001, pp 2). It was developed primarily for children in kindergarten to grade 3, but was also recommended for use with older children with learning difficulties in grades 4 to 9.

##### *4.4.3.1 Test description*

The CMB is composed of six separate subtests (see description of the six subtests in Appendix J). Materials used for the subtests include:

- Stimulus materials, namely 64 coloured blocks (yellow, red, green, blue), wooden plates (containing 9 windows) (See sample of stimulus in Fig. 4.4.3.1);
- Booklet of problems for examiners' use, which provides coloured illustration of the formation/arrangement of stimulus materials for each item in each subtest;
- Instruction manual detailing the administration procedures (for both the clinical and measurement versions) and suggestions for mediation. Specialised training in the use of the test and mediation techniques is required.

Fig. 4.4.3.1 : Stimulus Materials and sample Items for CMB Analogies Subtest



#### 4.4.3.2 Focus

As discussed in Section 4.4.2.1, for the purpose of the present thesis, the measurement version of the CMB, which provides quantitative index of cognitive modifiability, is more appropriate.

#### 4.4.3.3 Interaction

In the teaching phase, full mediation is given. Table 4.4.3.3 provides a summary of the overall mediation processes for the CMB, with specific examples of the strategies for the Analogies subtest.

Table 4.4.3.3 : Mediation Processes and Strategies for the CMB Analogies Subtest

Processes	Strategies
Preparation of the child for complex tasks by establishing pre-required thinking skills.	<ol style="list-style-type: none"> <li>1. Explicit teaching of pre-requisite tasks/concepts</li> <li>2. Motivational strategies to encourage/foster child's engagement in tasks.</li> </ol>
Self-regulation by planning and organization of the solution	<ol style="list-style-type: none"> <li>1. Intentional delay of children's response.</li> <li>2. Longer exposure to the problem.</li> <li>3. Systematic planning of the solution alternatives.</li> <li>4. Verbalization of the problem.</li> <li>5. Representation of the solution before pointing to the correct answer.</li> <li>6. <b>Highlighting impulsive behaviour.</b></li> </ol>
Enhancement of reflective, insightful and analytic processes.	<ol style="list-style-type: none"> <li>1. Focusing the child on the relation between his/her own thinking processes and the consequential cognitive performance.</li> <li>2. Encouraging reflective dialogues, e.g. 'What should we look at before we start to solve this problems? Or 'Why did you succeed in solving the problems that was so difficult for you to solve before?'</li> </ol>
Teaching specific contents that are relate to the task-specific contexts.	<ol style="list-style-type: none"> <li>1. Teaching use of terms top, bottom, left, right, same, different.</li> <li>2. Explicit teaching of transformational rules of analogy, e.g. 'Here (top) the yellow one becomes green. So what will happen to the yellow one here (bottom)?'</li> </ol>
Feedback on success or failure in the learning process.	<ol style="list-style-type: none"> <li>1. Providing specific praises for correct or partially correct answers.</li> <li>2. Providing feedback on wrong answers, providing reasons.</li> </ol>

However, because the judgment about the level and type of mediation given is left to the tester, the interaction between tester and testee in the teaching phase is relatively individualised. In the measurement version of the CMB, recording of the child's responses in the teaching phase is left open-ended.

#### *4.4.3.4 Target*

Although the six subtests use similar materials, namely coloured blocks and wooden plates with 'windows', the target skill involved in each task is different, ranging from the basic processes of ordering and copying (Seriation and Reproduction of Patterns subtests) to the more abstract processes involving analogical transformation (Analogies subtest). See Appendix J for a description of the target skills for all the 6 subtests.

No rationale was provided for the inclusion of these target skills in the CMB battery. Although Tzuriel (1995) made reference to Feuerstein et al.'s (1979) Modified Learning Experience (MLE) theory to guide the mediational strategies and approaches in the CMB, the selection and design of the items in the CMB was not rooted in the MLE theory. In his description of the development of the CMB tests, Tzuriel (1995, 2000) seems to place greater emphasis on the application and uses of the tests, rather than the theoretical basis for including these items in the test battery.

#### *4.4.3.5 Primary uses*

Tzuriel (2000) reported that since its publication in 1995, subtests from the CMB have been used in several studies looking at its use as a diagnostic tool and an intervention instrument for children with developmental delays (Lauchlan & Elliott, 1997; Jeffrey & Tzuriel, 1997, cited in Tzuriel, 2000). There are no published reports of the use of the CMB with children with autism.

#### *4.4.3.6 Scoring criteria*

In the CMB measurement version, two methods of scoring are available for children's scores in the pre and post-Teaching phases:

1. Method 1 – All or none: One point is given for each correct item.

2. Method 2 – Partial credit : Children’s responses for each item are analysed and one point is given for each dimension correctly solved. For example, in Analogies subtest, the maximum score for each item is four, corresponding to the dimensions of colour, number, height and position of the blocks.

For the present thesis, the decision to choose either Method 1 or 2 is based on the review of research on the reliability of the two scoring criteria and the field-test results, which are discussed in the next sections.

Children’s performance in the individual subtest is treated as discrete scores, i.e. there are no total/global scores for cognitive modifiability. This suggests that although they were developed and published as part of the same battery, each subtest is discrete and focuses on a particular skill area. It is also noted that in studies evaluating the reliability and validity of the CMB, the subtests were used as discrete or separate tests (Tzuriel, 1995; Tzuriel 2000).

#### *4.4.3.7 Analyses of equivalent norms*

From the children’s performance, three scores may be obtained:

1. Total score for pre-teaching items
2. Total score for post-teaching items
3. Gain score (i.e. difference between pre and post-teaching scores)

- *Total Pre & Post-Teaching scores*

These reflect the total number of correct responses made by the child. The CMB manual provided the means and standard deviations for 4 out of the 6 subtests.

The norms are grouped according to children’s education level – Kindergarten to Grade 3. Although the test is recommended for ‘older children with LD’ (Tzuriel, 1995, pp7), no norms are available for children beyond Grade 3. The grade equivalent scores may be more suited for clinical purposes, for example it can be used to gauge the extent to which a child’s pre or post-teaching scores are comparable to his/her peers. However, the utility of the grade equivalent scores for research purposes is less clear. In studies evaluating the reliability and

validity of the CMB (Tzuriel, 1995 & 2000) raw scores were used instead of grade equivalent scores.

- *Gain scores*

The CMB manual highlighted that ‘improvements (in pre and post teaching scores) were taken as indicators of cognitive modifiability’ (pp. 5). However, the criteria for interpretation are less clear. Unlike pre and post teaching scores, there are no norms for Gain scores, and no information to guide decisions regarding the adequacy or age-appropriateness of children’s Gain scores.

There are also some difficulties when Gain scores of children with different pre-teaching scores are compared. Within each subtest, the items have different difficulty levels: for example, in the Analogies subtest, the items progress from simple transformation involving one dimension (e.g. changes in colour), to more complex operations involving transformations along multiple dimensions (e.g. changes in size, colour and position). The same quantum in Gain scores may reflect very different progression in skills, depending on the difficulty level of the items that the child had mastered at baseline (i.e. pre-teaching scores).

To overcome this issue, Tzuriel (2000) used another variable derived from post-teaching scores, after removal of pre-teaching effects. These residual scores are based on a regression analysis of post by pre-teaching scores on each subtest (this procedure was first developed by Cronbach and Furby [1970]). Tzuriel argued that the adjusted, post-teaching score could be considered as an index of cognitive modifiability as it reflects the child’s post-teaching performance after controlling for the initial pre-teaching scores. One assumption behind its use is that the variance in post-test scores is mainly influenced by variance observed in pre-test conditions (Cronbach & Furby, 1970).

#### *4.4.3.8 Analysis of reliability*

Inter-item reliability of individual subtests all exceeded 0.6 alpha levels for adequate reliability (Tzuriel, 1995). Reliability coefficients for pre and post-teaching phases are very similar for all subtests, suggesting that the items used in pre and post-teaching phases have comparable reliability. In addition, reliability

coefficients based on the ‘partial credit’ scoring method are higher than for the ‘all or none’ method.

Reliability coefficients are also higher for subtests with higher abstraction/difficulty level, e.g. the Analogies subtests. The items in the Analogies subtest appear to have the highest ceiling, suggesting that it may be more appropriate for use with older children.

#### *4.4.3.8 Analysis of validity*

In a study to evaluate the predictive validity of the CMB, Tzuriel (2000) obtained Grade 1 children’s (n=35) responses to a reading comprehension test, a math test and selected subtests of the CMB (Seriation, Reproduction of Patterns, Analogies and Sequences). A step-wise regression analysis was used to predict the academic scores (math and reading comprehension) by CMB subtest scores. Results indicated that reading comprehension was predicted by post-teaching score for Analogies. Math scores were predicted by pre-teaching score for Reproduction of Patterns and post-teaching Seriation score.

Tzuriel interpreted the higher predictive power of the Analogies subtests as reflecting the fact “that the Analogies tap an abstract domain, which is closer to reading comprehension than Seriation” (Tzuriel, 2001, pp129). However, no empirical or theoretical rationale was provided for the inference. In the prediction of math scores, Tzuriel argued that the correspondence between math scores with the two CMB subtests was meaningful because “the Seriation and Reproduction of Patterns involves the need for precision and accuracy, a cognitive function that is required in math performance’ (Tzuriel, 2001, pp129). However, following his line of argument, one should expect that all CMB subtest will be predictive of math scores, as all the CMB subtests involve the need for accuracy and precision. This however, was not the case – three of the CMB subtests were not related to math scores. While there was some concurrence between children’s achievement scores (reading comprehension) and their scores on the Analogies subtest, the evidence to support Tzuriel’s conclusion that “the CMB scores are able to predict achievement scores” is inconsistent.

#### *4.4.3.9. Summary and conclusion*

Several features of the CMB make it appropriate for use in the present thesis:

- The availability of the measurement version provides a quantitative measure of children's improvement following mediation; this would enable a comparison between children's performance in CMB and other quantitative indicators of SEN used in the present thesis.
- The scripted interview protocol (without mediation) in the pre and post-teaching phases provides the standardisation that is required for a valid and fair comparison of children's scores;
- The internal reliability indices for most subtests were high; with higher reliability indices reported for the Analogies subtest, which also has the highest ceiling for item difficulty. This makes this subtest particularly suitable for use with children in the present study, who were slightly older than the population intended by the developers of the CMB. The 'partial credit' (Method 2) scoring criteria shows a stronger reliability than the 'all or none' method; and
- The Analogies subtest post-teaching scores have shown to be predictive of children's reading comprehension skills.

The administration of the full CMB battery would take about 8 hours per child; this is not feasible for research purposes. As each of the subtest is discrete, it would be possible to obtain an index of cognitive modifiability based on the administration of one subtest. Based on the review, the Analogies subtest appears to be most appropriate for use in the present thesis.

Although it has been used with older children with learning disability (i.e. developmental delays and speech pathologies), the CMB has not been used with children with autism. There may be some language and task demands that present particular difficulties for children with autism.

To address these issues, a field-test of the CMB Analogies subtest was carried out with a small sample of children with autism, age 8 to 12 years old (n=5) (See Appendix K). Results of the field-test provided some support for the use of CMB

Analogy subtest as a measure of children's cognitive modifiability. The children in the sample were able to cope with the language and task demands of the test. In addition, no floor or ceiling effects were noted. Children's scores based on the 'partial credit' method seem to better reflect variations in children's performances than the 'all or none' criteria. The adjusted post-teaching scores can be used as an outcome measure, as they reflect children's performance in the test after mediation, while removing the effects of children's scores at the pre-teaching phase.

#### **4.5 MEASURE FOR THEORY OF MIND (ToM) ABILITIES**

Unlike the indicators of SEN that has been reviewed thus far, there is currently no commercially developed standardised test for theory of mind (ToM).

Therefore in the present thesis, measures of theory of mind are selected based on experimental tasks that have been used in autism research. Consequently, in this section, which discusses the measure for ToM, a greater emphasis is placed on the analyses of content validity, and a small-scale study was conducted to obtain an initial gauge of reliability and validity.

##### **4.5.1 What is theory of mind; how is it measured?**

Theory of mind (ToM) refers to the ability to think about thoughts (or 'mentalising' ability). As discussed in Section 2.4.2.1, the theory of mind hypothesis suggests that individuals with autism lack this ability and consequently, are impaired in certain social, communication and imaginative skills (Baron-Cohen et al., 1985).

Studies of theory of mind have used two measurement approaches: the use of specific tests and the test-battery approach. In the former, specific tasks were designed to investigate the extent to which children with autism show impaired theory of mind abilities, compared to children with normal development. Some of these tasks were developed to assess children's concept of first-order false belief, i.e. the ability to think about another person's thoughts about an objective event (e.g. the Sally-Ann test [Wimmer & Perner, 1983; Baron-Cohen et al., 1985]); while others were developed to assess second-order false belief, i.e. the ability to think about another person's thoughts about a third person's thoughts

about an objective event (e.g. the Ice-Cream man story [Perner and Winner, 1985; Baron-Cohen, 1989]).

In the test-battery approach, instead of focusing on specific tasks, children's responses across a range of theory of mind tasks are aggregated; and it is this aggregate score that is used as an indicator of ToM abilities (e.g. Hughes & Dunn, 1998; Hughes, Adlam, Happe, Jackson, Taylor & Caspi, 2000). Although less commonly used, there are several advantages in the use of the test-battery approach. Studies using the test battery approach have found that the reliability of children's performance significantly increases with the use of aggregate scores. For example in Hughes and Dunn (1998) longitudinal study it was found that individual differences in aggregate scores were stable, with within-child correlations of at least 0.5 across two consecutive six month intervals. In another study, Hughes et al. (2000) investigated the internal and test-retest reliability of children's aggregate scores obtained from a battery of nine false-belief tests. Internal reliability of children aggregate scores was good (Cronbach alpha 0.84). Test-retest reliability of the nine items was between fair to moderate, with kappa coefficient ranging from 0.40 to 0.72. This is in contrast to studies that used the single task approach, where test re-test and internal reliability indices were reported to be poor (Mayes, Klin, Tercyak, Cicchetti & Cohen, 1996; Charman & Campbell, 1997). The improved reliability through the use of the test battery approach suggests that aggregate scores may smooth out 'noise' (or error variance) from task specific factors, thus giving a more consistent measure of the theory of mind ability. Unlike the specific task approach, where responses are coded as a pass or fail, in the task battery approach, children's performance is reflected in a range of aggregate scores, which may make it a more sensitive indicator of varying levels of ToM abilities.

In the present thesis, the aim is to see if variations in children's theory of mind abilities reflect variations in their levels of special educational needs. It is important that both variables, i.e. theory of mind and special education needs, are measured along a range of scores, with adequate variance. Thus, the use of aggregate scores would be more appropriate; however, there are other reliability and validity issues with this approach to measuring theory of mind.

## 4.5.2 Measurement issues

### 4.5.2.1 Reliability

Although there is evidence of improved internal and test-retest reliability with the test battery approach, these studies focused only on young children aged 5 to 6 years old, with no developmental or learning difficulties. Happe (1995), reviewing the data collected over 5 years from the several studies on theory of mind abilities, found that age, mental age and verbal ability correlated significantly with children's performance on false-belief tasks. In addition, children with autism appear to require a higher level of verbal ability to pass theory of mind tasks. Thus the high reliability in the test battery used by Hughes and Dunn (1998) and Hughes et al. (2000) with very young normally developing children, cannot be generalised to other populations, e.g. older children with autism.

### 4.5.2.2 Validity

The use of aggregate scores in theory of mind test battery is based on the assumption that the different tasks measure the same unitary construct of 'theory of mind'. The decision to include (or exclude) a particular test in the battery is often made based on empirical grounds, i.e. prior research using the test have demonstrated its usefulness in distinguishing children with and without theory of mind. There is no comprehensive theoretical framework that guides the selection of the tasks that are included in the test battery. As such, there may be other dimensions or aspects of theory of mind that have not been explored in research, which may have been excluded.

Although the test-battery approach seems more appropriate for use in the present thesis, evidence of its reliability and validity as a measure of theory of mind abilities for children with autism is needed. The test-battery used by Hughes et al. (2000) is not appropriate as it is intended for younger aged children with normal development. White et al. (in prep) conducted a study that involved the use of a variety of different mentalising tasks with children with autism in the target age range (8 to 12 years), and these were considered for the present thesis.

### 4.5.3 Theory of Mind Battery

#### 4.5.3.1 Analyses of content validity

In White et al.'s study, 8 tests for theory of mind were used (See Table 4.5.3.1.1).

**Table 4.5.3.1.1 Test Items in White et al.'s Study**

Test Items <sup>15</sup>	ToM Construct	Reference*
1. Sally-Ann	1 <sup>st</sup> order false belief	Baron-Cohen et al. (1985)
2. Smarties	1 <sup>st</sup> order false belief	Perner et al. (1987)
3. Ice-cream man	2 <sup>nd</sup> order false belief	Baron-Cohen (1989)
4. Birthday puppy	2 <sup>nd</sup> order false belief	Sullivan et al. (1994)
5. Cow	Interpretive diversity	Luckett et al. (2000)
6. Dalmation	Interpretive diversity	
7. Penny Hiding	Deception	Baron-Cohen (1992)
8. Strange stories: o ToM stories o Physical stories o Jumbled stories	Advance false belief	Fletcher et al. (1995)

\*Prior use of this version of the test item

- *Sally-Ann test*<sup>16</sup>

This is the standard first-order false belief test initially developed by Wimmer & Perner (1983) and adapted by Baron-Cohen et al. (1985). To pass this test, children were required to predict an action based on an attributed false belief and two memory control questions ("Where is the marble really?" and "Where did Sally put the marble?"). 80% of children with autism with normal IQ failed the Sally-Ann test, indicating a deficit in first order belief attribution (Leslie & Frith, 1988; Baron-Cohen et al., 1985).

- *Smarties test*

The 'Smarties' test is another first order false belief test. Perner et al. (1987) argued that experience and expectations play key roles in facilitating our understanding of other's false beliefs, i.e. children would be able to better comprehend others' false beliefs if they had a direct experience of making similar false-belief attribution. In the Smarties test, the child was exposed (first-hand) to the result of a false-belief attribution: he/she discovered that the box of Smarties actually contained a pencil. The child was then asked to predict how a friend would respond when faced with the same question/stimulus. Using the Smarties test, Perner 's (1987) showed that children with autism, despite passing

<sup>15</sup>The mentalising tasks used in the present thesis is presented in Table 4.5.3.2.

<sup>16</sup>For a full description of the adapted 'Sally-Ann' test, see Appendix M, Item 1: 'Jim-Rosie' test.

the control questions (i.e. they were aware and remembered that they had mistakenly thought the box had Smarties) were not able to realize that someone else would make that same error for the same reasons as they made it. At the same time, there were children who failed Sally-Ann test but passed the Smarties test, suggesting that some children with autism may be failing the Sally-Ann test due to deficits in understanding the pragmatics of the task, and not a deficit in first-order false-belief per se. The Smarties test appear to be useful in distinguishing children with autism who may have first-order false belief attribution, but who may not have the pragmatic skills to cope with task demands in the Sally-Ann test.

- *Ice cream man story*<sup>17</sup>

This test was initially developed (Perner & Winner, 1985) to distinguish children who are able to pass the more basic first-order false belief test, but show deficit in second level perspective taking, i.e. the ability to think about another person's thoughts about a third person's thoughts about an objective event. After listening to the Ice Cream Man story, in addition to the second-order false belief question, the child's responses to 3 control questions were noted: the child's awareness of reality; and memory for information. The control questions were added to ensure that in addition to accurate false belief attribution, the child has both the knowledge of the real location of the object and an accurate memory of its previous location. Studies indicated that while children with autism were as accurate as control groups in recalling and providing information about object location, only one child with autism passed the second order false-belief question in the story (Baron-Cohen, 1989). The Ice-Cream Man test seems to be able to differentiate between children with autism with first and second-order false belief attributions.

- *Birthday Puppy Story*<sup>18</sup>

The 'Birthday Puppy story' is a variation of the 'Ice-Cream Man' test and was developed to see the extent to which information processing demands could have

---

<sup>17</sup> For a full description of the adapted Ice Cream Man Story, see Appendix M, Item 4: 'Ice Cream Man' test.

<sup>18</sup> For a full description of the adapted Birthday Puppy Story, see Appendix M, Item 5: 'Birthday Present' test.

affected children's responses (Sullivan, Zaitchick & Tager-Flusberg, 1994). It was argued that in the original 'Ice-Cream Man' story by Perner and Wimmer (1985), in addition to the ability for second-order perspective taking, the children must also have high level of linguistic competence, namely an understanding of double embedded propositions, e.g. "Does John know that Mary know where the ice-cream man is?" In addition, it was felt that the original task required 'the child to keep track of a great deal of information and reason through long inferential chains' (Sullivan et al., 1994, pp397). The modified story (i.e. the 'Birthday Puppy') was shorter and with probe questions added to ensure that the child was actively processing the story and able to cope with its linguistic complexity.

Sullivan et al. (1994) compared the performance of preschoolers in the original (Ice Cream Man) and the adapted (Birthday Puppy) versions of the second-order false belief tests. Results showed that over 90% of children tested with the Birthday Party story passed the test at 5 ½ years old; far more than the proportion of children who passed the Ice Cream Man test at the same age. The test seems to be useful in discriminating children who may have second-order false belief attribution, but who may not have developed the linguistic competency and information processing capacity needed to pass the original Ice-Cream Man test.

- *Cow and Dalmatian tests*

These were designed to assess children's ability to demonstrate interpretive diversity, which is the term used by Carpendale and Chandler (1996) to refer to the fact that two people exposed to precisely the same stimulus may interpret it in quite different, but equally plausible ways. Carpendale and Chandler argued that an understanding of interpretive diversity is necessary to succeed in false belief tests. To test this hypothesis with a sample of children with autism, Luckett, Powell, Messer, Thornton and Schulz (2002) developed a battery of tasks, including the Cow and Dalmatian tests.

The Cow and Dalmatian tasks were designed to resemble, as closely as possible the cognitive demands in the 'Smarties' test, but differing only in that the stimuli

did not change throughout the task, but only the children's interpretation of it did change. In order to pass the test (thus demonstrating adequate interpretive diversity) the children must pass the first-order false belief question ("What will X say if I show him this picture, just like I showed you the first time?"), as well as questions that were added to control for memory ("When I first showed you the picture, what did you say was in the picture?"), and understanding of reality ("What's really in the picture?"). See Fig. 4.5.3.1.

Fig. 4.5.3.1: Stimulus for Interpretive Diversity Test (Cow)



Note: The child is first shown (a) and asked, 'What's in this picture?' S/he is then shown that it is picture of a Cow (b). Finally, s/he is asked to predict her/his friend's reaction if shown picture (a).

Luckett et al. (2002) found that overall; children with developmental delays and autism found the interpretive diversity tasks more difficult than the first-order false belief tasks. There was no correlation between children's scores on the interpretive diversity tasks and false belief tasks; and in addition, the sample included a sub-group of children who passed false belief tasks but failed interpretive diversity tasks. Luckett et al. (2002) suggested that this subgroup could be children with autism who were able to pass standard false belief task, only if these are presented in a 'contrived' setting (e.g. Sally-Ann test), but who have difficulties when the task was embedded in a naturalistic/social context (e.g. attributing false beliefs to their actual peers). Luckett et al. suggested that the poor performance of children with autism in the interpretive reality task may be reflecting their poor social understanding. However, no data was obtained regarding the children's level of social understanding, and hence there was no independent corroboration for this interpretation.

Other aspects of the task demands could also account for the poorer performance of the children with autism in the interpretive diversity tasks. The ‘figure-ground’ effect in the Cow/Dalmatian stimuli (see Fig. 4.5.3.1) were chosen because it has the characteristic of remaining unchanged throughout the task, but yet able to cause changes in the children’s interpretation of it (Luckett et al., 2002). The ‘figure-ground’ effect is very powerful, and once being shown that the stimulus (i.e. the ‘meaningless black spots’) is a picture of a cow, it is difficult for individuals to suspend or ignore the impact of the new meaning of the stimulus. The change is not merely in the individuals’ interpretation of the stimulus; a change in children’s visual perception had occurred, which may make it difficult for them to revert to the previous perception. It can be argued that the stimulus in the Cow/Dalmatian task did not totally ‘remain unchanged’, as intended by Luckett et al. (2002). Children who responded correctly to the Cow/Dalmatian tasks were not only able to attribute false beliefs to their peers; they had to be able to suspend/resist the effect of a competing visual perception, namely the perception of the stimulus as that of a Cow/Dalmatian. It could be argued that these additional demand characteristics could have introduced other sources of interference in the task.

- *Penny Hiding task*<sup>19</sup>

The Penny Hiding task (Baron-Cohen, 1992) was developed as a way to assess children’s ability to deceive. Deception entails manipulation of another person’s thoughts (i.e. making someone believe in something false) and as such, tasks involving deception can be seen as an extension of the ability for second-order perspective taking that occurs in everyday life. The Penny Hiding task, which is based on a commonly played children’s game, taps three indices of deception: 1) object occlusion, seen in terms of children’s ability to ensure that the other party (guesser) does not see the penny; 2) information occlusion, i.e. the child’s attempts to prevent the guesser from getting access to clues or information about the location of the penny; and 3) ‘advance deception’, i.e. attempts by the child to cheat during the test, for example by concealing the penny elsewhere and presenting two empty hands.

---

<sup>19</sup> For a description of the test, see Appendix M, Item 3: ‘Coin Hiding’ test.

Studies indicated that all subjects who show information occlusion show object occlusion, but the reverse did not apply; suggesting that information occlusion is a more developmentally advanced deception skill (Baron-Cohen, 1992). In the study, four children with autism who passed first-order false belief tests failed to show information occlusion. At the same time, no children passed the Penny Hiding game without passing first-order false belief tests, suggesting that first order false belief test is a necessary but insufficient condition for passing the deception task.

- *Strange Stories*

In most studies of theory of mind, there were consistently, a minority of children with autism that pass first, even second-order false belief tests. One hypothesis is that these children with autism pass the false belief test using an ‘alternative’ strategy (Frith et al., 1991 & 2003), such as using logical deduction/ inference. The strategy used by such children could be applied successfully in controlled (or contrived) testing situations where visual access and information are spelled out, as they were in the false belief tests. However, when applied to more realistic settings, these strategies may fail. The ‘Strange Stories’ test was developed to demonstrate this distinction between the theory of mind ability in standard false-belief tasks and context embedded false belief attribution (Happe, 1994; Fletcher, Happé, Frith, Baker, Dolan, Frackowiak, Frith, 1995). The Strange Stories used by White et al. (in prep) is based on the version used by Fletcher et al. (1995).

Children’s responses were coded for both accuracy and justification, and a higher score was given for children who provided justification that reflected awareness of mental states, compared with justification that were based on physical states. This scoring method is appropriate for the present thesis, as the aim is to identify test items that can reflect subtle variations (or gradations) in children’s theory of mind abilities (See Table 4.5.3.1.2).

**Table 4.5.3.1.2 : Scoring Criteria for Strange Stories**

Strange Story <sup>20</sup>	<p><i>Brian is always hungry. This morning he is going out for a picnic with his friends in school. His teacher has prepared special snacks for everyone. It is his favourite food: chicken burgers. He is a very greedy boy, and would like to have more chicken burgers than anyone else, even though his mother will have prepared him a nice lunch when he gets home! But everyone is allowed one burger and no more. When his teacher gives out the burger to him, he says, "Oh, please can I have three burgers, because I won't be having any lunch when I get home!"</i></p> <p><b>Accuracy Question: Is it true what Brain said? ( 0 or 1 point)</b>  <b>Justification Question: Why did Brian say that? ( 0 - 2 points)</b></p>
<b>Graded Justification</b>	
0 point response	Factually incorrect or irrelevant responses.
1 point response (Physical state attribution)	Reference to Brian's actual physical state, e.g. 'Because he's greedy or hungry; or the factual outcome of his actions, e.g. 'Because he will get more chicken burgers'.
2 point response (Mental state attribution)	Reference to the fact that Brain was trying to elicit sympathy or being deceptive, e.g. 'He lied because he was a greedy boy and wanted more food', and/or 'He wanted the teacher to feel sorry for him and give him more burgers'.

*Note: Total score is based on both Accuracy and Justification Questions*

#### **4.5.3.2 Summary of analyses of content validity**

Overall, the analyses suggest that the individual items in White et al.'s study could be considered for use in the theory of mind (ToM) test battery in the present thesis, provided that it meets the following criteria:

1. It has the appropriate level of difficulty, given the target sample (i.e. children with autism, age 8 to 12 years);
2. It contributes positively to the internal reliability of the test battery; and
3. Children's responses in the test items are relatively free from bias, i.e. characteristics extraneous to theory of mind abilities, such as age and general (non-verbal) abilities.

A small-scale study (described in Appendix L) was conducted to identify the items that met the above criteria. Based on data obtained from 27 children (8 to 12 years old), the study concluded that:

<sup>20</sup> The sample is one of the Strange Stories by Happe (1994) that has been adapted for use in the present thesis; see Appendix M, Item 6 : Strange Stories.

1. All items met adequate reliability criteria (alpha levels greater than 0.6). However, the reliability of the overall scale (i.e. the Aggregate ToM Score) was improved when the interpretive diversity items, namely the Cow/Dalmatian tests were removed. This suggests that these items did not contribute positively to the reliability of the overall scale and should be removed.
2. Except for the Strange Stories, none of the other items showed a ceiling or floor effect, indicating suitable levels of item difficulty. The Strange Stories were too difficult for some of the children, and the items needed be replaced with a version that is more suited for young children, such as those used by Happe (1995).
3. The children were able to respond adequately to the task demands of the items in the battery. Positive correlations were found between children's verbal abilities and their performance on the second-order and advanced false-belief subtasks. It was argued that this might reflect a characteristic that is unique to the sample of children in the study, i.e. children with autism. It was noted that, consistent with the findings from other studies, children with autism require a higher level of language skill to pass advance theory of mind tests (Happe, 1995; Fisher et al., 2005).

Based on the above findings, six items were identified and selected for use in the present thesis (See Table 4.5.3.2). Additional modifications were needed to make the content more culturally appropriate for children from Singapore, and these modifications and adaptations will be discussed as part of the methodology for Study 2 (See Chapter 5).

**Table 4.5.3.2 Structure of ToM battery Used in the Present Thesis**

Test Items (adapted)	ToM Construct	Reference	No. of Control Qs*	No. of test Qs.	Max score
Sally-Ann	1 <sup>st</sup> order false belief	Baron-Cohen et al., 1985	1	1	1
Smarties	1 <sup>st</sup> order false belief	Perner et al., 1987	1	1	1
Ice-cream man	2 <sup>nd</sup> order false belief	Baron-Cohen, 1989	2	2	2
Birthday puppy	2 <sup>nd</sup> order false belief	Sullivan et al., 1994	2	2	2
Penny Hiding	Deception	Baron-Cohen, 1992	-	6 (trials**)	6
Strange stories	Advance false-belief	Happe (1995)	-	24***	72
Composite	Aggregate ToM	NA (present thesis)	30	36	84

Note

\*Control Qs: These questions check children's understanding and/or memory of the tasks/stories. Children must pass the control questions in order to receive a full score for correct their responses in the test questions.

\*\*In the Penny Hiding test, the children's scores are based on the number of times (out of 6 trials) that they succeed in hiding the Penny without errors (e.g. failing to fully conceal the coin).  
\*\*\*In the Strange Stories test, responses to each 'Justification' question can be given a maximum of 2 points, depending on the type of justification given for the answers; see Table 4.5.3.1.2.

## 4.6 SUMMARY AND CONCLUSION FOR CHAPTER 4

Based on the review and series of field-tests, the following measures were selected for each of the indicators of SEN:

Table 4.6 Selected Measures for the Indicators of SEN

<b>Indicators of SEN</b>	<b>Selected measures (and authors)</b>
1. Intelligence	<ul style="list-style-type: none"><li>• Wechsler Intelligence Scales for Children (3<sup>rd</sup> Edition); Singapore Version, or WISC-III (Singapore)</li><li>• Ministry of Education (Singapore)</li></ul>
2. Executive function	<ul style="list-style-type: none"><li>• Behavioural Assessment for Dysexecutive Syndrome for Children, or BADS-C</li><li>• Emslie et al., 2003</li></ul>
3. Central Coherence	<ul style="list-style-type: none"><li>• Children's Embedded Figures Test, or CEFT</li><li>• Witkins et al., 1971</li></ul>
4. Cognitive Modifiability	<ul style="list-style-type: none"><li>• Analogies Subtest in the Cognitive Modifiability battery, or CMB</li><li>• Tzuriel, 1995</li></ul>
5. Theory of Mind	<ul style="list-style-type: none"><li>• Theory of Mind Battery, or ToM Battery</li><li>• Selected and adapted for use in the present thesis.</li></ul>

The five measures were selected based on evidence of reliability and validity, i.e. the test showed adequate reliability and validity as a measure for the respective intended constructs. However, none of the indicators have been directly evaluated in terms of their validity as indicators of special educational needs for children with autism. As discussed in Chapter 2, the strength of validity arguments for psychological assessment lie in the extent to which evidence of validity supports the interpretations of a test score for a given purpose and context of use. The two studies, which are reported in Chapters 5 and 6, aim to evaluate the extent to which information obtained from these five measures can be interpreted as reflecting the special educational needs of children with autism in Singapore.

## CHAPTER 5

### STUDY 2 : CRITERION VALIDITY OF THE INDICATORS OF SPECIAL EDUCATIONAL NEEDS FOR CHILDREN WITH AUTISM.

#### 5.1 INTRODUCTION

Indicators of special educational needs (SEN) can be defined as measures of specific domains of ability, which can be used as a predictor of future functioning or needs (as discussed in Section 2.3.8). For children with autism, five possible SEN indicators were identified, namely measures of intelligence, theory of mind, executive function, central coherence and cognitive modifiability. In the previous chapter, specific tests for these five indicators were identified. Based on the Standards for Psychological and Educational Testing (APA, AERA & NCME, 1999) or *Standards*, two aspects of validity are identified as critical in the evaluation of these indicators, namely criterion and treatment validity. The focus of the present study is criterion validity, i.e. the extent that children's performance on these indicators reflect their level of special educational needs.

#### 5.2 CRITERION VALIDITY

Criterion validity involves a comparison of an individual's performance in one test/indicator with another independent activity. The selection of a criterion measure is determined by the purpose of assessment<sup>21</sup> (Anastasi & Urbina, 1996); and in the Singapore context the immediate outcomes of the SEN assessment processes are: 1) the identification of children's level of special educational needs; and 2) decisions regarding their ability to cope with the demands of mainstream school. Based on these outcomes, the evaluation of criterion validity for the selected indicators is based on the following criteria:

- 1) The extent to which the indicators show good concurrence with children's level of special educational needs, based on an independent measure of SEN; and

---

<sup>21</sup> The issues relating to the evaluation of criterion validity is discussed in Section 2.2.3.5

2) The extent to which the indicators are accurate in distinguishing children with autism who can cope with mainstream schools, from those that need special schools.

The purpose of the present study is to evaluate the identified indicators of SEN against these two criteria.

### 5.3 PARTICIPANTS

In total, 52 participants aged 8 to 12 years, agreed to participate in the study: 13 from mainstream schools; 17 from special schools for children with mild learning difficulties; and 22 from special schools for children with moderate learning difficulties. Only children whose school placements were considered stable, i.e. they were not being considered for transfer to or from special schools, were included in the study. Table 5.3 shows the demographic profile for the sample.

Table 5.3 : Profile of Participants

Sex and Age	Mainstream	Special Sch (Mild)	Special Sch (Moderate)	N
Number	13	17	22	52
Boys	7	16	14	37
Girls	6	1	8	15
Mean age	10:00 yrs	9:09 yrs	10:00 yrs	9:10 yrs
Socio-economic status	Mainstream	Special Sch (Mild)	Special Sch (Moderate)	N
High	6	4	3	13
Middle	7	8	16	31
Low	0	5	3	8
Home languages	Mainstream	Special Sch (Mild)	Special Sch (Moderate)	N
English	13	17	22	52
Mandarin	13	14	19	46
Malay	0	2	1	3
Tamil	0	1	2	3
GARS Autism Quotient	Mainstream	Special Sch (Mild)	Special Sch (Moderate)	N
AQ $\geq$ 80	6	10	12	28
AQ < 80	7	7	10	24

As shown in Table 5.3, the children in the three groups are comparable in terms of mean age, sex ratio and language background. Overall, the total sample shows an over-representation of boys. This is consistent with the findings of other studies indicating a higher incidence of autism among males (Lord, 1982).

All the participants had a pre-existing diagnosis of autism or Autism Spectrum Disorder (ASD). The diagnoses were given by clinical/educational psychologists from health or education ministries, or psychologists in private practice. There may be variability in the diagnostic criteria used in the pre-existing diagnosis. To gauge this variability, the Gilliam Autism Rating Scale, or GARS<sup>22</sup> (Gilliam, 1995) checklist, was given via post to the parents of all participants. The parents completed the GARS by reporting the frequency of behaviours associated with autism. The standard score derived from the GARS, i.e. the Autism Quotient, represents the extent to which the child shows symptoms of autism. An Autism Quotient of  $\geq 80$  is regarded as showing significant characteristics corresponding to the DSM-IV (APA, 1995) diagnostic criteria of autism.

As shown in Table 5.3, there were proportionally more children in special schools who showed at least an average probability of meeting the DSM-IV criteria for autism, i.e.  $AQ \geq 80$ . Higher scores in the GARS checklist also correspond to higher incidence/frequency of autism characteristics. This was to be expected, as children who are able to cope in mainstream schools would show lower severity in terms of autism behaviours. At the same time, it could be noted that a sizable number (24) fell below the GARS cut-off criteria for autism characteristics. This should not be taken as indicating that these children are not autistic, but rather, as suggesting the increased use of the broader dimensional criteria for the diagnosis of autism spectrum disorder (Blaxill, 2002)<sup>23</sup>.

## 5.4 MEASURES

### 5.4.1 Criterion measure for SEN level

Based on the results of study 1 (reported in Chapter 3), the International Classification for Functioning, Disability and Health, or ICF (WHO, 2001) was found to have adequate reliability and validity as an independent measure of the special educational needs of children with autism in Singapore. The ICF checklist was completed based on parental reports of children's behaviour and functioning levels, obtained through investigator-based interviews. The child's

---

<sup>22</sup> A description and review of the GARS checklist is in Section 2.5.2.

<sup>23</sup> The issues and implications of using different diagnostic criteria for autism were discussed earlier in Section 2.4.1.

functioning level was rated along a 5-point scale ranging from '0', i.e. no impairment or activity limitation, to '4', i.e. severe impairment or activity limitation. The ICF SEN composite score was used as a measure of children's special educational needs and comprised the following component scores:

- Total ratings for impairments in body functions;
- Total ratings for limitation in activity and participation; and
- Total ratings for environmental barriers.

Higher ICF SEN ratings indicated greater severity of special educational needs. The aim of the evaluation for the first criterion is to identify the indicators that best predict children's ICF SEN ratings. One factor that would affect the patterns of result is the extent to which the abilities measured by the indicators of SEN overlap with the domains of functioning captured in the ICF checklist.

#### *5.4.1.1 Concurrence between the ICF domains and SEN indicators*

Table 5.4.1.1 presents an analysis of the concurrence between domains of functioning in the ICF, and the five identified indicators of SEN, namely measures of intelligence (IQ), executive function (EF), theory of Mind (ToM), central coherence (CC) and cognitive modifiability (CM). As all the indicators relate to children's cognitive functioning, there is a direct link between these indicators and the ICF domain for Mental functions. However, in terms of the domains relating to children's level of Activity and Participation, the patterns of overlap differs:

##### ICF domains and IQ

- Based on the review of studies on measures of intelligence (discussed in Section 4.1), the ICF domains which have been shown be linked to IQ are: Learning and Applying Knowledge; Communication; Self-care; and Education. Given that children's verbal skills may affect their ability to relate to others, there is also an inferred link between IQ and ICF domain for Interpersonal activities.

### ICF domains and executive function

- The ICF domain that is most clearly related to executive function skills is Learning and Applying Knowledge<sup>24</sup>. In addition because language is implied in the application of executive function processes, there is an assumed link with the ICF domain for communication. Also by inference, children with weak executive function skills are expected to have greater impairments in applying their skills in every day contexts such as Self-care, Education and Community/Social/Civic activities (see Table 5.4.1.1).

Table 5.4.1.1 Concurrence between ICF Domains and SEN Indicators

<i><b>ICF Domains</b></i>	<i><b>Indicators of SEN</b></i>				
	<i><b>IQ</b></i>	<i><b>EF</b></i>	<i><b>ToM</b></i>	<i><b>CC</b></i>	<i><b>CM</b></i>
<b><i>Body Functions/Structures</i></b>					
Mental					
Sensory					
Voice & Speech					
Cardiovascular, Heamatological etc					
Digestive, Metabolic etc					
Gentinourinary and reproductive					
Neuromusculoskeletal and movement					
Skin					
<b><i>Activities &amp; Participation</i></b>					
Learning & Applying Knowledge					
Communication					
Moving around					
Self care					
Interpersonal					
Mobility					
Social relationships					
Education					
Community, Social and Civic life					
<b>Key</b>					
	Direct relationship (based on published research findings)				
	Inferred relationship (based on theory)				
	No established relationship				
IQ : Intelligence					
EF : Executive Function					
ToM : Theory of Mind					
CC : Central Coherence					
CM : Cognitive Modifiability					

### ICF domains and theory of mind

- Based on studies on the theory of mind of children with autism (discussed in Sections 2.4 and 4.5) children's ratings in ICF domains for Communication

<sup>24</sup> The review of research on executive function is in Section 4.2.

and Interpersonal activities would be strongly related to measures of ToM. By inference, children with poor theory of mind would be expected to show greater limitations in activities related to Learning and Application of Knowledge, Social Relationships, Education and Community/Social/Civic life.

#### ICF domains and central coherence

- Although few studies have directly evaluated the relationship between central coherence and children's adaptation in daily life activities, based on the theory, a diminished central coherence should be reflected in children's ratings for the ICF domains for Learning and Applying Knowledge, Communication, and participation in activities related to Education and Community/Social/Civic life.

#### ICF domains and cognitive modifiability

- Based on the literature (discussed in Section 4.4), there is a direct link between ICF domains for Learning and Applying Knowledge and cognitive modifiability. By inference, children who are less modifiable (i.e. show less gains after teaching) would be expected to experience greater limitations in the ICF domains for activities involving Communication, Education and Community/social/civic life.

Overall, based on the analysis, there is ample concurrence between ICF domains and all the five indicators of SEN; and this provides some basis for predicting a significant relationships between the variables of interest. Compared with the other SEN indicators, central coherence appears to have the least direct relationship with the content of the ICF domains (See Table 5.4.1.1).

#### **5.4.2 Criterion for the ability to cope with mainstream school**

As discussed in Chapter 1, children with autism in Singapore remain in mainstream schools only if their needs are not severe and they are perceived to be able to cope with the demands of the mainstream schools. In this context, a reflection of the children's ability to cope with mainstream school is their stable

placement in mainstream schools, i.e. with no plans or considerations for transfer to special schools. The aim of the evaluation for the second criterion is to identify the indicators that best classify children according to their school, i.e. mainstream, special school for mild difficulties, or special school for moderate difficulties.

One key advantage of this criterion is that it is based on pre-existing, independent information about the child. Unlike the first criteria, i.e. the ICF SEN score which was obtained by the present researcher, decisions pertaining to the children's suitability for mainstream or special schools were based on decisions made by independent professionals, such as educational psychologists and therapists who had worked directly with the child, in consultation with the child's parents and teachers. Hence there is some form of independent corroboration (albeit by inference) that the needs of the children in mainstream were perceived to have been met without transfer to special schools.

However, one possible drawback with this criterion is that the decision for a child to be transferred to a special school is often based on multiple factors, and not purely an objective evaluation of the child's special educational needs. For example:

1. Teachers' and parental perceptions and expectations of the child, as well as peer acceptance have been shown to be important precursors to successful integration of children with SEN in mainstream primary schools (Farrell, 1997 & 2004; Frederickson & Woolfson, 1987; Frederickson, Dunsmuir, Lang & Monsen, 2004).
2. From the teachers' perspective, behavioural concerns appear to be the most important dimension influencing their receptivity to children with SEN in mainstream schools (Ward et al., 1987; Zetlin, 1987).

In the Singapore context, given that there is no policy for compulsory inclusion, the decision for the transfer of a child from mainstream to special schools are often initiated by the school/teachers, in consultation with parents. The children's behaviour and functioning that affects teacher's/schools' perceptions of their ability to cope with the demands of mainstream school would have a

major role in determining their school placements. For example, a child with autism with severe needs but who is ‘passive’ (e.g. showing signs of selective mutism), may be less likely to be considered for transfer to special schools than another child with autism who may have less severe learning needs but displays behaviour which are perceived as ‘active but odd’<sup>25</sup>.

### **5.4.3 Predictor variables (Indicators of SEN)**

The selection of the specific tests and measurement issues for each of the five indicators of SEN were discussed in Chapter 4. Based on the review and a series of field-tests, the five measures shown in Table 5.4.3 were selected for each of the respective indicators: intelligence; executive function; central coherence; cognitive modifiability; and theory of mind. The following subsections outlines adjustments to the measures considered in relation to age and cultural differences.

#### *5.4.3.1 Age adjustments*

Two of the outcome measures were based on standardised normative scores: Full Scale IQ (FSIQ) based on the WISC-III (Singapore); and total scaled scores for the BADS-C. However, in three of the tests children’s raw scores are used, namely theory of mind (ToM Battery), central coherence (CEFT) and cognitive modifiability (CMB). Preliminary checks for possible developmental/age effects would need to be made, by investigating the correlation between the children’s score in these three measures and their chronological age. If age-effects were found, adjustments to the scores would need to be made.

---

<sup>25</sup> Wing & Gould (1979) described three categories of children with autism based on their social characteristics: 1. ‘aloof’ (indifferent to social approaches, although physical contact might be enjoyed), 2. ‘passive’ (accepting social approaches and following other’s lead but not initiating any contact), and 3. ‘active but odd’ (making active social approaches but in an inappropriate manner).

**Table 5.4.3 Selected Measures for the Indicators of SEN**

<b>Indicators of SEN</b>	<b>Selected Tests</b>	<b>Outcome Measure</b>
1. Intelligence	Wechsler Intelligence Scales for Children (3 <sup>rd</sup> Edition); Singapore Version, or <b>WISC-III (Singapore)</b>	Full-Scale IQ (age-norms), obtained from 8 subtests: Picture Completion; Information; Coding; Similarities; Picture Arrangement; Arithmetic; Block Design; and Vocabulary.
2. Executive function	Behavioural Assessment for Dysexecutive Syndrome for Children, or <b>BADS-C</b> .	Total scale score (age and IQ based norms); obtained from all six subtests in the <b>BADS-C</b> .
3. Central Coherence	Children's Embedded Figures Test, or <b>CEFT</b> .	Total raw score*
4. Cognitive Modifiability	Analogy Subtest in the Cognitive Modifiability battery, or <b>CMB</b> .	Adjusted Post-teaching score (Post-teaching score, after removal of Pre-teaching effects)
5. Theory of Mind	Theory of Mind battery, or <b>ToM Battery</b> which was selected and adapted for use in the present thesis.	Total raw score*

\* adjusted for age (if needed)

#### *5.4.3.2 Cultural adjustments*

The WISC-III (Singapore) was developed for use with Singaporean children; hence no further adaptations were needed. For the tests of executive function (BADS-C), central coherence (CEFT) and cognitive modifiability (CMB), no adaptations were felt necessary as these did not involve content that would be unfamiliar to children from Singapore.

Items selected for the measure of theory of mind (ToM Battery) were redesigned and adapted to make the stories more culturally appropriate for Singapore children, such as:

1. The names of some characters were changed to names which were more familiar to local Singapore children, e.g. 'Mrs Peabody' became 'Mrs Tan'; 'Sally-Ann' dolls were replaced with the 'Jim' and 'Rosie' characters;

2. ‘Halloween party’ was changed to ‘costume party’, as the former is not a holiday celebrated in Singapore;
3. A character’s birthday gift was changed from a puppy to a skateboard, which was hidden by his mum in the storeroom of his flat (instead of the basement of the house, as in the original version). These changes were felt necessary as a large majority of Singaporean children live in high-rise apartments where there are no basements. It is also not common practice to give dogs or cats as birthday presents.
4. The 3-dimensional scenes used in the Ice-Cream Man and Birthday stories were designed to reflect a typical housing estate in Singapore.

(See Appendix M for a full description of the materials used in the ToM Battery).

## **5.5. PROCEDURE<sup>26</sup>**

Invitations to the participants were sent out to the parents of children aged 8 to 12 years old with a pre-existing diagnosis of ASD. Additional checks were made with the schools to ensure that at the time of the study, the children’s educational placements were deemed appropriate by parents and professionals who were working with the child; none of the mainstream children in the study were being considered for transfer to or from special schools.

From the initial 20 children from Study 1 who met the criteria, 11 agreed to participate in Study 2. An additional 50 ‘open invitation letters’ were sent out through schools and the Autism Resource Center<sup>27</sup>; and 42 parents responded (84%). Upon receiving informed parental consent, arrangements were made to conduct the individual testing of the children and the parent interviews (See Appendix N for a sample of the open invitation letter and informed consent forms).

### **5.5.1 Individual testing**

The assessments were carried out in the children’s school by the researcher, who is a qualified educational psychologist and experienced in conducting

---

<sup>26</sup> Procedure for the study was approved by UCL ethics committee for non-NHS human research (Project ID number 0038/002)

<sup>27</sup> The Autism Resource Centre is a non-profit organisation that provides support and therapy services for children with autism in Singapore.

standardised assessments with children with autism. The researcher has also undergone the specialised training required for users of the CMB test. Prior to Study 2, the researcher familiarised herself with all measures used in the study by using the tests with a small sample of children with autism. All the testing was carried out using standard Singapore English. The order of the tests<sup>28</sup> was counter-balanced and the testing for each child was spread over 2 to 3 days. Breaks ranging from 15 to 45 minutes were given during each testing session to minimise fatigue effects. Table 5.5.1 shows the sample format and average testing time for each child. It should be noted that the total testing time does not include the time used for establishing rapport and break/rest periods.

**Table 5.5.1 : Sample Testing Format and Average Testing Time per Child**

Day/test sessions	Tests	Testing Time (mins)	No. of breaks/rests (15 to 45 mins each)
Day 1	WISC-III	60	2
	CEFT	15	
Day 2	BADS-C	30	2
	Theory of Mind	60	
	CMB – pre teaching	15	
Day 3	CMB - teaching	60	1
	CMB – post teaching	15	
<b>Average total testing time</b>		<b>255min (4hrs 15 mins)</b>	5

### **5.5.2 Parent interview**

Face to face parent interviews were conducted using the ICF checklist, with each interview lasting for about two hours. The parent interview was conducted after the individual testing for the child concerned was completed. This was to ensure that detailed information about children's functioning (from the interview) was not available to the tester prior to the individual testing session.

### **5.5.3 Feedback**

After completion of the individual testing and parent interview, a summary of individual children's responses was given to parents who had requested feedback. This procedure was included in accordance with UCL's ethical guidelines on research involving minors and vulnerable individuals. The option for parents and caregivers to request feedback on children's responses in the

<sup>28</sup> The order for Pre, Teaching and Post phases of the CMB were conducted in the required sequence.

study was offered as part of the process of obtaining informed consent (See Appendix N).

## 5.6 RESULTS

Multivariate methods were used to investigate the concurrent validity of the indicators in terms of the two criteria:

- 1) The extent to which the indicators show good concurrence with children's levels of special educational needs, based on an independent measure of SEN, i.e. ICF SEN scores; and
- 2) The extent to which the indicators are accurate in distinguishing children with autism who can cope with mainstream schools, from those that need special schools.

### 5.6.1 Preliminary analyses

Table 5.6.1.1 shows the descriptive statistics for the criterion variable, i.e. children's ICF SEN score, and the five predictor variables. Prior to the analysis, all variables were examined for accuracy in data entry, missing values, and fit between their distributions and the assumptions of multivariate analysis. No missing values were detected and the distribution (i.e. z score) of the variables did not exceed the '3.29 value for extreme skewness and kurtosis' (Tabacknick & Fidell, 2001, pp. 90).

Table 5.6.1.1 Descriptive Statistics

Criterion variable	Mean	SD	Std error	Min	Max	z-score (skewness)	z-score (kurtosis)
ICF SEN <sup>a</sup>	137.85	85.40	11.84	6	377	3.13	2.62
<b>Predictor variables</b>							
FSIQ <sup>b</sup>	80.23	24.84	3.44	40	132	1.35	-1.03
ToM <sup>c</sup>	21.94	19.30	2.68	0	70	2.53	-0.71
BADS-C	30.02	11.04	1.53	6	51	0.64	-0.89
CEFT <sup>e</sup>	8.98	5.50	0.76	1	25	2.78	0.72
CMB <sup>f</sup>	65.73	15.20	2.11	29.60	98.61	-1.53	-0.58

<sup>a</sup>Total ICF ratings for Special Educational Needs (SEN), collated from ratings for impairments in body function, limitations in activity & participation and environmental barriers.

<sup>b</sup>Full-Scale Intelligent Quotient obtained from WISC-III (Singapore).

<sup>c</sup>Aggregate Theory of Mind score obtained from the ToM Battery

<sup>d</sup>Total scaled score obtained from the Behavioural Assessment of Dysexecutive Syndrome for Children.

<sup>e</sup>Total score for the Children Embedded Figures Test.

<sup>f</sup>Total Post-teaching score for Cognitive Modifiability Battery (Analogies subtest), adjusted for Pre-teaching scores.

A correlation analysis was conducted to estimate the possible effects of age on the predictor variables which were based on raw scores, namely measures of theory of mind (ToM), central coherence (CEFT), and cognitive modifiability (CMB). As shown in Table 5.6.1.2, only the indicator for central coherence (CEFT) showed a significant age effect ( $r = 0.47, p \leq 0.01$ ). A regression-based adjustment was carried out, and it was the age-adjusted CEFT scores that were used in subsequent analyses.

**Table 5.6.1.2 : Correlation Matrix (Pearson's  $r$ ) Between Age and Un-Normed Predictor Variables**

Children's Age <sup>a</sup>	ToM <sup>c</sup>	CEFT <sup>e</sup>	CMB <sup>f</sup>
	0.08	0.47**	0.18

\*\* $p \leq 0.01$

**Key**

<sup>a</sup> Children's age in months.

<sup>c</sup> Aggregate Theory of Mind score obtained from the ToM Battery.

<sup>e</sup> Total score for Children's Embedded Figures Test.

<sup>f</sup> Total Post-teaching score for Cognitive Modifiability Battery (Analogies subtest), adjusted for Pre-teaching scores.

### **5.6.2 First criterion: Concurrence with an independent measure of SEN, i.e. ICF**

To evaluate the relative concurrence of each predictor variable (i.e. indicators of SEN) with the independent measure of SEN (ICF SEN), a multiple regression analysis was chosen. This would enable the evaluation to go beyond the effects of individual predictors and investigate the combinations of indicators that would best reflect children's level of special educational needs. Two methods of regression analyses were considered, namely stepwise and hierarchical regression.

In stepwise regression, the order of entry of the predictor variables is based solely on statistical criteria. Prior to considering its use, following the methods recommended by Tabachnick and Fidell (2001), the data was examined to address two fundamental questions: 1) what is the size of the relationship between the criterion variable (ICF) and each of the predictors; and 2) how much

inter-correlations are there between the predictor variables, i.e. are there any multi-collinearity effects?

Table 5.6.2.1 : Correlation Matrix (Pearson's  $r$ ) for ICF SEN Ratings and All Predictor Variables (i.e. Indicators of SEN)

	ICF SEN	FSIQ	ToM	BADS-C	CEFT	CMB
ICF SEN <sup>a</sup>	1.0	-0.78**	-0.74**	-0.77**	-0.01	-0.55**
FSIQ <sup>b</sup>	-	1.0	0.88**	0.83**	-0.6	0.68**
Agg. ToM <sup>c</sup>	-	-	1.0	0.83**	0.07	0.58**
BADS-C <sup>d</sup>	-	-	-	1.0	0.01	0.62**
CEFT <sup>e</sup> adjusted	-	-	-	-	1.0	0.19
CMB <sup>f</sup>	-	-	-	-	-	1.0

<sup>a</sup> ICF SEN ratings, reflecting level of special educational needs.

\*\* $p<0.01$

<sup>b</sup> Full-Scale Intelligent Quotient obtained from WISC-III (Spore).

<sup>c</sup> Aggregate Theory of Mind score obtained from the ToM Battery.

<sup>d</sup> Total scale score obtained from the Behavioural Assessment of Dysexecutive Syndrome for Children.

<sup>e</sup> Total score for Children Embedded Figures test, adjusted for age.

<sup>f</sup> Total Post-teaching score for Cognitive Modifiability Battery (Analogies subtest), adjusted for Pre-teaching scores.

From Table 5.6.2.1, it can be seen that except for the scores on the Children's Embedded Figures Test (CEFT adjusted), all other predictors showed a strong negative correlation with children's ICF SEN ratings. Children's level of special educational needs (ICF SEN) showed the strongest negative correlation with the Full Scale IQ index ( $r=-0.78$ ), followed by the score for Behavioural Assessment of Dysexecutive Syndrome ( $r=-0.77$ ), and the score for Aggregate Theory of Mind ( $r=-0.74$ ). A relatively weaker, but nonetheless significant, negative correlation was found between children's ICF SEN ratings and their score on the Cognitive Modifiability Battery ( $r=-0.55$ ). The results indicate that the predictors that could be used in subsequent regression analysis were as follows: 1) Full Scale IQ (FSIQ) ; 2) Aggregate Theory of Mind (ToM), 3) Behavioural Assessment of Dysexecutive Syndrome (BADS-C); and 4) Cognitive Modifiability Battery (CMB) scores.

However, there were also strong correlations between individual predictors (see Table 5.6.2.1). Full Scale IQ correlated very strongly with children's scores for Aggregate Theory of Mind ( $r=0.88$ ) and Behavioural Assessment for Dysexecutive Syndrome ( $r=0.83$ ), and correlated moderately with children's score on the Cognitive Modifiability Battery ( $r=0.68$ ). In addition, children's

score for Aggregate Theory of Mind and Behavioural Dysexecutive Syndrome showed a strong positive correlation ( $r=0.83$ ).

The strong inter-correlation between the predictors may pose a problem for stepwise regression analyses. High levels of collinearity (i.e. bivariate correlations close to 0.9) between predictors mean that statistically, the effects of the two highly correlated variables are interchangeable. This will increase the probability that a good (i.e. meaningful) predictor of the model might be found to be non-significant and rejected, i.e. a type II error (Tabachnick & Fidell, 2001). Based on these preliminary considerations, it was decided that a stepwise regression analysis was not suitable, and a hierarchical regression method was chosen.

In a hierarchical regression the order in which the predictors are entered into the regression model is based on theory or information from other studies which reflect the relative importance of the predictors. This method overcomes the problem associated with stepwise regression, where the order of variables is determined solely on statistical criteria. However, the frequently cited challenge in the use of hierarchical regression methods is to obtain theoretical or research evidence that can meaningfully guide the ordering of variables in terms of their relative importance (Tabachnick & Fidell, 2001; Field, 2000). In the present thesis the findings from Study 3 (described in Chapter 6), which investigated the treatment validity of these same variables as indicators of SEN, can be used to inform the relative importance of the predictors. Given that another important purpose of SEN assessment is to use the information derived from the various tests to plan interventions, it can be argued that one critical aspect of the relative importance of the predictors is the extent to which they provide information that can be used for intervention, i.e. their treatment validity.

The results of the use of hierarchical regression analyses for the first criterion in the present study is reported as a follow-up analysis in Chapter 7. The rest of the sections in the present chapter focus on results relating to the evaluation of the second criterion, namely the extent to which the indicators are accurate in

distinguishing children with autism who can cope with mainstream schools, from those that need special schools.

### 5.6.3 Second criterion: Accuracy in distinguishing children who can cope with mainstream from those that need special schools.

Stepwise discriminant function analyses were carried out to identify the predictors that could best classify the children according to their type of school, i.e. mainstream school ('Mainstream'), special school for mild learning difficulties ('Special Mild'), and special school for moderate learning difficulties ('Special Moderate'). Discriminant function analysis is useful for classifying individual cases in a dataset into the values of a categorical, dichotomous criterion groups. If the discriminant function analysis is effective for a dataset, the classification of the cases will yield a high percentage accuracy (Tabachnick & Fidell, 2001).

Preliminary analyses indicated children's mean scores for all the variables, except central coherence, were significantly different for the three groups (See Table 5.6.3).

Table 5.6.3 : Means (and S.D) Scores for All Predictor Variables (by School-types)

	Mainstream	Special Mild	Special Moderate	F value
<sup>a</sup> FSIQ	109.70 (19.3)	86.41 (10.9)	58.05 (10.4)	64.11**
<sup>b</sup> ToM	47.92 (12.2)	20.82 (13.2)	7.45 (7.0)	58.82**
<sup>c</sup> BADS-C	41.08 (7.2)	33.47 (8.0)	20.81 (6.6)	34.94**
<sup>d</sup> CMB	75.30 (9.9)	74.54 (8.6)	53.26 (12.9)	24.87**
<sup>e</sup> CEFT	8.83 (2.7)	8.81 (2.3)	9.18 (2.8)	0.12

\*\* $p \leq 0.01$

#### Key

<sup>a</sup> Full-Scale Intelligence Quotient obtained from WISC-III (Spore).

<sup>b</sup> Aggregate Theory of Mind score obtained from the theory of mind battery.

<sup>c</sup> Standardised total scale score obtained from the Behavioural Assessment of Dysexecutive Syndrome for Children.

<sup>d</sup> Total Post-teaching score for Cognitive Modifiability Battery (Analogies subtest), adjusted for Pre-teaching scores.

<sup>e</sup> Children's Embedded Figures Test scores, adjusted for age differences.

In discriminant function analyses, the examination of group differences goes beyond the effects of individual variables; the five predictor variables were evaluated in terms of the significant change in Rao's V observed when they were added to the model. Rao V is an index of the extent to which the discriminant function discriminates (or distinguishes) between the three criterion groups, and is used to determine if adding a predictor variable to the model will significantly improve the classification of individual children into the three different school-types.

A series of stepwise discriminant function analyses was carried out in two stages, and the results are shown in Tables 5.6.3.1 to 5.6.3.3. In the first stage, the predictors were evaluated for the ability to distinguish the children across the 3 criterion groups, and in the second stage, further comparisons were made in turn between two criterion groups.

#### *5.6.3.1 Mainstream VS Special (Mild) VS Special (Moderate) Schools*

Results of the first analysis are presented in Table 5.6.3.1.1. Out of the five predictor variables, the combination of indicators that produced significant changes to the values of Rao V were Full Scale IQ (FSIQ), Theory of Mind (ToM) and the Cognitive Modifiability Battery (CMB) scores. Based on the children's scores on these three predictor variables, 43 children (82.7%) were correctly classified according to their school type (See Table 5.6.3.1.2).

**Table 5.6.3.1.1 : Stepwise Discriminate Function Analyses of Predictor Variable on Children's Group by School Type**

Predictor variable	Rao's V	Change in Rao's V	Groups Means		
			Mainstream	Special Mild	Special Moderate
Full Scale IQ <sup>a</sup>	128.23	128.23**	109.69	86.41	58.05
Theory of Mind <sup>b</sup>	158.02	29.79**	47.92	20.82	7.45
CMB <sup>c</sup>	175.04	17.02**	75.30	74.54	53.26
CEFT <sup>e</sup>	182.11	1.28	8.83	8.81	9.18
BADS-C <sup>d</sup>	183.39	8.35	41.08	33.47	20.81

\*\*p<0.01

**Key**

<sup>a</sup> Full-Scale Intelligence Quotient obtained from WISC-III (Spore).

<sup>b</sup> Aggregate Theory of Mind score obtained from the theory of mind battery.

<sup>c</sup> Total Post-teaching score for Cognitive Modifiability Battery (Analogies subtest), adjusted for Pre-teaching scores.

<sup>d</sup> Standardised total scale score obtained from the Behavioural Assessment of Dysexecutive Syndrome for Children.

<sup>e</sup> Children's Embedded Figures Test scores, adjusted for age differences.

Table 5.6.3.1.2 : No. of Children Classified According to Actual Group (i.e. School Type) and Predicted Group (Based on FSIQ, ToM and CMB Scores)

Actual Group (School Type)	Predicted Group			Total
	Mainstream	Special Mild	Special Moderate	
Mainstream	10 (77%)	2 (15%)	1 (8%)	13
Special Mild	2 (12%)	14 (82%)	1 (6%)	17
Special Moderate	0 (0%)	3 (14%)	19 (86%)	22

### 5.6.3.2 Mainstream VS Special (Mild) Schools

A second discriminant function analysis was conducted, contrasting the children from the 'Mainstream' group with children from the 'Special Mild' group. In the analysis, Theory of Mind was found to be the predictor that made significant contributions to the separation between children from 'Mainstream' and 'Special Mild' groups (See Table 5.6.3.2.1). The accuracy rate of classifications based on ToM scores for these two groups was 83.3% (See Table 5.6.3.2.2).

Table 5.6.3.2.1. : Group Contrasts Between Mainstream and Special Schools (Mild)

Predictor variables	Rao's V	Change in Rao's V	Groups Means	
			Mainstream	Special (Mild)
Theory of Mind	33.09	33.09**	47.92	20.82
Full Scale IQ	33.68	0.59	109.70	86.41
CMB	34.59	0.91	75.30	74.54
BADS-C	34.71	0.12	41.08	33.47
CEFT	34.77	0.06	8.83	8.81

\*\* $p \leq 0.01$

Table 5.6.3.2.2 No. of Children Classified According to Actual Group (i.e. School Type) and Predicted Group (Based on ToM Scores)

<b>Actual Group (School Type)</b>	<b>Predicted Group</b>		<b>Total</b>
	<i>Mainstream</i>	<i>Special Mild</i>	
Mainstream	10 (76.9%)	3(23.1%)	13
Special Mild	2(11.8%)	15 (88.2%)	17

### *5.6.3.3 Special (Mild) VS Special (Moderate) Schools*

A third discriminant function analysis was conducted, contrasting the children from the two special schools, i.e. the ‘Special Mild’ and the ‘Special Moderate’ group. In this analysis, Full Scale IQ score was found to be the predictor that best distinguishes children from the two groups (see Table 5.6.3.3.1). The accuracy rate of classifications based on FSIQ scores for these two groups was 89.7% (See Table 5.6.3.3.2).

Table 5.6.3.3.1. : Group Contrasts Between Special Schools : Mild and Moderate

<b>Predictor variable</b>	<b>Rao's V</b>	<b>Change in Rao's V</b>	<b>Groups Means</b>	
			<b>Special (Mild)</b>	<b>Special (Moderate)</b>
Full Scale IQ	67.86	67.86**	86.41	58.05
CEFT	67.91	0.05	8.81	9.18
BADS-C	68.06	1.39	74.54	53.26
Theory of Mind	69.45	1.59	33.47	20.82
CMB	76.35	6.9	20.82	7.45

\*\* $p \leq 0.01$

Table 5.6.3.3.2 No. of Children Classified According to Actual Group (i.e. School Type) and Predicted Group (Based on FSIQ Scores)

<b>Actual Group (School Type)</b>	<b>Predicted Group</b>		<b>Total</b>
	<i>Special Mild</i>	<i>Special Moderate</i>	
Special Mild	15 (88.2%)	2(1.8%)	17
Special Moderate	2(9.1%)	20 (90.9%)	22

#### **5.6.4 Summary of results**

For the first criterion, i.e. concurrence with ICF SEN levels:

1. Out of the initial five predictors, four showed significant correlations with children's level of special educational needs: measures of intelligence (FSIQ); theory of mind (ToM); executive function (BADS-C); and cognitive modifiability (CMB).
2. Children's scores on the test for central coherence (CEFT) did not show any significant correlations with SEN levels after the effect of age was accounted for.
3. There were also high inter-correlations among some of the predictors, especially FSIQ, BADS-C and ToM. This indicated a need for a hierarchical regression analysis in order to identify the indicators that best predict children's levels of SEN. As hierarchical regression analysis requires information regarding the relative importance of the predictors, this is planned as a follow-up analysis, which will incorporate the findings from Study 3 (See Chapter 7).

With reference to the second criterion, the results indicated that:

1. Overall, the combinations of indicators that best distinguished children from the three school-types were measures of intelligence (FSIQ), theory of mind (ToM) and cognitive modifiability (CMB).
2. For individual group contrasts:
  - a. Children's scores on the Theory of Mind battery (ToM) provided the best discrimination between children in the mainstream and special schools for mild difficulties (i.e. 'special mild' group). Based on ToM, the accuracy rate of classifying the children in these two groups was 83.3%.
  - b. Full Scale IQ scores provided the best discrimination between children in special schools for mild and moderate difficulties. Based on FSIQ, 89.7% of the children would be correctly classified.

## 5.7 DISCUSSION

### 5.7.1 Evaluation of the first criterion

*The first criterion relates to the extent to which the indicators show good concurrence with children's levels of special educational needs, based on an independent measure of SEN.*

As discussed in Section 5.4.1, the analysis of the contents of the ICF domains concluded that there was a strong concurrence with aspects of functioning related to intelligence, theory of mind, executive function and cognitive modifiability.

The results of the present study, which indicated strong correlations between children's level of SEN and these four indicators of SEN supported this conclusion.

The measure of central coherence did not show a significant relationship with children's SEN levels. This may be partly due to the lack of content concurrence between central coherence and the ICF domains (as discussed in Section 5.4.1.1). There is also a methodological problem that may account for the lack of observed relationship, namely the lack of sensitivity in the outcome measure used in the test for central coherence, i.e. the CEFT (this is discussed in greater detail in Section 5.7.3).

The high inter-correlations between measures of intelligence, theory of mind and executive function may reflect the theoretical and empirical overlap between these three constructs. As discussed in Chapter 2, it has been shown that children's performance in standardised intelligence tests has a strong relationship with their performance in theory of mind and executive function tasks. It has also been suggested that difficulties experienced by children with autism in theory of mind tasks could be attributed to deficits in executive function (Ozonoff, 1995).

The argument is that because of an impaired ability to use an internal representation, children with autism solve the mentalising problem, e.g. Sally-Ann task, using the best available data, namely what they actually know and see in the box. Thus their responses are driven by the external cue in the environment. So, "what appears to be an inability take another's perspective may actually be a failure to disengage from the external context and use an internal

representation (e.g. someone else's belief) to solve the problems" (Ozonoff, 1995; pg 208). As discussed earlier, one of the key denominators of executive function is the ability to disengage from immediate context (Shallice, 1991; see section 4.2.3.2). The high inter-correlation between children's executive function and theory of mind scores could be interpreted as reflecting the theoretical and conceptual overlap in the two constructs.

At the practical level however, a more pertinent question is whether the information obtained from one measure provides a more meaningful account or explanation of the severity of SEN in children with autism. An evaluation concerning the relative validity of the indicators of SEN is explored in the follow-up analyses, which is described in Chapter 7. The follow-up analyses utilised information on the relative importance of the selected indicators of SEN obtained from Study 3.

### **5.7.2 Evaluation of the second criterion**

*The second criterion relates to the extent to which the indicators are accurate in distinguishing children with autism who can cope with mainstream schools from those that need special schools.*

Results indicated that best way of distinguishing the children in the three groups is by the use of a combination of indicators, namely measures of intelligence, theory of mind and cognitive modifiability. IQ score is one of the entry requirements for admission to special schools in Singapore; it is therefore not surprising that measure of intelligence was found to a significant discriminator. The finding that the measure for theory of mind added significantly to the categorisation of children in the three groups is consistent with the observation that the utility of IQ is strongly improved when combined with other sources of information (Flanagan et al., 1997).

At the same time, the present results highlighted the fact that at different thresholds of special educational needs, different indicators may be more relevant. For children in the lower threshold of SEN, i.e. children in the mainstream and special schools for mild difficulties, theory of mind scores

provided the best discrimination. On the other hand, for children at the higher threshold of SEN, i.e. those more severe needs in the special schools, IQ appears to be more important.

The interaction between the criterion validity of the indicators and the threshold for special educational needs may be explained by the factors that influence the ability of children with autism to cope with mainstream school (as discussed in Section 5.4.2). It was highlighted that children's ability to successfully integrate in mainstream school were strongly related to behavioural and social competencies (Ward et al., 1987; Zetlin, 1987; Farrell, 1997 & 2004; Frederickson et al., 1987 & 2004). Theory of mind abilities reflect a mentalising function, namely the ability to think about other people's thoughts, which has a direct bearing on children's ability to interact or function in social situations (Baron-Cohen, 1988). It could be argued that children with autism who have stronger mentalising abilities, would be able to maintain some form of social relationship with their peers (albeit taking on a passive role) and may be perceived as having less difficulties adjusting to the social and behavioural demands of the mainstream school.

Another dimension in theory of mind abilities is language and comprehension skills (De Villiers, 2000; Fisher et al., 2005). It could be argued that children with strong theory of mind abilities would encounter less difficulty in accessing verbal instructions in class and may be perceived as having a greater ability to cope with the language demands in the mainstream curriculum. Consequently, there may have been less pressure for children with autism who have relatively intact theory of mind abilities (or have some ability to compensate for their deficit in theory of mind), to consider a transfer to special schools.

On the other hand, for children with autism who have more severe difficulties, and who are struggling with the demands of basic self-care and adaptive functions, theory of mind abilities show weaker relationships to children's SEN levels. For these children, the more 'global' measure of intelligence appears to have stronger criterion validity.

There are some methodological issues that need to be considered in interpreting the findings of the present study.

### **5.7.3 Methodological issues**

#### *5.7.3.1 Choice of instruments*

The choice of the specific tests (described in Chapter 4) were based on several factors, such as the extent to which the test has been reported to be a reliable and valid measure of the respective indicator, as well as the appropriateness of the task and language demands for children with autism. However, with the exception of the theory of mind battery, which has been compiled and adapted for the present study, all other tests were selected from commercially available standardised instruments. This has several advantages: firstly, the technical properties of published tests are available for inspection; and secondly, there is a greater likelihood that the tests have been used in other studies involving children with autism or special educational needs and as such, information from these studies could be useful in interpreting the results of the present study.

However, for some of the indicators, the choice of available standardised instruments was very limited, for example the CEFT was the only available test for central coherence that has been standardised for children. The focus of the CEFT is relatively narrow and its content validity may be somewhat limited, i.e. it may not have captured the universe of abilities that are associated with central coherence. The items and tasks in the CEFT might not have adequate breadth and sensitivity to reflect the range of aptitudes and impairments experienced by children with autism in the sample, whose abilities range from significantly below average to superior levels (IQ range for the sample was from 40 to 132 [see Table 5.6.1.1]). The use of reaction time in the CEFT (instead of a pass-fail criteria) as an outcome measure might have provided a more sensitive indicator of children's central coherence, in particular for children with high IQ, where differences in performance may be more visible in terms of the children's speed in detecting the embedded figures. This could be one of the reasons for the absence of a significant correlation between children's scores on central coherence and their level of special educational needs (measured by ICF).

#### *5.7.3.2 Investigator's prior knowledge*

The same researcher carried out all the individual assessments and parent interviews. On the one hand, this minimised the variability that might arise from different styles and quality of assessment/interview. Several efforts were made to minimise the possible bias: standardised administration procedures were strictly adhered to for all data collection; and parent interviews were conducted after all the individual assessments were completed. However, as the investigator was not blind to the children's school placement, the possibility that this information may have affected the quality of interaction between the participants and investigator is still present. The use of multiple investigators for the present study was not feasible and would have introduced other sources of variance, such as inter-tester inconsistencies.

#### *5.7.3.3 Sample selection*

The participants' parents volunteered their involvement in the study. The higher proportion of children from high SES, English-speaking families (See Table 5.3) suggests that parents of children who may not be confident in English might have been less motivated to respond to the invitation letters. Thus, the finding of the present study may have limited generalisability when applied to other groups, for example children from non-English speaking families.

#### *5.7.3.4 Criterion bias*

For the second criterion, children's ability to cope with the demands in mainstream school is based on pre-existing school placement decisions. This has the advantage as a criterion that is completely independent of the present research. However, as IQ scores would have been one of the factors that influenced the decision for a child to be transferred to a special school, there was an inherent bias in favour of IQ scores in the discriminant analyses. Nonetheless, the finding that theory of mind was the indicator that best distinguished children with autism in mainstream schools from those in special schools for mild learning difficulties suggests that, for these groups, the effects of theory of mind scores outweighed any pre-existing bias in favour of IQ scores.

## **5.8 Conclusion**

The aim of this study was to investigate the criterion validity of the selected indicators of SEN. The findings suggest that criterion validity is strongest when multiple indicators are used. This reaffirms the widely held view that the use of multiple sources of information is crucial in ensuring the validity of psychological assessments (APA, AERA & NEMC, 1999; Meyer et al., 2001).

The results of the present study highlight the fact that the criterion validity of specific indicators differs for children at different thresholds of special needs: theory of mind abilities which arguably relate most strongly to children's behavioural and social competencies, show the strongest criterion validity for distinguishing children with autism who can cope with mainstream school. For children with more severe special needs who require special schools, indicators of general intelligence seem to show stronger criterion validity in discriminating between available types of special school.

At a broader level, the findings of Study 2 reaffirm the view that evaluation of validity is very much context dependent (APA, AERA & NEMC, 1999); the criterion validity of an indicator depends not only on the purposes of assessment, but also on the characteristics of specific groups in the target population. The present study has focused on one purpose of SEN assessments, namely educational placements. Another important purpose of SEN assessment is the implementation of appropriate intervention to address the needs of children with autism. Thus, for a comprehensive evaluation of the validity of SEN indicators, criterion validity per se, may not be fully adequate; another critical aspect is treatment validity, which is the focus of the next study.

## CHAPTER 6

### STUDY 3: TREATMENT VALIDITY OF THE INDICATORS OF SPECIAL EDUCATIONAL NEEDS FOR CHILDREN WITH AUTISM

#### 6.1 INTRODUCTION

The aim of the present thesis is to identify valid indicators of special educational needs (SEN) for children with autism. Based on a review of research on special educational needs and autism (presented in Chapter 2), several indicators of SEN were identified, namely measures of intelligence, theory of mind abilities, executive function, central coherence and cognitive modifiability. As discussed, there are theoretical bases for using these indicators as a reflection of the level of functioning and impairments in children with autism. However, as they are intended for use in high stakes educational assessment, theoretical support *per se*, is insufficient; these indicators must be evaluated against established criteria for the evaluation of psychological assessment, such as the *Standards*' (APA, AERA & NCME, 1999). Two level of evaluations were identified: firstly, evaluations that relate to the psychometric properties of a particular measure, such as analyses of reliability, content and criterion validity; and secondly, evaluations that focus on the social and practical consequences of using the measures as indicators of SEN, such as analysis of treatment validity.

Study 2, described in the previous chapter, focused on two aspects of criterion validity: 1) the extent to which the indicators showed good concurrence with children's level of special educational needs, based on an independent measure of SEN; and 2) the extent to which the indicators were accurate in distinguishing children with autism who can cope in mainstream schools, from those that need special schools. The results from study 2 suggested that, except for central coherence, all other indicators showed strong concurrence with children's SEN level, as measured by the ICF. It also provided some evidence for the psychometric (i.e. criterion) validity of three indicators of SEN, namely measures of intelligence, theory of mind and cognitive modifiability (i.e. the combination of scores from these three measures provided the best classification of children

from the available school-types). However, as discussed earlier, in high stakes assessment evaluations based on psychometric validity *per se* are inadequate; there is also a need for evaluation that focuses on the social and practical consequences of using the measures as indicators of SEN, i.e. treatment validity.

## **6.2 TREATMENT VALIDITY**

Broadly defined, treatment validity (sometimes referred to as treatment utility or instructional validity) is the degree to which any assessment procedure contributes to beneficial outcomes for individuals (Cone, 1989). Three key components of treatment validity have been identified (Gresham, 2002):

1. Treatment utility, i.e. the extent to which there is evidence for test use, particularly as it relates to positive consequences and utility;
2. Cost-benefit analysis, i.e. evaluation of the features or properties of the test that could either facilitate or limit its utility;
3. Incremental validity, i.e. the extent to which the use of a test improves existing assessment procedures.

Gallery and Hofmeister (1978) highlighted three factors that affect the evaluation of treatment validity: 1) the variability in the relationship between the test and treatment; 2) variability in the quality of treatment; and 3) the variability in the relationship between the proposed treatment and the actual curriculum. In most cases, it is not feasible to control all of these factors; hence evaluations of treatment validity tended to focus on the first issue, namely the relationship between test information and treatment. This focus may be justified, given the fact that an important pre-requisite for treatment validity is the extent to which test users are able to make a clear link between information derived from the test and the decisions about the interventions/treatment programmes that will address children's needs. The present study focuses on this important pre-requisite and seeks to evaluate the treatment validity of the selected SEN indicators, based on the views of the intended test users, i.e. practitioners who are directly involved in providing assessment and advice on the SEN of children with autism.

Based on the findings of Study 2 (See Chapter 5), four indicators were identified as having a significant correlation with children's SEN levels, namely measures of intelligence, theory of mind, executive function and cognitive modifiability.

There is a wide variation in the extent to which the treatment validity of these indicators has been previously studied and discussed in the literature. These will be discussed in the following sub-sections, with reference to the specific tests that were selected for the present thesis.

### **6.2.1 Treatment validity for the measure of intelligence**

The lack of evidence for treatment validity has been one of the strongest criticisms raised against IQ tests (McGrew, 1993; Siegal, 1989). For example, although the developers of the Wechsler Scales claim that information from the test has high utility for clinical and educational uses, the burden of proof for treatment validity is left to the individual users (as discussed in Section 4.1). In the literature, there are reports of the use of the Wechsler Scales for diagnoses, identification of cognitive strengths and weaknesses, and for instructional planning. These are discussed in turn.

- *Diagnostic utility*

It was reported (Hynd, Cohen, Ricio & Arceneaux, 1998) that one of the most important uses of intelligence tests in the United States is in the diagnosis of specific disorders. For example, the DSM-IV's (APA, 1995) definition for mental retardation (MR) includes an IQ-based criterion<sup>29</sup>.

In the diagnosis of autism, Klin et al. (1997) singled out intelligence (IQ) as a key aspect of cognitive assessments for children with autism. This is linked to the DSM-IV and ICD-10 diagnostic criteria, where general ability is a key factor in the differential diagnosis for Asperger Syndrome (See Section 2.2.1).

However, as discussed in Chapter 2, some commentators disagree with this categorical framework for defining autism and argue that Asperger's Syndrome can occur in individuals of all levels of ability (Wing, 1981; Ehlers & Gillberg, 1993).

---

<sup>29</sup> One of the criteria for mental retardation in the DSM-IV (APA, 1995) is the 'presence of significantly sub-average intellectual functioning - an intelligence quotient (IQ) of approximately 70 or below'.

- *Identification of cognitive strength and weaknesses*

One of the claims of the WISC-III is that subtest scores and factor-based indices can be used to identify ‘cognitive strengths and weaknesses’ that can be the basis for educational planning. As discussed in Section 4.1, although the empirical basis for this claim is weak, the practice of interpreting Wechsler factor scores and subtest profiles as indications of children’s relative strength and weaknesses is commonly used in practice. Critics have pointed out that studies which looked at the reliability of the subtest scale scores and factor based indices have consistently shown that the WISC-III results are better interpreted (i.e. show stronger validity) as a single factor, namely the *g* factor, rather than two, three or four factors; and the classifications based on discrepancies or differences in subtest scores are unstable (Barnett & Macmann, 1992). Contrary to the assumption that large variations in subtests scores are clinically significant (Wechsler, 1993), it has been shown that scatter (i.e. differences among Wechsler factor or subtest scores) is more common (i.e. normal) than flat profiles or very small scale differences (Glutting, McDermott, Watkins, Kush, & Konold, 1997). Nonetheless, the lack of evidence for validity does not seem to deter practitioners from recommending the use of factor-based indices and subtests profiles to describe children’s cognitive strengths and weaknesses (Rechley, 1997; Kaufman, 1979).

- *Instructional planning*

A methodology for instructional planning commonly espoused by some practitioners is modality matching, which aims to match children’s cognitive strength in processing information (in either the auditory, visual or kinesthetic mode) with instructional approaches that utilise the child’s strength (Reynolds, 1992). For example, in children with autism, superior performance in non-verbal IQ tasks is often used as a basis for recommending the use of strategies that utilise the visual modalities (Hodgdon, 2000). However, a key assumption in this approach is the existence of a treatment by aptitude interaction (ATI), i.e. there must be an interaction between the presumed aptitude strength and instructional methodology (e.g. that students will do less well if the instructional methodology

is not matched to their strengths). However, reviews of the ATI research conclude that there is little evidence to support the existence of ATI for IQ scores. More importantly, matching presumed cognitive strengths with instructional methodologies does not seem lead to demonstrable differential gains in academic achievement, regardless of whether the aptitude strengths are conceptualised as modality preferences (Rechley & Gresham, 1989; Reschly, 1997).

### **6.2.2 Treatment validity for the measure of cognitive modifiability**

In contrast to intelligence measures, the utility of dynamic assessments<sup>30</sup> (DA) for planning intervention has often been reported as one of its significant strengths (Tzuriel, 2001; Lidz, 1991; Sternberg & Grigorenko, 2000). Although DA is not intended for use in the diagnoses of specific disorders, a key claim by proponents of DA is that it is designed to identify deficient cognitive functions as well as to change or improve these deficiencies. As discussed in Section 4.4, in DA tests for cognitive modifiability, during the mediation phases, key components affecting children's learning needs are identified by means of a detailed analysis of: 1) within-child cognitive factors, e.g. ability to make comparisons, to generate hypothesis, motivation and attention; and 2) external inputs, i.e. the contents of the assessors mediation, e.g. the modality in which the task is presented (i.e. verbal, visual, pictorial), the level of task complexity, and specific cognitive strategies required to bring about successful performance. Proponents of DA argued that these analyses can provide the insights that can result in useful recommendations for teachers and parents (Lidz, 1991).

There are several assumptions underpinning the treatment utility of DA; the most critical is that cognitive processes can be inferred from responses to test items in DA and consequently, instructions can be designed to overcome the weakness to produce more sophisticated cognition. Evidence to date suggests that the empirical basis for this assumption is not convincing (Reschly, 1997). Instead, accurate classification and identification of cognitive processes based on responses in DA seems to be problematic. For example, studies by Haywood,

---

<sup>30</sup> The terms dynamic assessments (DA) and tests of cognitive modifiability are used interchangeably.

Brown and Wingerfeld (1990) suggest that even among 'world-class' dynamic assessment experts, the inter-judger agreement regarding the identification of categories of cognitive deficits is low to moderate. It can be argued that if the interpretations linking children's responses in DA and the underlying cognition is unreliable, the claims regarding the validity of interventions that are based on such interpretations would be suspect.

Although it has yet to take root in mainstream practice, among educational psychologists who have had exposure to dynamic assessments there appears to be a strong perception that tests for cognitive modifiability have high utility for intervention purposes (Deutsch & Reynolds, 2000). More specifically, the key benefits of measures of cognitive modifiability that were identified by these educational psychologists (N=88) were as follows:

- DA's methods provide practical advice for teachers and parents (51%);
- DA's focus on strengths rather than weaknesses enhances the self-esteem for the child (46.6%);
- DA is perceived as a 'superior alternative' to psychometric approaches (21.6%);
- DA provides 'rich information' that is perceived to be unique, i.e. not easily obtained in other ways (20.5%).

At the same time, the same group of EPs reported that the main limitations (or difficulties) of using DA were:

- The additional amount of time needed for assessment, report writing and feedback (30.7%);
- The difficulty of linking DA to the curriculum and classroom practices (24%);
- Preference by local education authorities (LEAs) for standardized tests over DA procedures (20.5%);
- Subjectivity inherent in the interpretation of DA results (19.3%).

Based on the above findings, Deutsch and Reynolds (2000) concluded that the difficulties were primarily due to the lack of training and practical expert support in the use of DA approaches, and argued for a greater LEA investment in and

commitment to the training and use of DA methods in the UK. This seems to imply that the lack of practical utility of DA methods is ‘blamed’ on the users’ lack of skills and expertise. However, some the difficulties expressed by EPs in the study can be seen to reflect problems inherent to the DA methods (e.g. subjectivity in interpretation and lack of direct relation to the curriculum). It can be argued that it would be premature to call for further investment/resources for DA approaches before establishing the incremental validity of the measures for cognitive modifiability, i.e. demonstrating that DA methods can provide reliable information that is not obtainable by other modes of assessments; and that the benefits of its use, outweigh the additional demands in time and resources.

### **6.2.3 Treatment validity for measures of executive function.**

Tests for executive function, such as the Behavioural Assessment for Dysexecutive Syndrome are reported to be widely used by clinicians for diagnostic purposes, e.g. in the diagnoses of Traumatic Brain Injury (TBI) and Attention Deficit Hyperactivity Disorder (ADHD) (Emslie et al., 2003).

- *Diagnostic utility*

In the BADS-C’s technical manual, data regarding the performance of children with ADHD, dyslexia and pervasive developmental disorder (including autism) were included, implying the test’s utility for these clinical population. However, the mean scores for the different clinical groups did not show any clear clinical pattern that can guide differential diagnosis. Moreover, impairment in executive function is not one of the diagnostic criteria for ADHD, dyslexia or autism. In the Diagnostic Interview for Social and Communication Disorder, or DISCO (Wing, 1999; see Section 2.5.3 in the present thesis), there are a small number of items relating to children’s repetitive routines and resistance to change which may be seen as reflecting impairments in executive function. However, these are based on the child’s behaviour in daily life activities (e.g. insistence on sameness, asking repetitive questions) and not on direct observations of the children’s responses to cognitive tasks.

As highlighted by Hill (2004), there are two key problems in using executive functions as a diagnostic marker for autism. Firstly, while there is clear evidence

that individuals with autism experience deficits in areas of executive functioning, there is a lack of clarity as to which aspects of executive function are typically problematic in autism. Secondly, difficulties in executive function may not be a universal feature of autism; some studies have found that the tests of executive function that they have employed have not been problematic for all individuals with autism, especially those with normal IQ levels (Baron-Cohen et al., 1999; Russell & Hill, 2001; Hill & Russell, 2002). It can be argued that until there is a clearer fractionation of executive function skills and identification of the specific skills that are impaired in children with autism, the diagnostic utility of measures of executive function may be limited.

- *Treatment planning*

There have been some studies that demonstrate the usefulness of rehabilitation focusing on executive functions in individuals with brain injury (Marshall, Karow, Morelli, Iden, Dixon, & Cranfill, 2004; Burgess & Aldermann, 1990; Fasotti, 2003). However, these re-training programmes were intended for patients, for whom the individual component skills (e.g. memory) have already fully developed and what is lost or impaired is the ability to initiate their use. The applicability of these findings for children with autism may be limited because in these children, the impairment stems from the inability to acquire executive functions.

Reviewing the use of cognitive-neuropsychological tasks in the assessment of young children with autism, Klinger and Renner (2000) reported that in practice, there is an increase in the use of cognitive treatment programs for autism focusing on impairments in executive function, theory of mind, selective attention, and abstraction. However, it was noted that the adoption of these treatment plans were not integrated or linked to the information obtained during diagnostic / assessment processes.

#### **6.2.4 Treatment validity for measures of theory of mind**

The test-battery approach in the measurement of theory of mind (ToM) abilities that is used in the present thesis have only been used for research purposes. However, some of the well known mentalising tests (for example, the Sally-Ann

test described in Section 4.4) are increasingly being used by some practitioners working with children with autism.

- *Diagnostic utility*

Although both theory of mind and executive function are linked to causal theories of autism, in practice there appears to be a clearer relationship between theory of mind and the diagnostic approaches for autism. In some of the diagnostic tools for autism (described in Section 2.5), behavioural characteristics that reflect deficits in theory of mind abilities are included. For example in the DISCO, there are many items relating to impairments in theory of mind, such as the child's use of one-sided communication, literal use of language, limited appreciation of humor, one-sided social approaches, lack of awareness of others' feelings, inability to share others' happiness or distress, limited sharing of interests, and lack of reciprocal friendship. However, the use of theory of mind as a diagnostic marker is currently limited by the fact that an impaired performance in mentalising tasks is not universal for children with autism, i.e. there are a very small sub-group of children with autism that appear to have an intact theory of mind (as discussed in Section 2.4.2.1).

- *Treatment planning*

There have been some attempts to remediate the absence of theory of mind in individuals with autism to improve empathy or social perspective taking (Klin & Volkmar, 2000). Despite findings that show improvements on experimental theory of mind tasks, the results have not generalised broadly to conversational skills (Hadwin, Baron-Cohen, Howlin & Hill, 1997) or social competence in real life situations (Ozonoff & Miller, 1995).

Another treatment approach that is linked to theory of mind is the use of Social Stories. These innovative stories, developed by Gray (1994), have been widely used to teach higher functioning individuals with autism the social understanding needed to interact in a variety of social situations. The stories typically depict a social situation that an individual is likely to encounter, and include opportunities to learn the appropriate ways of responding to the social situation, including

learning the appropriate perspective statements (i.e. statements which provide information on others' perspectives and likely reactions). Opportunities for practice and transfer of the skills to real situations are built into the treatment programme. A few research studies have demonstrated that social stories can effectively enhance social interaction in higher-functioning individuals with autism (Norris & Dattilo, 1999; Swaggart, Gagnon, Bock, Earles, Quinn, Myles, & Simpson, 1995; Sansosti, Powell-Smith, & Kincaid, 2004). However, perspective taking is only one of the components of the Social Stories intervention programmes; and it is difficult to link its efficacy specifically to children's theory of mind.

Despite the emergence of specific interventions for children with autism that are aimed at theory of mind abilities, what remains unclear is the extent to which information about a child's theory of mind abilities has been utilised in the planning of the treatment goals/outcomes. This issue is of greater concern in 'packaged' treatment programmes that are aimed at specific clinical groups, e.g. the Social Stories. As these 'packaged' interventions become increasingly popular, it may well be that a diagnosis of autism *per se* would initiate the child's access to such treatments. Thus, the reported increase in the use of such treatment programmes (National Research Council, 2001) does not necessarily reflect evidence of their treatment utility. Further evidence would be needed to see the extent to which practitioners are utilising information on children's theory of mind (if at all) to determine treatment goals and approaches.

### **6.2.5 Summary**

There are three components in the evaluation of treatment validity: analyses of utility; incremental validity; and cost-benefit analyses.

In terms of utility for assessment:

- The utility of IQ scores seems to be linked to the diagnostic criteria that are used for the various disorders. For example in the DSM-IV criteria for mental retardation (MR) and Asperger's Syndrome, children's IQ is specified as one of the criteria for differential diagnoses. However, for practitioners that do not subscribe to DSM-IV's categorical framework for diagnosis, the utility of IQ scores for diagnoses may be limited.

- Despite their widespread use, there appears to be very little evidence supporting the validity of cognitive profiles based on IQ sub-subtests scores.
- Although tests for cognitive modifiability are designed to capture information regarding cognitive processes and yield rich descriptions about children's learning, the reliability of the interpretation/judgment regarding children's deficient cognitive functions is problematic. This limits the validity of the treatment plans that are based on these judgments.
- Both executive function and theory of mind are not part of the diagnostic criteria for autism; however there appears to be a greater inclusion of behaviours relating to theory of mind in some of the leading diagnostic instruments for autism. In addition, although both impairments in executive function and theory of mind are not universal in children with autism, the potential use of theory of mind as a diagnostic marker may be more plausible given the fact that deficits in theory of mind are held to be unique to children with autism (Baron-Cohen et al., 1995; Hill & Frith, 2004).

In terms of utility for treatment planning:

- Although widely used, there is very little evidence of the validity for using IQ scores as the basis for 'modality matching', i.e. prescribing instructional approaches that utilise a child's dominant learning mode.
- There is evidence for the utility of re-training programmes for executive function, but these were based on executive functions that were lost through brain injury. They may not be applicable for children with autism where the impairment in executive function skills is developmental.
- While there is evidence that measures of cognitive modifiability can be used to improve children's performance in the specific DA tasks, there are doubts regarding the applicability of these skills to the curriculum or classroom activities.
- Treatment programmes focusing on theory of mind have shown limited utility: programmes addressing deficits in theory of mind in children with autism have shown some effectiveness but limited generalisability; and the utility of Social Stories intervention packages cannot be singularly attributed to improvements in skills which are related to theory of mind.

For all four measures there seems to be little emphasis on evaluation of incremental utility or cost-benefit analysis. From this discussion, three factors affecting the treatment validity emerge:

1. The links between the test information and the criteria/processes used for diagnoses;
2. The validity of the descriptions of children's strength and weaknesses based on the test scores; and
3. The validity of the treatment plans.

These factors would need to be considered in the evaluations of treatment validity of the selected measures.

## **6.3 METHODS**

### **6.3.1 Objectives of Study 3**

The aim was to obtain practitioners' responses to the following questions concerning each of the four indicators of SEN, namely the Weschler Intelligence Scale for Children (WISC-III), Behavioural Assessment for Dysexecutive Syndrome for Children (BADS-C); Theory of Mind Battery (ToM Battery), and Cognitive Modifiability Battery (CMB)<sup>31</sup>:

1. In what ways is the test information useful for children with autism (i.e. does it have treatment utility)?
2. In what ways would the use of the test improve current assessment practices for children with autism (i.e. does it have incremental validity)?
3. What are the benefits of using the test (i.e. cost-benefit analysis)?
4. What are the costs or difficulties of using the test (i.e. cost-benefit analysis)?

These questions were derived from the three components of treatment validity defined by Gresham (2002), as discussed in Section 6.2. The views were obtained using focus group discussion. Focus groups are planned sessions where individuals discuss ideas and perceptions focused around a topic of interest (Kruegar, 1988). There are advantages of using a focus group over individual

---

<sup>31</sup> Each of the measures is described in detail in Chapter 4.

interviews that makes this approach particularly useful for the purpose of the present study:

- Vaughn, Schumm and Singub (1996) highlighted that participants in focus groups can see that a diverse range of views are welcomed and valued. There is no requirement or pressure for an individual to answer every question, so the responses made are more genuine and substantial. It was also argued that individuals are likely to feel more supported, relaxed and confident than in an individual interview with an unfamiliar interviewer. These advantages are particularly relevant for the present study, as the research questions involved the expression of views that may indirectly expose the participants' shortcomings and lack of knowledge in the assessment of children with autism.
- Focus group discussion (FGD) is particularly useful when eliciting views and perceptions on topics/issues which are relatively new, and where there may be wide variability in respondents' knowledge and experiences. Other than the measure of intelligence, the indicators of SEN used in the study were relatively less well known to psychologists in Singapore. The FGD format provided an opportunity for all the measures to be described and demonstrated before participants' views about their treatment validity were sought.

### **6.3.2 Structure of the focus groups**

A scripted protocol was developed for the focus groups, which included:

1. Opening remarks by the moderator, which included statements about the purpose and context of the focus groups, and assurances of confidentiality and anonymity for all participants.
2. Stimulus presentation by the moderator, which included descriptions of the individual tests and demonstrations of the BADS-C, CMB and ToM Battery<sup>32</sup>.
3. Focus group questions, as well as probes that can be used by the moderators to facilitate the discussions.

---

<sup>32</sup> The WISC-III was not demonstrated as all the participants have extensive experience of its use.

(See Appendix O and P for power-point slides for the stimulus presentation and moderators' script respectively).

### 6.3.3 Field trial

A field trial of the stimulus presentation materials and moderators' script was carried out with two educational psychologists in training as participants of the FGD. The field trial FGD was audio recorded and was used as the basis for moderators' training which involved reviews of the audiotape of the focus group field-trial and rehearsal of the moderators' script. The training aimed to minimise variability in the style and approach used by the two moderators.

### 6.3.4 Participants

Two groups of educational psychologists and clinicians participated in the focus groups. They were selected to represent the key public agencies that provide assessment and advice for children with autism in Singapore. Both groups, FGD team 1 and FGD team 2 had similar numbers (n=7), but FGD1 had more participants with more than 5 years practical experience (See Table 6.3.4).

Table 6.3.4 : Information About Participants of Focus Group Discussion (FGD) Teams

Context /settings of practice	FGD Team 1	FGD Team 2	Total
Practitioners working with children in mainstream schools (primary, secondary and higher education colleges)	2	1	3
Practitioners working with children in special school for mild disabilities	1	2	3
Practitioners working with children in special school for moderate to severe disabilities	2	2	4
Practitioners working with children across all educational contexts/settings, e.g. EPs; CPs & psychiatrists in the health/medical agencies.	2	2	4
<b>Total Number</b>	<b>7</b>	<b>7</b>	<b>14</b>
No. of participant with less than 5 years of practice	2	7	9
No. of participant with 5 years of practice or more	5	0	5

### **6.3.5 Procedure<sup>33</sup>**

Invitations for the participants were sent out to 24 educational and clinical psychologists working in all public agencies in Singapore that provided assessment and advice services for children with autism. 16 (67%) responded positively to the invitation, but two were unable to attend the session. Upon arrival at the venue, all 14 participants independently completed the informed consent forms and were grouped according to their respective teams (see Appendix Q for a sample of the informed consent form). An attempt was made to ensure that, as far as possible, participants with more than 5 years practical experience were grouped together, while keeping the number of participants in each group the same. This is to ensure that less experienced participants would not be intimidated by the more experienced practitioners who may be more vocal in expressing their views.

The focus groups started with a lecture-style presentation and demonstration of the individual tests, which was attended by all participants. During the presentation, participants were given the opportunity to clarify information about the tests. This was then followed by discussions in the respective FGD teams, with one moderator facilitating each team. The moderators were experienced educational psychologists, with extensive experience in facilitating multi-disciplinary group discussions. In addition one of the moderators has extensive experience conducting focus group research. Four focus discussions, each lasting between 45 and 60 mins., were conducted with each FGD team. All 8 focus group sessions were audio recorded and transcribed to capture participants' verbatim responses.

### **6.3.6 Data Analysis**

The transcripts were analysed using a qualitative approach, described by Vaughn et al. (1996), as follows:

1. Key themes or 'big ideas' were identified. In the present study, a 'top-down' approach was used, and the key themes were derived from the three key

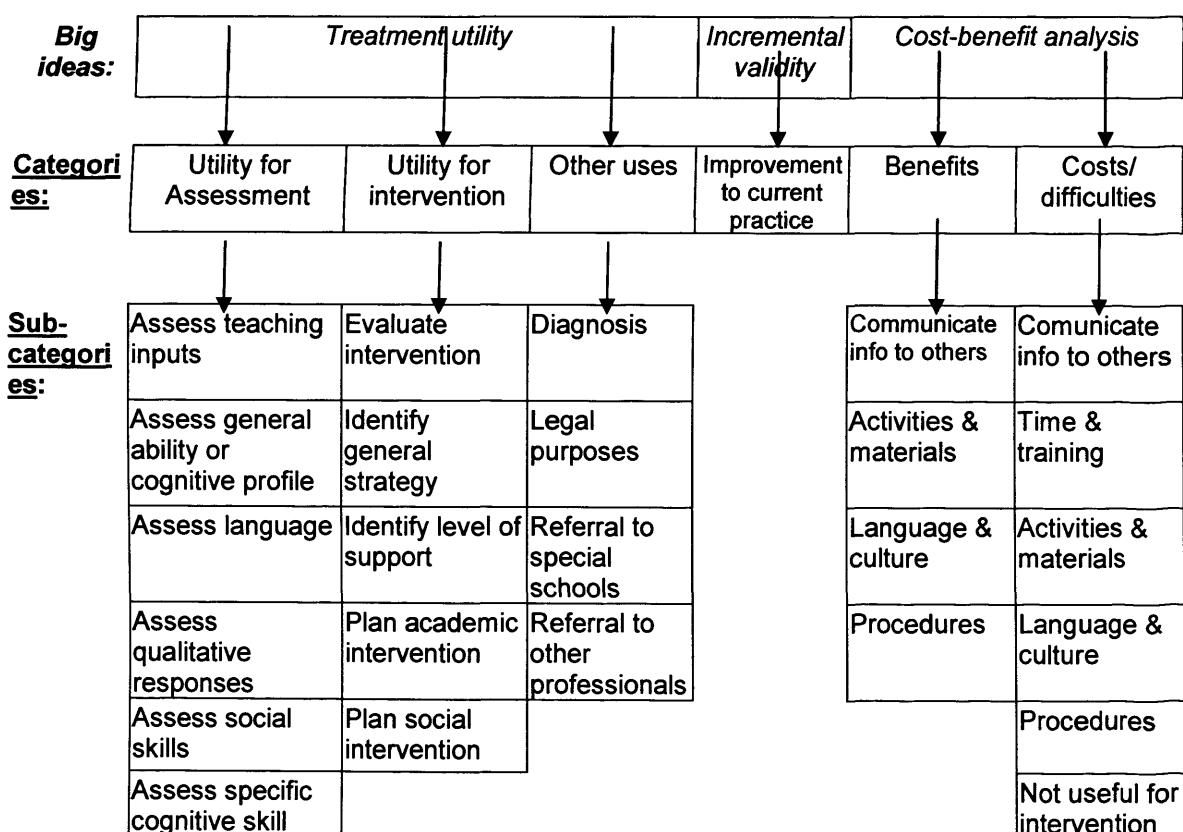
---

<sup>33</sup> Procedure for the study was approved by UCL ethics committee for non-NHS human research (Project ID number 0038/002)

components of treatment validity as identified by Gresham (2002): treatment utility; incremental validity; and cost-benefit analysis.

2. Units of information (i.e. phrases and /or sentences in the transcript) relevant to the research questions were identified and highlighted. Category headings, or 'codes', were selected to sort and group these units of information. The units of information were assigned to the codes to enable most of the units to be placed within a category. A software package for qualitative data analysis, winMAX (Kuckartz, 1998) was used.
3. The codes were refined to establish category headings that most economically accommodated the relevant units of information. Following Miles and Hubermann (1988), codes/categories used in analysis were organized along a structure that was derived from the research questions. As the process of unit assignment proceeded, sub-codes were added to reflect the ideas and views that emerged during the FGD. These newly evolved sub-codes were mapped onto the coding structure (See Fig. 6.3.6).

**Fig. 6.3.6 Structure for Coding Categories for Comments on Treatment Validity**



### 6.3.6.1 *Definition of codes*

During the process of coding, operational definitions of codes were developed and refined. To ensure clarity and minimise overlaps in the assignment of units, the operational definition of each code was specified along 3 parameters:

1. Description: A statement describing the criteria used for unit assignment.
2. Inclusion: An elaboration of the range of comments that met the criteria and are likely to be included in the code.
3. Exclusion: Examples of comments that are, although closely associated with the target units in the code, better accommodated with other coding categories available in the code structure.

Table 6.3.6.1 presents an example of the operational definition for the codes. A full list of the definitions for codes appears in Appendix R.

**Table 6.3.6.1 An Example of Operational Definition for Codes**

<b>Code</b>	<b>Utility for Intervention</b>
1. Description	Use of test information to plan and design intervention for children.
2. Inclusion	Use of test information to identify general approaches to intervention, identify level of support needed; planning academic, cognitive, social and other interventions; planning Individual Education Plans (IEP); and evaluate intervention outcomes.
3. Exclusion	Use of test for diagnosis, referral to special schools, referral to other professionals (consider alternative code 'Other Uses').

### 6.3.6.2 *Reliability*

To evaluate the consistency with which the respective codes were assigned to individual units, an independent rater (an educational psychologist in training) coded segments of the transcripts across two focus groups. Inter-rater reliability, based on the calculations of percentage agreements between the two coders, was found to be 81% across 98 (16%) coded units.

## **6.4 RESULTS**

### **6.4.1 Method of summary and interpretation**

Krueger (1988) identified two broad ways in which the units/information elicited from FGDs can be summarized: 1) qualitatively, i.e. by describing all the categories of comments that emanated from the FGDs; and 2) quantitatively, i.e. by reflecting the frequency of comments generated under each of the categories.

The former, qualitative method is useful for the purpose of obtaining information on the range or type of comments that reflected participants' views in the FGD; while the latter method can be useful in reflecting the categories of comments that were most prevalent (or dominant) in the FGDs. The qualitative method has the advantage of providing rich information on the subject of interest (i.e. treatment validity); but its main limitation is that the categories would not distinguish isolated comments from individual participants, from consensus views expressed in the FGDs. On the other hand, while the use of frequency analysis can provide a sense of the most prevalent/dominant views, its main danger is that the numbers and percentages associated with frequency analysis may convey the (false) impression that the results can be generalised to a population. As highlighted by Krueger (1988), the two methods represent two end-points in the 'analysis continuum' for focus groups (pp 109); and depending on the purpose of the study, both methods (either in combination or independently) could be appropriate.

In the present study, the focus groups were used to elicit the range of categories which reflected the treatment validity of each of the four tests, e.g. the different ways in which the test can be used to plan interventions, assessments, and to improve current practice. In addition, as the information will also be used to compare the treatment validity of the four tests, information that reflects the dominant/prevalent categories is also sought. Thus, in the present study, results are based on an analysis of the type and range of categories elicited as well as the number of comments/units falling into each category heading or code.

#### **6.4.2 Treatment validity of the measure of intelligence (WISC-III)**

Table 6.4.2 shows the categories of comments/units from the FGDs on the WISC-III related to the three 'big ideas' in treatment validity (cf. Fig. 6.3.6), namely treatment utility, incremental validity and cost-benefit analysis.

### *Treatment Utility*

Comments related to the utility of the WISC-III for 'Other Uses' and 'Assessments' dominated participants' responses. These include its use in the

Table 6.4.2. Summary of Categories Elicited for WISC-III

<b>Big Ideas</b>	<b>Categories &amp; Sub-categories</b>	<b>Dominance</b>	
<b>Treatment Utility</b>	<b>Utility for Assessment</b>	<b>35</b>	<b>22%</b>
	Assess Gen. Cognitive Ability	12	8%
	Assess Language Function	5	3%
	Assess Qualitative Responses	15	9%
	Assess Social Skills	1	1%
	Assess Specific Cognitive Skill	2	1%
	<b>Utility for Intervention</b>	<b>10</b>	<b>6%</b>
	Identify General Strategy	5	3%
	Plan Academic Intervention	5	4%
	<b>Utility : Other Uses</b>	<b>36</b>	<b>23%</b>
	Diagnosis	10	6%
	Referral	22	14%
	Referral To Other Professionals	4	3%

<b>Incremental validity</b>	<b>Improvement to Current Assessment Practice</b>	<b>0</b>	<b>0%</b>
-----------------------------	---	----------	-----------

<b>Cost-benefit analysis</b>	<b>Benefits</b>	<b>2</b>	<b>1%</b>
	Communicate Info To Others	2	1%
		<b>77</b>	<b>48%</b>
	<b>Costs/Difficulties</b>		
	Communicating Info To Others	6	4%
	Time And Training	10	6%
	Activities/Materials	4	3%
	Language/Culture	11	7%
	Procedures	23	14%
	Not Useful For Intervention	23	14%
	<b>Total Comments/Units</b>	<b>160</b>	<b>100%</b>

referral to special school, diagnosis and referral to other professionals (See Table 6.4.2; Code 'Other Uses'). Participants frequently commented that the utility of IQ scores was predetermined by the admission criteria for special schools in Singapore (See Table 6.4.2.1, example no. 9). As discussed earlier, among psychologists in the US, it was noted that the diagnostic utility of IQ scores was strongly influenced by 'external' factors that place high weighting on IQ scores, e.g. the DSM-IV diagnosis of MR and Asperger's syndrome. This is considered an 'external' consideration as it does not reflect the clinical/ educational utility inherent to the test. In the Singapore context, the utility of IQ scores seems to

also have been influenced by similar 'external' considerations, namely the admission policies for special schools<sup>34</sup>. Participants' comments in the FDG reflected a situation where the use of IQ tests is considered only after an initial decision regarding the child's need for special school is made, i.e. to confirm eligibility for special school. It does not seem that IQ scores are used to inform decisions about a child's need for special school a priori.

Another dominant category that emerged in the FGDs is the utility of the WISC-III for assessment, in particular the assessment of children's general cognitive ability or profile (See Table 6.4.2.1; example no. 1). As discussed in Section 4.1 the use of the WISC-III for measurement of general ability is consistent with the intended purposes of the test; however its use for describing profiles of cognitive strengths and weaknesses rests on very limited evidence of validity. However, like other practitioners in the US and UK (Rechley, 1997; Kaufman, 1994), the lack of validity evidence does not seem to hinder the participants' perception that IQ scores are useful for such purposes. This can be partly attributed to the large amount of data provided in the WISC-III technical manual relating to inter-subtests discrepancy analysis (implying value), without the research evidence that supports its validity.

**Table 6.4.2.1 Examples of Coded Units (Comments) for the WISC-III**

<b>Coding categories</b>	<b>Examples of units</b>
<b>Utility for Assessment</b>	
Assess general cognitive ability/profile	1. For kids with autism, (information on) the discrepancy between VIQ & PIQ score and the type of scatter (of scaled scores) is useful.
Assess language	2. For example, in language. A lot of them can do the Vocabulary - where a lot of the answers are 'normal' (straightforward answers) and a lot of the questions are short. But when you move on to Comprehension, you find that your sentences and questions get longer. And the answers (from the child) get more disjointed. You can then pick up on how the child is coping with extended questions or length of sentences or sentence constructions and their (language) expression.
Assess child's qualitative responses	3. How the child responds to the task presentation of the WISC. (For example), how the child approaches the visual presentation of the task, how they respond to demonstration, how they then approach the task (after the demonstration). These qualitative aspects (are

<sup>34</sup> In all special schools, an IQ based criteria is one of the pre-conditions for entry, e.g. a child who seeks admission to a special school for moderate learning difficulties must have an IQ score of between 50 to 70

	useful).
Assess social function	4. (Its useful) for social information, such as in Picture Arrangements - what (social) cues do they look out for?
Assess specific cognitive skill	5. Another (useful) info is the auditory processing issue, from the digit span subtest.
<b>Utility for intervention</b>	
Identify general strategy	6. (Its usefulness is) not so much for intervention per se, but more on the kinds of general directions (i.e. guides) for teachers.
Plan academic intervention	7. For intervention purposes, the one that is most easily translatable into academic work is (info from) the Verbal subscale. For example, the child has a high score for Information and a low score for Comprehension, (but) the child is failing his school exam. So the scores show that the child may have a good store of general knowledge but may not be able to apply or consolidate (the knowledge).
<b>Other uses</b>	
Diagnosis	8. ... The typical profile of (a child with) very high block design (scores) and very low object assembly scores, gives useful diagnostic information. For diagnostic formulation, this gives one source of evidence that can support other sources of information.
Referral to special school	9. For special schools, the key reason, 90% of the time, IQ test is used for referral purposes. So after you do the IQ tests, you just need to check that this child is eligible for the next school that the parents wants him to go. So in a way, if we confirm (eligibility), then we prepare the report for referral out.
Referral to other professionals	10. But if your verbal IQ is below 50 and your performance IQ is 120, then, in this case, they would try and refer the child to a speech and language therapist
<b>Benefits</b>	
Communicate info to others	11. We also find the information useful to feedback to parents, for e.g. we can say, "Your child is this age but actually he is functioning at this level". It makes more sense to them (parents).
<b>Costs</b>	
Communicate to others	12. Teachers don't really understand or don't really know how to interpret the (IQ) scores.
Time & training	13. In the clinic setting, because of the time constraints that we have, it may be difficult to do the modifications that you all suggested.
Activities & materials	14. That's the thing (issue) about testing children with anxiety problems - if you don't break it down and structure the activities, you won't get the best responses.
Language & culture	15. In terms of questions, even more difficult is the Comprehension questions - some of which are super political. A lot of the things our kids would not have picked up in school. It's the culture-loading factor.
Procedures	16. I agree. For kids with autism, I don't think I've done a single standard administration or demonstration before ((laughter)).
Not useful for intervention	17. Generally, we don't look at it (i.e. WISC scores) in terms of planning or intervention. The intervention does not somehow appear high-up (in our priority of uses).

### *Cost-benefit analysis*

Overall, very few comments were made regarding the benefits of using the WISC-III for children with autism. In contrast, a high proportion of comments referred to the costs and difficulties. Among the difficulties cited were its lack of usefulness for planning interventions and the rigid administration procedures; see Table 6.4.2.1, examples no. 17 and 16 respectively. Other difficulties cited were related to language and culture biases in the items, the time involved in using the test, and difficulties in communicating information about IQ scores to parents and teachers; see Table 6.4.2.1, examples nos. 15, 13 and 12 respectively. This suggests that although there is some perceived utility, the difficulties of using the WISC-III for children with autism outweigh its possible benefits.

### *Incremental validity*

The FDGs on the WISC-III did not elicit any comments that could be coded as reflecting incremental validity, i.e. the extent to which the use of the WISC-III would improve current procedures for the assessment of SEN (See Table 6.4.1).

The code 'Improvement in Current Practice', is defined as follows:

Table 6.5.4.2 Operational Definition for 'Improvement to Current Practice'

<b>Code</b>	<b>Improvement to Current Practice</b>
<b>1. Description</b>	Comments expressing the view that the use of the test could improve current assessment practices for children with autism, i.e. which are not currently in practice.
<b>2. Inclusion</b>	The ways in which the test could introduce new insights/dimensions of functioning that are ignored/overlooked in current assessment practices., e.g. use of test in improving (expanding) focus of assessment; highlighting aspects of children's functioning that is currently overlooked in assessments.
<b>3. Exclusion</b>	Usefulness of test information for planning intervention, assessment and features of the test that are beneficial for children with autism (Consider alternative codes: Utility for Intervention; Utility for Assessment; Benefits).

It could be noted that based on the definition (See Table 6.4.2.2) a unit is coded as indicating 'Improvement In Practice' only if it related to examples of test use which extended the scope of current practice, i.e. introduced new insights/dimensions for assessment. As the WISC-III has been extensively used in Singapore, much of the comments concerning current practice were related to the costs and difficulties of using of the test.

### 6.4.3 Treatment validity of the measure for cognitive modifiability (CMB Analogies Subtest)

Table 6.4.3 shows the categories elicited for the treatment validity of the CMB and the extent to which these categories dominated the views of the participants in the FGDs.

Table 6.4.3 Summary of Categories Elicited for CMB

<b>Big Ideas</b>	<b>Categories &amp; Sub-categories</b>	<b>Dominance</b>	
<b>Treatment Utility</b>	<b>Utility for Assessment</b>	<b>30</b>	<b>22%</b>
	Assess Teaching Input/Environment	6	4%
	Assess Gen. Cognitive Ability	6	4%
	Assess Qualitative Responses	8	6%
	Assess Social Skills	2	1%
	Assess Specific Cognitive Skill	8	6%
	<b>Utility for Intervention</b>	<b>13</b>	<b>9%</b>
	Identify Level Of Support	5	4%
	Plan Academic Intervention	8	6%
	<b>Utility : Other Uses</b>	<b>3</b>	<b>2%</b>
<b>Incremental validity</b>	Diagnosis	1	1%
	Referral	2	1%
<b>Cost-benefit analysis</b>	<b>Improvement In Current Assessment Practice</b>	<b>2</b>	<b>1%</b>
<b>Costs/Difficulties</b>	<b>Benefits</b>	<b>19</b>	<b>14%</b>
	Communicate Info To Others	8	6%
	Activities/Materials	4	3%
	Language/Culture	3	2%
	Procedures	4	3%
	<b>Costs/Difficulties</b>	<b>71</b>	<b>51%</b>
	Time And Training	23	17%
	Activities/Materials	9	7%
	Language/Culture	4	3%
	Procedures	18	13%
<b>Total Comments/Units</b>		<b>138</b>	<b>100%</b>

#### *Treatment utility*

In terms of treatment utility, the dominant comments related to the CMB's utility for assessment; as shown in Table 6.4.3. Within this category heading, participants' comments related to the utility of the CMB for assessing specific cognitive skills, children's qualitative responses in learning situations, and general cognitive ability/profile; see Table 6.4.3.1, examples nos. 5, 3, and 2

respectively. The CMB elicited many comments pertaining to its utility for assessing the quality of teaching inputs/environments; see Table 6.4.3.1, example no. 1.

**Table 6.4.3.1 Examples of Coded Units (Comments) for the CMB**

Coding categories	Examples of units
<b>Utility for Assessment</b>	
Assess teaching inputs/environments	1. It's not just whether the child can learn or not. Also the type of teaching that the child responds to.
Assess general cognitive ability/profile	2. I like the fact that this test can tell you the child's maximum potential, after scaffolding.
Assess child's qualitative responses	3. The child's strategies when learning - how does he approach the task, e.g. does he play around with the blocks, does he see different options, is he impulsive? Is he flexible when you teach him the answers? These kinds of behaviours can come out in the testing.
Assess social function	4. (You can use the test to see whether) ..the child's behaviour gets more difficult as the items get more difficult. So you can link (the child's) thought processes with what you observe in his behaviour.
Assess specific cognitive skill	5. From what I've heard about this test, it seems that you can also find out about the child's flexibility when solving problems. For example, when the dimension changes, how fast can a child learn to apply changes in colour, size, position etc. How fast can they switch?
<b>Utility for intervention</b>	
Identify level of support	6. If the child needs more than one session to show modifiability, then this highlights that there is a need to either work on earlier (pre-requisite) levels, or increase the time available for support.
Plan academic intervention	7. I think our current math curriculum and method of teaching math incorporates much of this (analogical thinking) already, (and there is a high) focus on problem solving skill. Teachers are supposed to mediate and provide the intervention needed - if they can't do this, then (this test can tell us) what aid or help do they (the children) need? So this will fit in very nicely with the math curriculum here.
Improvement to current practice	8. You also teach the child in this (test). At the start he may not know about colours or size, and in the teaching, you are suppose to try and teach the child these concepts. So, at the end of it, the child leaves the testing (session) with some learning. So there is opportunity for the child to experience some level of success and learning (that is not present in other types of assessments).
<b>Other uses</b>	
Diagnosis	9. (Expressing agreement) The interaction pattern (between tester and testee) can also add more info to the diagnostic formulation.
Referral to special school	10. But one way, perhaps is to see which child needs more support or mediation and (use that info) to gauge his ability to cope with mainstream. For example, if the child needs a lot of individualised probing and prompting, then he may be suited to academic settings that matches that level of need.
<b>Benefits</b>	

Communicate info to others	11. For example, at one level, if a child is not doing well, and the modifiability is high, then what recommendations do we give (to parents)? Here (in Singapore) parents only think of tuition as the main (academic) intervention. Then (one use of the test would be ) the straight advice we can give to parents, (which would be) focused on the natural mediation strategies.
Activities & materials	12. I am thinking of the possibility of using this (test) for the screening procedure (in my special school). Because currently there is a lot of paper and pencil work, and (a heavy emphasis on) tasks to see how ready they are to go into a classroom. Some of the things that we do, are for example, sums like 3+5. So if the child does not know how to do this, we will try and teach the child. And then, if the child can be taught this simple addition concept, we can get an idea of the child's potential for learning. So probably for younger child, or a child with ASD, we don't want to do '3+5', but something else that's more abstract (like this).
Language & culture	13. But this non-verbal, so less language issue, and (hence) it's better for ASD children.
Procedures	14. Compared to other tests, where we've said that (other tests) are inflexible, and so can't really show what the child can do. So this (flexibility in administration) is a plus point in this test.
<b>Costs</b>	
Time & training	15. I think time is the main concern. 5 hours for one test?! (strong agreement by others)
Activities & materials	16. It's also dependent on how good you are with these kinds of (analogical) problems. So your learning potential is mainly concerned about these types of problems, and may not necessarily relate to other (academic) stuff.
Language & culture	17. To do this (CMB Analogies) your cognitive function needs to be quite high, as well as language function.
Procedures	18. But I am not sure we can actually use it in a standard way - the way you do it (mediate) and I do it is different! With you, the child maybe more modifiable!"
Not useful for intervention	19. Its sounds alluring at first, but...? (expressing doubt/reservations) But I have reservations about the instrument itself, whether it can do what it supposed to do (i.e. measure potential). Can it really show the difference between the child who needs more or less support? 20. A big issue for me is how much of this applies to academic work. (Because) I work with kids in mainstream, many are able, so they are beyond (understanding) colours, size etc. So, if I want to share this with teachers, they want to know if the strategies for this can be used in Math, Science etc.

### *Cost-benefit analysis*

As shown in Table 6.4.3, the most frequently expressed difficulty in using the test was the length of time required to complete the test and the training time. Although time constraints were also expressed as a concern for the other tests used in the present study, comparatively, the administration time for the CMB was the longest. There were also many comments relating to difficulties concerning test procedures. Participants expressed reservations about the reliability of the test due to the fact that variability in testers' mediation skills could affect the children's performance (See Table 6.4.3.1, example no. 18). The FDGs on the CMB emitted a relatively high number of comments expressing

reservations about its utility for planning intervention; see Table 6.4.3.1, example nos. 19 and 20.

There were some favourable comments on the benefits of using the CMB for children with autism. In particular, participants highlighted the ease with which information from the CMB can be communicated to parents and teachers, the suitability of the materials for young children and those with low language abilities, and the flexibility of the administration procedures (See Table 6.4.3.1, example nos. 11 to 14).

Overall, in terms of the cost-benefit analysis for the CMB, participants highlighted the difficulties/costs of using the CMB more than its potential benefits.

There appear to be strong parallels between the views concerning the costs and benefits of DA methods expressed by the participants in the present FDGs with the UK EPs interviewed by Deutsch and Reynolds (2000). It should be noted that while the EPs in the UK study all underwent at least 3 days' training in DA, most of the participants in the present FGD had not themselves used the CMB. Despite this difference, in both studies, the key advantage of the use of DA tests expressed by both groups related to the information and advice that can be communicated/provided to teachers and parents; and the main difficulties concerned the time demands in the use of the test. At the same time, while both groups of practitioners acknowledged the advantages of the test procedures used in DA, concerns were raised regarding its subjectivity/reliability (See Table 6.4.3.2).

**Table 6.4.3.2 Perceptions on the Benefits and Difficulties of Dynamic Assessments\***

<b>Singapore Practitioners (present FGDs)</b>	<b>UK Practitioners (Deutsch &amp; Reynolds, 2000)</b>
<p><b>Benefits</b></p> <ul style="list-style-type: none"> <li>• Communicating info/advice to parents &amp; teachers</li> <li>• Activity/Materials engaging for children</li> <li>• Flexibility in test procedures</li> <li>• Able to accommodate language/cultural barriers in testing situation.</li> </ul>	<p><b>Benefits</b></p> <ul style="list-style-type: none"> <li>• Practical advice for teachers &amp; parents</li> <li>• Focuses on strengths, enhances self-esteem</li> <li>• Superior alternative to psychometric approaches</li> <li>• Provide unique, rich information abt the child</li> </ul>
<p><b>Costs</b></p> <ul style="list-style-type: none"> <li>• Time &amp; training demands</li> <li>• Procedures – lack of reliability</li> <li>• Reservations abt usefulness for school-based interventions</li> <li>• Activities/materials too narrow in focus.</li> </ul>	<p><b>Costs</b></p> <ul style="list-style-type: none"> <li>• Time demands</li> <li>• Difficulty relating to classroom practices</li> <li>• LEA preference for standardized tests</li> <li>• Subjectivity in test interpretation</li> </ul>

*\*Note: The items are ranked in order of dominance/frequency of the views expressed*

The concerns regarding time demands and subjectivity that resonated strongly across the two groups of practitioners suggest that these are problematic issues that may be inherent in the DA methodology, and not easily overcome by more training, as suggested/implied by Deutsch and Reynolds (2000) (See discussions in Section 6.2.2).

#### *Incremental validity*

There were very few comments in the category coding for ‘improvement to current practice’; see Table 6.4.3. These tended to focus on the social consequence of using the test with a child, e.g. a child leaving the testing session having gained some learning experiences during the teaching phase. It could be argued however, that in such cases the focus of use was on the CMB as an intervention tool, rather than for measurement/assessment purposes. There were no comments on the extent to which the use of the CMB would improve the process of SEN assessment, per se.

#### **6.4.4 Treatment validity of the measure for executive function (BADS-C)**

Table 6.4.4 shows the categories for treatment validity that was elicited for the BADS-C.

##### *Treatment utility*

In terms of its treatment utility, the comments regarding the utility of the BADS-C for planning interventions, especially academic intervention, dominated the FGDs; see Table 6.4.4. In the definition of the category heading ‘planning academic intervention’, the development of pre-requisite skills that are needed for academic achievement are included. This could be, for example, problem solving strategies that are relevant across all curriculum areas, such as planning skills needed to complete a practical project (e.g. science project) and special exam accommodations. The prevalence of comments for this category reflects the participants’ view that the information on executive function that could be derived from the BADS-C is useful for identifying the kinds of provisions that children with autism would need in order to cope with academic demands and curriculum activities in school.

Table 6.4.4 Summary of Categories Elicited for the BADS-C

<b>Big Ideas</b>	<b>Categories &amp; Sub-categories</b>	<b>Dominance</b>	
<b>Treatment Utility</b>	<b>Utility for Assessment</b>	<b>28</b>	<b>19%</b>
	Assess Teaching Input/Environment	3	2%
	Assess Qualitative Responses	2	1%
	Assess Social Skills	3	2%
	Assess Specific Cognitive Skill	20	14%
	<b>Utility for Intervention</b>	<b>49</b>	<b>33%</b>
	Evaluate Intervention	7	5%
	Identify General Strategy	10	7%
	Identify Level Of Support	3	2%
	Plan Academic Intervention	21	13%
<b>Utility : Other Uses</b>	Plan Social/Adaptive Prog.	8	5%
	<b>9</b>	<b>6%</b>	
	Diagnosis	6	4%
<b>Incremental validity</b>	Referral	3	2%
	<b>Improvement in Current Assessment Practice</b>	<b>3</b>	<b>2%</b>
<b>Cost-benefit analysis</b>	<b>Benefits</b>	<b>17</b>	<b>11%</b>
	Communicate Info To Others	8	5%
	Activities/Materials	1	1%
	Language/Culture	1	1%
	Procedures	7	5%
	<b>Costs/Difficulties</b>	<b>42</b>	<b>28%</b>
	Communicating Info To Others	11	7%
	Time And Training	16	11%
	Activities/Materials	6	4%
	Language/Culture	5	3%
<b>Total Comments/Units</b>		<b>148</b>	<b>100%</b>

Another dominant category related to the utility of the BADS-C for assessment purposes, in particular for assessing specific cognitive skills (See Table 6.4.4.1, example nos. 8 and 4 respectively). Since the intended focus of the BADS-C is the assessment of a group of cognitive skills known as 'executive function', it is somewhat expected that this aspect of cognitive assessment would emerge strongly in participants' comments concerning its utility.

Despite the fact that the BADS-C was relatively new to the participants, there appears to be a fairly positive acceptance of the 'face-validity' of the test for treatment planning. In the literature, the treatment of executive functions have focused on the rehabilitation of specific cognitive skills. However, the FGD

participants identified its utility for development of skills which were seen as pre-requisite for academic activities, e.g. planning skills; project work; exams skills. This emphasis may due to the fact that recently in Singapore, practical project work has been included as of one of the compulsory components in high-stakes examinations in schools.

Table 6.4.4.1: Examples of Coded Units (Comments) for the BADS-C

Coding categories	Examples of units
<b>Utility for Assessment</b>	
Assess teaching inputs	1. I think it (BADS-C) would be helpful in helping to determine how much structure in the environment is needed for the child to cope,
Assess child's qualitative responses	2. (What's useful from this test?) Info such as the planning time, the total time taken by the child (are useful). (As well as info on) self-correction - its important to note that the child is able (flexible) enough to try again in the same way or (has he) changed, and have these reported in the score sheet.
Assess social function	3. Some teachers may tend to see these problems (in EF) as the child being naughty or difficult or oppositional, so this info maybe useful to help them better understand the child's behaviour.
Assess specific cognitive skill	4. We infer their planning skills when we do the Block Design (in the WISC-III). With this (BADS-C), we are much more certain. We have clearer view of their levels (of planning skills).
<b>Utility for intervention</b>	
Evaluate intervention	5. (It could be useful) as a pre-measure of the things that you are planning to do (with the child). As a target for the specific cognitive processes for e.g.. for inhibition, training in strategies to improve inhibition. What do you think? For e.g. before starting a behaviour mod. programme (we could use the BADS-C) as a pre-measure that will give us information on some of the issues.
Identify general strategy	6. For our school, because we use structured teaching and (focus on) organizational skills as part of the curriculum, this could be info to say that 'Hei, his (EF skills) are really bad, and so, you need some more structured (approach)'.
Identify level of support	7. Yes, (its could be useful to determine) how much more structure you need (for a child) and (alternatively) how much less you need.
Plan academic intervention	8. Possibly, because one of the exam accommodations for national exams is providing one to one invigilation. So there are children with ASD, for e.g. who may not know how to turn pages. The invigilator can give a visual prompt for the child to move on. This (test) would provide stronger data (to support the request for one-to-one invigilation).
Plan social intervention	9. This would be useful for secondary aged pupils. They get into trouble in sec. school because these (difficulties) becomes more of a social issue (at that age). (Partly) because they have more project groups, where planning skills becomes important.
Improvement to current practice	10. Very often we just see the child as 'autistic' or 'not autistic', then (consider only) what's the IQ? ((laughter)). This info on EF helps us see the child in a less rigid, narrow way.
<b>Other uses</b>	

Diagnosis	11. Firstly for diagnosis: for example for flexibility - a lot of times (to test for flexibility) we have to create situations so that the child has to react in different ways. So this (test) gives a direct way of assessing that aspect, and compare it with the normal population. So this has good value for diagnosis.
Referral to special school	12. "A lot of ASD kids do well in mainstream in primary, but when they go to secondary they fail - not because they are not good in academics, but the teacher has group projects and they can't do that part, and (when that happens) the teacher wants to send them immediately to other schools. A lot of this (EF skills) gets in the way as they grow older.
<b>Benefits</b>	
Communicate info to others	13. It's also good for having something concrete to share with parents. To highlight things that they need to do, and these are the gaps that the child has. Because they (parents) often see problem solving skills as (simply) 'Can you solve problems or not?' but not knowing actually how to teach their kids (and) how to address the gaps. So this (BADS-C) has some promise as a tool for us to communicate with parents.
Activities & materials	14. The nature of the tasks, more hands-on. So it a bit more natural. Also more interesting.
Language & culture	15. On one hand, it is good to include information and instructions so that the child understands (the task).
Procedures	16. The observation reported in the test scoring sheet - for me, is a big plus.
<b>Costs</b>	
Communicate to others	17. The key (issue) is that presently, when you say 'executive function' to teachers, they really don't understand. And parents, too.
Time & training	18. For the clinic setting, another issue is how much time is available for this test. Is it going to add to cost? It is expensive? Whether the info will actually be used?
Activities & materials	19. (Difficulties) in terms of the materials. In the (Six Part test), some of the things are so tiny (e.g. sorting tasks). So our younger kids may not be able to even pick up such things. So I think it's quite unsuitable for our younger kids.
Language & culture	20. But a lot of it is very verbal. I wonder how they (children with autism) would respond?
Not useful for intervention	21. This EF skills (to teachers), is very vague because it's all over - problem solving applies to all aspects of our lives; all situations require problems solving skills. So this information has no clear lines to let the school know if the child is suitable for this (particular) special school.

### *Cost-benefit analysis*

In terms of the 'costs' of using the BADS-C, comments regarding the time constraints in using the test dominated the FGDs (See Table 6.4.4). As the interpretation of children's scores in the BADS-C was based on IQ levels, participants expressed the concern that the test can only be used in addition to a full IQ test, hence extending the length of time needed for assessment purposes.

In terms of communicating the test information to parents and teachers, the FGDs on the BADS-C emitted fairly similar proportions of positive and negative comments (see Table 6.4.4.1, example nos. 13 and 17 respectively). As the concept of executive function is relatively new, participants expressed

reservations about schools' and teachers' readiness to address deficits in children's executive function (See Table 6.4.4.1; example no. 21).

The FGDs on BADS-C emitted favourable comments related to the benefits of using the test for children with autism, in particular the relevance of information on executive function for parents and teachers, and the administration procedures that included specific guides for testers to observe and record behaviours related to executive function (See Table 6.4.2.1; example nos. 13 and 16, respectively).

#### *Incremental validity*

There were only two comments related to 'improvement to current practice' in which participants highlighted the fact that BADS-C could be a useful addition to the IQ test. This could be seen to indicate that the BADS-C would be a useful enhancement to current practice only when used in conjunction with WISC-III. However, on its own, the BADS-C was not reported as providing a new or novel insight into the functioning of children with autism. Participants also expressed the view that the information on executive function may not be 'unique', in that current assessments methods can also provide similar information, e.g. by observations or inference (See Table 6.4.4, example no. 4).

#### **6.4.5 Treatment validity of the ToM Battery**

Table 6.4.5 shows the categories of comments related to the treatment validity of the ToM Battery.

#### *Treatment utility*

In terms of treatment utility, the most dominant comments in this category related to the utility of the ToM for planning interventions, in particular planning social intervention; see Table 6.4.5.1, example no. 10; and its utility for assessing children's cognitive function, in particular the ability for perspective taking; see Table 6.4.5.1, example no. 5. This may be seen to reflect participants' acceptance of the link between mentalising abilities, perspective taking and the development of social skills, which is consistent with the theory of mind hypothesis of autism.

Table 6.4.5 Summary of Categories Elicited for ToM

<b>Big Ideas</b>	<b>Categories &amp; Sub-categories</b>	<b>Dominance</b>	
<b>Treatment Utility</b>	<b>Utility for Assessment</b>	<b>35</b>	<b>21%</b>
	Assess Gen. Cognitive Ability	1	1%
	Assess Language Function	4	2%
	Assess Qualitative Responses	1	1%
	Assess Social Skills	10	6%
	Assess Specific Cognitive Skill	19	11%
	<b>Utility for Intervention</b>	<b>52</b>	<b>31%</b>
	Evaluate Intervention	1	1%
	Identify General Strategy	3	2%
	Identify Level Of Support	7	4%
<b>Utility : Other Uses</b>	Plan Academic Intervention	9	6%
	Plan Social/Adaptive Prog.	32	19%
	<b>12</b>	<b>7%</b>	
	Diagnosis	7	4%
<b>Incremental validity</b>	Legal Purposes	2	1%
	Referral	3	2%
	<b>11</b>	<b>7%</b>	
<b>Cost-benefit analysis</b>	<b>Benefits</b>	<b>27</b>	<b>16%</b>
	Communicate Info To Others	10	6%
	Activities/Materials	7	4%
	Language/Culture	7	4%
	Procedures	3	2%
	<b>Costs/Difficulties</b>	<b>30</b>	<b>18%</b>
	Communicating Info To Others	2	1%
	Time And Training	4	2%
	Activities/Materials	9	5%
	Language/Culture	12	7%
<b>Total Comments/Units</b>	Procedures	1	1%
	Not Useful For Intervention	2	1%
<b>167</b>		<b>100%</b>	

As described in Section 2.4.2.1, the theory of mind explanation of autism suggests that individuals with autism lack the ability to think about thoughts (or mentalising abilities) and are therefore specifically impaired in social interaction skills in which the ability for mentalising is required (Baron-Cohen et al., 1985 & 2000; Leslie & Frith, 1988; Frith, 2003).

**Table 6.4.5.1 Examples of Coded Units (Comments) for the ToM**

Coding categories	Examples of units
<b>Utility for Assessment</b>	
Assess general cognitive ability/profile	1. Yes. It ties in well with all the other info from the other tests, i.e. WISC , and helps to better understand the child's general ability.
Assess language	2. The instructions in the test get increasingly longer, so it would let us know if the child has (adequate) verbal reasoning.
Assess child's qualitative responses	3. "More on the questions and responses that the child shows during testing. For example, some kids when given ToM tasks, are very intrigued to know 'Why he said that', or Why did he do that', and they often ask you (the tester) back these questions. It's not the direct information score from the test, but more of the qualitative questioning that may surface during testing.
Assess social function	4. This (test) really highlight the difficulties (in social skills) which maybe very marked (in ASD kids).
Assess specific cognitive skill	5. You can (use it to) check the child's ability to take another person's perspective.
<b>Utility for intervention</b>	
Evaluate intervention	6. We can then monitor (children's) progress not just in terms of language, but also ToM levels, and plan for interventions appropriately.
Identify general strategy	7. Now (from the test) that we know the child's level of mentalising ability, this can help us - therapists - to know which strategy we should use and which ones we should stay away from.
Identify level of support	8. So the ToM battery can help you decide what level the child is and help you decide where you are moving the child towards.
Plan academic intervention	9. In the higher level (older age) it would be useful for teachers especially teaching literature, religion, moral education - (all of which) need perspective taking.
Plan social intervention	10. It (the test info) can help to guide me on how I would use the knowledge about the child's grasp of other's perspective to determine the approach that I would use for social skills training.
Improvement to current practice	11. It can shift the focus (of assessment) a little bit more on perspective taking issues. I think we tend to ignore (this aspect). ...at the moment it (assessment) is focused more on language and social interaction issues. (ToM) is an important aspect that we don't usually include in our assessments.
<b>Other uses</b>	
Diagnosis	12. So I find this aspect of ToM very interesting, and I've read research that show that at 3 years, children learn (develop) ToM skills, and even Down Syndrome kids can do it. But not autistic kids. It might be a good way to distinguish Asperger's kids from normal kids, because their verbal ability is there, but it's specifically a ToM deficit. So this can help identify Asperger's kids.
Legal purposes	13. For legal and NS purposes, this (tool) can be useful. (In such cases) the medical officer's (conducting the assessment) priority is to firstly establish whether the incident is malingering or is it a real deficit. So this test can provide good data for that purpose.
Referral to special school	14. Even for special school (placements), this can be used to assess a child's readiness for integrating into a larger group (mainstream),
<b>Benefits</b>	

Communicate info to others	15. Information on the degree of mentalising impairment is very useful for us to then explain to teachers and parents why the child think about certain things in a particular way.
Activities & materials	16. But in this test, the plus factor is that the verbal instruction is accompanied with verbal cues (the prompts) and concrete items, so it's much easier for kids with ASD to understand. (agreement by others).
Language & culture	17. But it's about the (applicability of) context. Because we know that the 'basement story' will never be relevant to our children, so we never really use it. Whereas this, it's more localised, and the motivation (to use it) is there.
Procedures	18. ...I think the best thing about this test is that at any time you can pause, and check their understanding. There is the possibility to do that, compared to WISC-III where you can't do that, i.e. stop and check understanding.
<b>Costs</b>	
Communicate to others	19. I wonder how parents react when you tell them your child has no ToM. So perhaps we need more info to help the teachers interpret the information.
Time & training	20. It would add-on more time to the assessment (process) as a whole. We would still need to do the WISC and get IQ. This has to be an add-on.
Activities & materials	21. There maybe also gender differences. For the guys, they may like the army stories involving double bluff, but when you give that to an Asperger's girl, she may not get it. The lack of motivation (on the subject matter) can also affect the child's responses.
Language & culture	22. Before you administer this, you would want to look at their level of understanding of verbal instructions (agreement by others). If the verbal comprehension skills are not there - you wouldn't carry on with the testing.
Procedures	23. Yes, (having discontinuation criteria) would be good (addition). Because this (test involves) is social and high language load. So for our ASD kids (who are impaired in these skills), if they can't do it and we make them do all items, they will get frustrated!
Not useful for intervention	24. The kids themselves are very different in terms of general abilities. In my school, maybe none of them can pass the ToM tasks, so if we select kids for social skills training based on this, then none will be no participants for the social skills training (laughter). So, I wonder, how useful is this in terms of differentiating (overall) functioning level?

### *Cost-benefit analysis*

Participants highlighted the potential benefit of sharing information on theory of mind with parents and teachers (see Table 6.4.5, example no. 15), and the suitability of the materials for children with autism, and its culturally appropriate contents; see Table 6.4.5, example no. 16 and 17. Although culture bias is potentially an issue for all psychological and educational tests (APA, AERA & NCME, 1999), it could be argued that in constructs that are related to social skills, the issue of cultural relevance is of greater significance. This is because, compared with other developmental skills, social behaviour is arguably more strongly influenced by cultural norms and expectations. Consequently, the appropriateness of the contents of the ToM Battery, which was adapted to suit

the cultural and social context of Singapore, emerged as one of the key benefits of the test.

Difficulties in using the ToM were also expressed, especially concerns related to the language demands that may be present in the test items; see Table 6.4.5, example no. 22. It was felt that the language demand would limit the validity of the test with children of younger ages or those with lower language abilities. In this respect, participants' views were consistent with research findings that have indicated a strong relationship between theory of mind and language abilities (Fisher et al., 2005). However, the overall responses from the FGD suggest that the benefits of using the test were seen to outweigh the potential difficulties.

#### *Incremental validity*

FDGs on the ToM emitted many comments relating to incremental validity (See Table 6.4.5; and Table 6.4.5.1, example no. 11). Participants expressed the view that the information from the ToM Battery addressed a dimension of functioning that has been frequently ignored or overlooked in current assessment practices, i.e. perspective taking. It was noted that this dimension is especially important for children with autism. As discussed in Section 6.2.4, compared with executive function, behaviours related to ToM are more clearly included in diagnostic instruments for autism, e.g. DISCO. The DISCO is widely known among EPs in Singapore, and perhaps the familiarity with the concept and its link to autism behaviours may have influenced the perceived utility.

## **6.5 DISCUSSION**

### **6.5.1 Summary of FGDs**

Overall, there appears to be little evidence for the treatment validity of the measure of intelligence, i.e. WISC-III: participants' views reflected low treatment utility, low incremental validity, and high costs in relation to the potential benefits of using the test for children with autism. This perception seems to echo a commonly expressed view in the literature on the debate concerning the limited functional utility of IQ tests for children with special educational needs (e.g. Flanagan et al., 1997).

For the measure of cognitive modifiability, i.e. the CMB, participants' responses reflect ambivalence and a lack of clarity: on the one hand, some utility and benefits of the test for assessment purposes were recognised. On the other hand, participants also expressed reservations about the reliability of the assessment procedures, and its utility for planning interventions that are directly related to academic outcomes. As discussed, there appear to be strong parallels between the views concerning the cost and benefits of DA methods expressed by the participants in the present FDGs with the UK EPs interviewed by Deutsch and Reynolds (2000). In both studies, the key advantage of the use of DA tests expressed by both groups related to the information and advice that can be communicated/provided to teachers and parents. However, the concerns regarding time demands and subjectivity of test interpretation that resonated strongly across the two groups of practitioners suggest that these are problematic issues that may be inherent in the DA methodology, and not easily overcome by more training, as suggested/implied by Deutsch and Reynolds (2000).

Participants' views on the treatment validity of the measure of executive function, i.e. the BADS-C were positive: the treatment utility was felt to be strong, especially for assessment and planning of academic skills; and the cost-benefit analysis highlighted more benefits than potential difficulties. There appears to be a fairly positive acceptance of the 'face-validity' of the test for treatment planning. Evidence for incremental validity, however, was rather minimal. This may be because information relating to children's planning abilities in academic settings, e.g. skills in the implementation of projects and exam-taking skills, are available through other sources, such as teachers' reports or direct observations.

Participants' comments on the treatment validity of the ToM Battery were strongly positive: the ToM Battery indicated high treatment utility for the assessment and planning of social skills, and the benefits of using the test outweigh potential difficulties. It also showed the strongest indication of incremental validity, i.e. it is seen to provide 'unique' information about children with autism that is not easily accessible through other methods which are currently used in practice. This perception is consistent with the widely held

view that a deficit in theory of mind abilities may be unique among children with autism (Baron-Cohen et al., 1985 & 2000).

### 6.5.2 Comparison of the treatment validity of the four tests

As discussed in section 6.2, analyses of treatment validity in this study are based on the three components of treatment validity as identified as Gresham (2002), namely treatment utility, cost-benefit analysis and incremental validity.

Integrating the results from all the FGDs (see Table 6.5.2), the practitioners' views suggest that the measure for theory of mind ability (i.e. the ToM Battery) appears to show the strongest evidence for treatment validity, followed by the measure for executive function (i.e. BADS-C). The evidence of treatment validity for the measures for cognitive modifiability (i.e. CMB) and intelligence (WISC-III) were both weak, but the measure of intelligence appears to be the weakest. The conclusions regarding the comparative treatment validity of the four tests will be used in the follow-up analyses for Study 2 (reported in Chapter 7; see section 7.1)

Table 6.5.2 Comparison of the Treatment Validity of the Four Tests

Test	Components of Treatment Validity			Comparative treatment validity*
	Treatment utility	Cost-benefit analysis	Incremental validity	
1. WISC-III	Poor	Costs outweigh benefits	Nil	Weakest
2. CMB	Ambiguous	Costs outweigh benefits	Nil	Weak
3. BADS-C	High utility	Benefits outweigh costs	Minimal	Strong
4. ToM	High utility	Benefits outweigh costs	Strong	Strongest

*\*This refers to the comparison of the treatment validity of the four tests, taking into account practitioners' views based on the three components of treatment validity, i.e. treatment utility, cost-benefit analysis and incremental validity.*

### 6.5.3 Factors affecting evaluation of treatment validity

In addition specific information about the four measures, the focus group discussions highlighted some key issues concerning the evaluation of treatment

validity and its practical implications for the assessment of children with special educational needs.

As discussed in the introduction, Gallery and Hofmeister (1978) highlighted the variability in the relationship between the test and treatment as a key factor that affects the evaluation of treatment validity of psycho-educational tests. They asserted that the greater the clarity and empirical support linking the test results and the prescribed educational treatment, the stronger is the evidence for treatment validity.

Based on the specific literature for the four measures (discussed in Section 6.1), the factors that appear to influence the treatment validity of these tests are:

1. The link between test information and the criteria and processes used for diagnosis and classification;
2. The reliability and validity of descriptions of children's strength and weaknesses based on test scores;
3. The validity evidence for the specific treatment programmes.

Based on the FGD, the measures for which the strongest utility was indicated were theory of mind and executive function. Because of the specificity of the constructs, it can be argued that the link between these measures and the proposed treatments are clear and unambiguous: they are linked to treatment for theory of mind and executive functions respectively. There is currently little research evidence supporting the use of test for executive function for the education interventions for children with autism. However, there is some clarity in the link between the components skills associated with executive functions (e.g. planning) and some academic skills (e.g. planning project work, exam taking and self-monitoring skills). This clarity offer some 'face validity' in the relationship between the test and treatment, and may have contributed to the perceived utility for the treatment of children with autism.

For the measure of theory of mind, an added advantage for its perceived utility is the availability of a treatment approach (Social Stories) that focuses on the same

construct, i.e. perspective taking. This added clarity may be one of the factors that had contributed to the perceptions of high treatment validity.

## **6.6 Conclusions**

The findings of study 3 provided some support for the treatment validity of the measures of theory of mind (i.e. ToM Battery) and executive function (i.e. BADS-C) as indicators of SEN. Comparatively, the measure for theory of mind appears to show the strongest evidence of treatment validity. Practitioners were ambivalent about the treatment validity for the measure for cognitive modifiability (i.e. CMB), and this may reflect the prevailing reservations regarding the generalisability and reliability of the information obtained from dynamic assessments. The measure of intelligence (i.e. WISC-III) appears to show the weakest treatment validity as an indicator of SEN, and this view is consistent with the frequently expressed criticism concerning the lack of treatment validity of IQ tests.

The results of the FGD also seems to reflect the observations made by Gallery and Hofmeister (1978) that key factors affecting the treatment validity are the clarity in the relationship between the measure and the identified treatment, and the clarity between the proposed treatment and the actual curriculum. From the present study, measures that focused on the more specific aspects/domains of functioning, namely theory of mind and executive function, were seen to have greater utility for assessment and planning interventions. For the measure that indicated the strongest treatment validity, i.e. theory of mind, there was also the additional clarity between the identified treatment (i.e. interventions for social skills) and the availability of a curriculum which aims to develop perspective taking and social skills in children with autism.

In the next chapter, the results from this study will be integrated with the follow-up analysis for Study 2.

## CHAPTER 7

### FOLLOW-UP ANALYSES

The aim of the present thesis is to evaluate the validity of indicators of special educational needs (SEN) for children with autism in Singapore. The indicators identified were measures of intelligence, theory of mind, executive function, central coherence and cognitive modifiability. These were evaluated based on two types of evidence: psychometric, i.e. criterion validity (Study 2); and treatment validity (Study 3).

#### 7.1 FOLLOW-UP FROM RESULTS OF STUDY 2

In Study 2, two criteria were evaluated: the extent to which the indicators show good concurrence with children's levels of special educational needs, based on an independent measure of SEN, i.e. ICF; and the extent to which the indicators are accurate in distinguishing children with autism who can cope with mainstream schools, from those that need special schools. To evaluate the relative concurrence of each predictor variable (i.e. indicators of SEN) with the independent measure of SEN (ICF), it was argued that a hierarchical regression analysis was the most appropriate (See discussion in Section 5.6.2).

This method would enable the evaluation to go beyond the effects of individual predictors and investigate the combinations of indicators that would best reflect children's level of special educational needs. At the same time, the use of hierarchical regression ensures that the order in which the predictors are entered into the regression model is based on information from other studies which can provide a meaningful reflection of their relative importance. This method would also overcome the problems associated with other types of regression analyses, e.g. stepwise regression, where the order of variables is determined solely on statistical criteria, and where strong inter-correlations between predictors can cause problems in the statistical analyses (i.e. multicollinearity).

The use of hierarchical regression methods requires theoretical or research evidence that can meaningfully guide the ordering of variables in terms of their

relative importance. In the present thesis the findings from Study 3 (described in Chapter 6), which investigated the treatment validity of these same variables as indicators of SEN, can be used to provide information regarding their relative importance. Given that another important purpose of SEN assessment is to plan interventions, it can be argued that one critical aspect of the relative importance of the SEN indicators is the extent to which they provide information that can be used for intervention, i.e. their treatment validity.

The findings from Study 3, which were based on a qualitative evaluation of the SEN indicators, suggest that, based on practitioners' views, the measure for theory of mind ability appears to show the strongest evidence for treatment validity, followed by the measure for executive function. The evidence for treatment validity for the measures for cognitive modifiability and intelligence were both weak, but the measure of intelligence appeared to be the weakest. (see Section 6.5.2)

The aim of the present follow-up analyses was to integrate the results from the two studies and evaluate the extent to which the indicators show good concurrence with children's level of special educational needs, based on an independent measure of SEN, i.e. the ICF.

Based on the comparative treatment validity of the indicators (see Section 6.5.2, Table 6.5.2), the order of predictors used in the analysis was as follows:

1. Theory of mind (ToM);
2. Executive function (BADS-C);
3. Cognitive Modifiability (CMB); and
4. Intelligence (FSIQ).

(Note: Central coherence was not included in this follow-up analysis because in the preliminary analysis of Study 2, children's scores on the measure of central coherence, i.e. CEFT, were not related to their SEN levels; see Section 5.6.2).

## 7.2 RESULTS

**Table 7.2 : Summary of Hierarchical Regression Analysis Based on Study 3**

Order of predictors*	R Square	Adjusted R Square	Std. Error of the Estimate	Sig. F Change
1 ToM	.553	.544	57.69	.000
2 BADS-C	.623	.608	53.49	.004
3 CMB	.628	.605	53.70	.439
4 FSIQ	.655	.625	52.27	.062

\*Model (combinations of predictors)

- 1 (Constant), ToM
- 2 (Constant), ToM, BADS-C
- 3 (Constant), ToM, BADS-C, CMB
- 4 (Constant), ToM, BADS-C, CMB, FSIQ

As shown in Table 7.2, on its own ToM accounted for a significant proportion of the variance in children's ICF scores, i.e. 55% ( $R^2=0.609, p \leq 0.01$ ). The addition of BADS-C scores produced a small, but nonetheless significant, increase in the predictive power of the model i.e. 61% ( $R^2=0.608, p \leq 0.01$ ). However, when the CMB scores were included, there was no significant increase in the  $R^2$ . When FSIQ was added to the prediction model, there was slight increase in the variance of ICF SEN scores accounted for, i.e. 66%, but this was not significant ( $R^2=0.655, p \geq 0.05$ ).

The results suggest that using treatment validity as a criterion for the prediction model, children's ICF SEN scores were best predicted by the combination of ToM and BADS-C scores.

## 7.3 DISCUSSION<sup>35</sup>

This follow-up analysis was part of a two-pronged evaluation of the criterion validity of the indicators of SEN (i.e. Study 2). The first criterion related to the concurrence between children's SEN level (based on ICF), and the scores on the different measures/indicators; the second criterion related to the ability of the

---

<sup>35</sup> As the data for the follow-up analyses was based on Study 2, some of the methodological issues discussed earlier in Chapter 5 are also applicable to the present, e.g. the limited generalisability of the results for children with autism from non-English speaking homes from Singapore.

indicators to distinguish children with autism who can cope with mainstream schools from those that require special schools.

Based on the first criterion, measures of children's theory of mind and executive function appeared to have the strongest concurrence with children SEN level: theory of mind accounted for 55% of the variance in children's SEN, while executive function accounted for a further 7% (62%).

Based on the second criterion, theory of mind provided the best discrimination between children with autism who can cope with mainstream school and those assessed as requiring special schooling (accuracy rate 83.3%). At the same time, discrimination between children in special schools for mild and moderate learning difficulties was best achieved based measures of intelligence (accuracy rate of 89.7%; see discussion in Section 5.7.2).

The results indicated that although each of the indicators were strongly correlated with children's level of SEN, the strength of criterion validity for individual indicators was not equivalent. The *combination* of specific indicators of SEN showed the greatest evidence for criterion validity. This resonates strongly with the view that the validity of decisions for special educational needs is strengthened when information from one source is augmented with data from other sources (Meyer et al., 2001; Flanagan et al., 1997). However, often these 'other' sources of information are unspecified, and the decision regarding the type and nature of additional information that could improve the validity of SEN assessment is left to practitioners' subjective judgments. Results from this follow-up analysis provided some guides regarding the specific test-based information that can improve the validity of inferences regarding the SEN of children with autism, i.e. by including measures that focused on aspects of cognition which have been linked to core impairments in autism, namely theory of mind and executive function.

The findings also reflected the fact that evaluations of validity are strongly dependent on the context of assessment: for different purposes (i.e. different criterion indicators) and for different target groups (children with mild and

severe learning difficulties) the strength of validity evidence for a measure can differ significantly. For example, for the purpose of distinguishing children with severe SEN, on its own IQ scores showed adequate criterion validity (Study 2; see Section 5.7.2).

From the results of Study 2 and this follow-up analysis, theory of mind appears to show good validity based on both criteria, i.e. concurrence with SEN level and identification of children with autism who can cope with mainstream schools. There are several factors which may underlie the validity of inferences based on theory of mind for children with autism, and these factors are discussed in turn in the next sub-sections.

### **7.3.1 Focus on specific core impairments in autism**

Firstly, unlike assessments approaches that focus on 'general abilities' (e.g. measures of intelligences), there is a theoretical link between children's theory of mind and specific impairments that are observed in children with autism, namely the theory of mind hypothesis (Baron-Cohen et al., 1985). As discussed, one of the problematic issues in use of measures of intelligence is the absence of a clear theoretical basis; the 'g' factor in intelligence tests is a psychometric construct. In contrast, there is a clear theoretical basis for assessing the theory of mind of children with autism.

### **7.3.2 Interactional analysis of SEN**

Theory of mind has a direct impact on the child's ability to succeed in his/her social environment. As discussed in Section 5.7.2, children with autism with relatively intact theory of mind are better able to process information about others' perspective and show the behaviours that are perceived as reflecting the ability to cope or comply with the demands of the mainstream schools. Seen within the interactional perspective of SEN that is used in this thesis, it can be argued that theory of mind is an important factor in determining the SEN level of children with autism, as it is an aspect of the child's functioning that affects his/her ability to participate and interact successfully in his/her social environment/contexts.

### 7.3.3 Methodological factors

The use of the test battery approach for measuring theory of mind meant that the test was able to show wide variations in children's mentalising ability. This ability is important, as SEN assessments often involve children with vast differences in terms of their severity of functioning and needs. Measures which are less able to reflect differences in performance for either the highly able children, or those with very low functioning, may show reduced validity when used as a criterion measure of children's SEN levels. As discussed, one of the methodological issues in the use of the measure for central coherence (CEFT) is its lack of sensitivity to reflect differences in children with high abilities (See Section 5.7.3.1); for the measure of executive function (BADSC-C), the absence of appropriate norms for children with IQ levels below 70 meant that the interpretation of these children's scores in the test may not be valid (See section 4.2.3.3).

In contrast, the Theory of Mind Battery included items which can be used with children with very low functioning (e.g. the Coin Hiding and first-order false-belief tasks), as well as those which can discriminate the more able children, e.g. the Strange Stories). However, more research and development work would be needed before the ToM Battery can be used as a practical tool in SEN assessments, e.g. the development of age appropriate norms for children across a wider age range.

The practical implications of the findings from this follow-up analysis, as well as those of the other studies in the thesis, will be discussed in the next chapter.

## CHAPTER 8

### CONCLUSION

The primary aim of the present thesis was to identify methods that can improve the assessment of special educational needs (SEN) for children with autism in Singapore. This involved the evaluation and comparison of validity evidence of several different measures or indicators of special educational needs. In this final chapter, the validity arguments for each of the identified measures are discussed in relation to the findings of the present thesis. Possible implications for the assessment practices in Singapore are considered, as well as implications for SEN provisions. Finally strengths and limitations of the research reported in this thesis and proposals for future research are considered.

#### 8.1 VALIDITY ARGUMENTS FOR MEASURES OF SEN

Validity arguments refer to the process of constructing and evaluating arguments for and against proposed test interpretations and uses. As discussed in Chapter 2, several aspects of validity arguments are noted (APA, AERA & NCME, 1999):

- Validity arguments relate to the interpretation of test scores, and not to the test itself; hence, it is not the test itself that is evaluated, but the interpretations and inferences made based on the test scores.
- Arguments for validity relate to the specific contexts and intended uses of tests/measures; hence the validity arguments that are established for one population/context cannot be generalised to other populations/contexts.
- Validity arguments are based on different types of evidence, namely evidence based on content; response processes; internal constructs; relations to other variables; and consequences of testing (as discussed in Section 2.2.2). In this chapter, the validity arguments for the measures of SEN are based on the evidence from the various studies conducted as part of this thesis, which are summarised in Table 8.1.

**Table 8.1. Summary of Findings**

<b>Study 1</b> (Chapter 3)	<p><b>Focus:</b> Reliability and validity of the ICF as a measure of SEN for children with autism</p> <p><b>Findings:</b></p> <ul style="list-style-type: none"> <li>• ICF showed adequate content validity;</li> <li>• Internal consistency and inter-rater reliabilities were high.</li> <li>• ICF showed adequate criterion validity in relation to a parallel measure of functioning for children with autism, namely the DISCO.</li> <li>• ICF showed adequate criterion validity in relation to its ability to discriminate between children with autism in mainstream and special schools.</li> </ul>
<b>Study 2</b> (Chapter 5 & follow-up in Chapter 7)	<p><b>Focus:</b> Criterion validity of indicators of SEN, namely measures of intelligence, theory of mind, executive function, central coherence and cognitive modifiability.</p> <p><b>Findings:</b></p> <p>In relation to an independent measure of SEN, i.e. ICF:</p> <ul style="list-style-type: none"> <li>• Measure of theory of mind showed the strongest concurrence with SEN level, accounting for 55% of the variance in children's SEN levels;</li> <li>• Measure of executive function contributed a further 7% in accounting for the variance in children's SEN levels;</li> <li>• All other measures had no significant effect, after children's theory of mind and executive function scores were accounted for.</li> </ul> <p>In relation to distinguishing children with autism who can cope with mainstream school, from those that require special schools:</p> <ul style="list-style-type: none"> <li>• Measure for theory of mind provided the best discrimination between children with autism in mainstream and those in special schools for mild learning difficulties (accuracy rate of 83.3%);</li> <li>• Measure of intelligence provided the best discrimination between children in the special schools for mild and moderate learning difficulties (accuracy rate of 89.7%);</li> <li>• Overall, the categorisation of children into the three school types, i.e. mainstream, special (mild) and special (moderate) was optimised when it was based on the combined measures of theory of mind, intelligence and cognitive modifiability.</li> </ul>
<b>Study 3</b> (Chapter 6)	<p><b>Focus:</b> Treatment validity of measures of SEN, namely measures of intelligence, theory of mind, executive function, and cognitive modifiability. This is a qualitative study, based on practitioners' views obtained from focus group discussions.</p> <p><b>Findings:</b></p> <p>Measure for theory of mind showed the strongest indication of treatment validity:</p> <ul style="list-style-type: none"> <li>• Its utility for assessment and intervention were high, in particular for use in planning interventions for social skills; cost-benefit analysis indicated that benefits outweighed potential difficulties; and compared with other measures, its incremental validity was the strongest, i.e. practitioners felt that it provided unique information about children's functioning that is not easily available or adequately emphasised in current practice.</li> </ul> <p>Measure for executive function showed strong indication for treatment validity:</p> <ul style="list-style-type: none"> <li>• Its utility for assessment and intervention were high, in particular for use in planning interventions for academic skills; cost-benefit analysis indicated that benefits outweighed potential difficulties. Incremental validity however was perceived to be poor.</li> </ul> <p>Measure for cognitive modifiability indicated weak treatment validity:</p> <ul style="list-style-type: none"> <li>• Utility for assessment and intervention was low (practitioners were unconvinced of the objectivity/reliability of the methods used and its links to school-related skills); cost-benefit analysis indicated that the cost incurred in terms of time and training outweighed the potential benefits; and incremental validity was perceived to be poor.</li> </ul> <p>Measure for intelligence indicated the weakest treatment validity:</p> <ul style="list-style-type: none"> <li>• Utility for assessment and intervention were low (the utility of IQ scores were related to external factors, i.e. the requirement imposed by special schools' admission criteria); cost-benefit analysis indicated that the difficulties strongly outweighed potential benefits; and incremental validity was perceived to be poor.</li> </ul>

### **8.1.1 Validity arguments for the use of ICF**

From the first study, the International Classification for Functioning Disability and Health, or ICF (WHO, 2001) was identified. There are several validity arguments in favour of the use of ICF in the SEN assessment of children with autism. In terms of its content, the ICF shows comprehensive coverage of the important components in children's SEN levels, including the interaction of within-child and environmental factors.

Secondly, in terms of its internal structure, information from the ICF showed adequate consistency and reliability. When used with a clearly defined and structured interview protocol, the judgements made based on the ICF showed adequate consistency across different raters. At the same time, the use of the investigator-based interview format ensured that judgements about the severity of children's impairments (e.g. extent to which they deviate from normal expectations/developmental milestones) lie with the trained practitioner; and no assumptions were made regarding parents' knowledge and understanding of complex behaviours related to children's social and communication skills.

Finally in relation to other variables, children's ICF scores were able to distinguish children with autism with severe special needs in the special schools, from those with less severe SEN in the mainstream schools; and ratings of children's functioning based on the ICF concurred with other sources of developmental assessments, namely those based on the DISCO.

On the one hand, the ICF may provide a useful structured framework for incorporating parental inputs in the evaluation of children's SEN. However, the validity of SEN interpretations based on the ICF may be limited by the fact that it is based on one source of information, namely parental reports. Its validity could be diminished in cases where the parents and practitioner do not share the same agenda in the ICF interview, e.g. some parents may be keener to reflect greater severity in their child's functioning in order to secure greater resources. Because of the inherent subjectivity and selectivity involved in parental reports and the high-stakes decisions that are involved in SEN assessments, information from the

ICF would need to be augmented with other independent sources of information. Study 2 evaluated a range of such sources.

### **8.1.2 Validity arguments for the measure of intelligence**

Findings from Study 2 indicated that the measure of intelligence provided the best discrimination between children in special schools for mild and moderate learning difficulties, with an accuracy rate of 89.7%. This suggests that measures of intelligence have adequate criterion validity for identifying the subgroup of children with autism who have the severest learning needs. However, the use of IQ scores to identify children with autism who can cope with mainstream schools showed very weak evidence of criterion validity. Other sources of information, in particular measures of theory of mind, showed stronger criterion validity for this purpose.

Results of Study 3 indicated that there appears to be little evidence for the treatment validity of measures of intelligence. Participants' views in the focus groups reflected low treatment utility, low incremental validity, and high costs in relation to the potential benefits of using IQ tests for children with autism. The feedback from practitioners indicated that the primary utility of using IQ scores in the Singapore context seems to be related to external factors, i.e. the special schools' current admission policies which require IQ scores as a condition for entry. There appears to be very little utility in the information inherent in the IQ test itself for deciding on and planning SEN provisions for children with autism.

The findings from the present thesis concur with the conclusions drawn from other empirical evaluation on the validity of IQ tests (based on studies conducted with other clinical groups, i.e. children without autism): when used for a clearly specified purpose and with specific populations, interpretation based on measures on intelligence do show some criterion validity; however, the evidence for its treatment validity is weak (Lincoln et al., 1995; Flanagan et al., 1997; Ortiz & Dynda, 2005). The findings from the present study highlighted the specific subgroup of children with autism for whom measures of intelligence demonstrate adequate criterion validity, namely children with autism who have severe learning needs.

### **8.1.3 Validity arguments for the measure of cognitive modifiability**

Based on the findings from Study 2, the measure of cognitive modifiability did not account for any significant variance in children's SEN level. At the same time, the measure for cognitive modifiability was not useful in distinguishing any specific groups of children in study, e.g. mainstream or special schools. The only significant findings relating to the measure of cognitive modifiability was that when combined with measures of theory of mind and intelligence, children's scores on the test for cognitive modifiability can slightly increase categorisation accuracy of children with autism from the three types of school placement: mainstream school, special school for children with mild learning difficulties, and special school for children with moderate learning difficulties. These findings suggest that on its own, the measure for cognitive modifiability has very little criterion validity as an indicator of SEN for children with autism. In contrast to claims by its proponents that dynamic assessments are a useful 'alternative' for IQ tests (e.g. Tzuriel & Feuerstein, 1992; Lidz, 2000; Lidz & Elliott, 2000; Tzuriel, 2001), the findings from the present thesis indicated that measures of cognitive modifiability might be useful only as additional information to descriptions of children based on IQ; its use as a replacement for information obtained from standardised IQ tests is unsubstantiated.

Another common claim by the proponents of dynamic assessments is that measures of cognitive modifiability have greater utility for intervention than standardised (or 'static') tests. However, the findings from Study 3 indicated that despite its emphasis on intervention/meditation, the measure of cognitive modifiability did not show strong evidence of treatment validity. One of the main limitations stems from the fact that practitioners were unconvinced of the reliability and objectivity of the interpretations based on the test. In addition, based on a cost-benefit analysis, the time and resources needed for the use of the test were not judged to be commensurate with the quality of additional information that was provided. Information derived from the test of cognitive modifiability was not perceived to have incremental validity: practitioners seemed to believe that the same type and quality of information is already available through other means of assessment, e.g. qualitative observations, curriculum assessments.

Commentators on the use of tests for cognitive modifiability have noted that the 'potential of such tests are not fully realised' in practice (Sternberg & Grigorenko, 2002; Elliott & Lauchlan, 1997; Deutsch & Reynolds, 2000). The findings of the present thesis indicated that despite the claims, its potential for use in high-stakes SEN assessment might remain limited if the critical issues concerning the validity of the methodology used in the measures of cognitive modifiability are not adequately addressed.

#### **8.1.4 Validity arguments for the measure of executive function**

The findings from Study 2 indicated that the measure of executive function accounted for an additional 7% in the variance of children's SEN levels (after accounting for the effects of theory of mind). However, the usefulness of the measure for executive function for identifying children with autism who can cope with mainstream schools was limited. Although children from the three school-types differed significantly in terms of their executive function scores, the measure for executive function did not improve the discrimination between children in mainstream and special (mild) schools, after the effects of theory of mind have been accounted for.

The findings from Study 3 indicated that among the practitioners who participated in the focus groups, there appears to be a fairly positive acceptance of the 'face-validity' of the measure of executive function for treatment planning. Its treatment utility was felt to be strong, especially for use in the assessment and planning of interventions for academic skills; and the cost-benefit analysis highlighted more benefits than potential difficulties. Evidence for incremental validity however, was minimal.

The measure of executive function focuses on a specific aspect of cognition which is believed to be related to a core deficit in autism, namely executive dysfunction. This specificity and link to a causal theory in autism may have contributed strongly to its perceived utility for use in assessment and treatment. However, because the information derived from measures of executive function was felt to be not 'unique', its perceived incremental validity is poor, i.e. behaviours related to children's planning and inflexibility can be obtained from

other sources of evidence, e.g. qualitative observations, school-based assessments.

### **8.1.5 Validity arguments for the measure of theory of mind**

The findings of Study 2 indicated that the measure of theory of mind accounted for the most variance in children's SEN level (55%), and it provided the best discrimination between children with autism who can cope with mainstream schools and those that require special schools. These findings suggest that interpretations of SEN of children with autism that are based on measures of theory of mind have adequate criterion validity – both in terms of predicting children's level of needs, and in identifying children with autism who are most likely to be able to cope with the demands of mainstream schools. As with other measures, the evidence of criterion validity is optimised when information from the measure of theory of mind is combined with other information, in particular those based on executive function and (for children with severe learning needs) IQ scores.

The findings of Study 3 indicated that compared with other measures examined in this thesis, theory of mind showed the strongest treatment validity for children with autism: practitioners believed that it has high utility for assessing and planning of social skills; and the benefits of using the test outweighed potential difficulties. It also showed the strongest indication of incremental validity, i.e. it was seen to provide information about children with autism that is not easily accessible through other methods that are currently used in practice.

Measures of theory of mind focus on another specific aspect of cognition that is held to be unique in children with autism, namely impairments in mentalising abilities. The clear link between the test construct, the diagnosis of autism, and the intended treatment (i.e. social intervention) may have contributed to the strong evidence for the treatment validity of measures of theory of mind.

## **8.2 METHODOLOGICAL AND EPISTEMOLOGICAL ISSUES**

Specific methodological issues relating to the individual studies conducted as part of this thesis have been discussed in the respective chapters (See Chapters 3,

5, 6 and 7). In this section, a broader analysis of the paradigmatic issues relevant to the present thesis is presented.

This thesis has been conducted primarily within a positivist paradigm, in that emphasis has been placed on data that are measurable, observable and replicable. It also holds an ideal of scientific knowledge as being value-free and independent (Hayes, 2000). However, some would argue that SEN is purely a social construction (Tomlinson, 1982) and is best examined within a phenomenological approach which emphasises understanding of events through the meaning and interpretation of the people who are directly/actively involved in them.

In this thesis, attempts were made to incorporate where appropriate, concepts and methods that are based on phenomenological approaches. For example, the conceptualisation of special educational needs is based on the interactionist paradigm of SEN, where children's special educational needs are viewed as an interaction between the strengths and weaknesses within the child, the level of support available for the child, and the appropriateness of the education provided. As discussed in Section 2.1.3 as a paradigm for the conceptualisation of special educational needs, the interactional approach has some key advantages: it presents a multi-dimensional analysis of learning and development, and posits the view that each dimension interacts with another. As such, this avoids the partisan/dichotomous view of special educational needs as being either determined by 'within-child' or 'environmental/societal' factors. It was also argued that the balance of arguments supporting the different paradigms for SEN indicated that the interactional perspective most closely reflects current knowledge about the factors underlying special education needs, i.e. that the effect of both within child and environmental/social factors are apparent.

Closely associated with the positivist paradigm is the empiricist methodology, i.e. the view that valid knowledge comes only from the kind of experiences that can be directly perceived through the senses, and where the finding of causal explanation of external events is emphasised. This can be contrasted with interpretivism, which emphasises the need to explain events through 'grasping the meaning of the individual's experience of and in the world' (Hayes, 2000,

pp7). In the various studies in this thesis, methodologies which are related to both the empiricist (Studies 1 and 2) and interpretive (Study 3) approaches were used. The different methodological approaches were chosen because they were felt to be suitable given the focus of the individual studies. Where the research questions required psychometric evidence regarding the validity of specific measures (namely in Studies 1 and 2), empiricist methods were felt to be most appropriate. However, where the emphasis was to obtain evidence relating to the social and educational consequences of using particular measures (Study 3) it was argued that the interpretive method employed in the focus group discussions was more appropriate. In addition, in the thesis, data obtained from both approaches were integrated in the follow-up analyses (Chapter 7). This incorporation of the different paradigms in research methodologies serves to strengthen the evaluation of the validity evidence for the indicators of special educational needs.

### **8.3 IMPLICATIONS FOR RESEARCH AND PRACTICE**

#### **8.3.1 Implications for the SEN assessment practice and policy in Singapore**

In Chapter 1, it was highlighted that one of the practical issues that had motivated the research reported in this thesis is the strong reliance on measures of intelligence in the assessment of special educational needs in Singapore. Underlying this practice is the assumption that interpretations of scores based on IQ tests are valid indicators of children's special educational needs. The findings of this thesis suggest that this assumption cannot be held true for all children with autism. The thesis indicated that depending on the purpose of testing and specific sub-groups of children with autism (e.g. those with mild or severe learning needs), different indicators have different validity. As the SEN provisions in Singapore develop to incorporate a wider continuum of provisions for children across a range of learning needs, it would seem important to review the current emphasis on IQ scores, and consider the use of alternative indicators that show stronger evidence of validity for specific groups of children with autism.

This thesis has sought to identify alternative measures of SEN and systematically evaluate their validity based on established standards. Based on the evaluations,

it can be concluded that for specific groups of children, namely those with very severe needs, measures of intelligence may provide information that can be used to predict SEN levels; however, the validity of the interpretation based on IQ can be enhanced by augmenting the information with data from other measures, specifically theory of mind and executive function. In making the fine-grained and arguably more difficult distinction between children with autism 'who can cope with the (demands of) regular schools' (MOE Press Release, 2004) and those who need more intensive specialised provisions, theory of mind appears to have the strongest evidence for validity.

In terms of policy implications, the findings from the present thesis highlight the need to broaden the criteria used for deciding access to specialised interventions, by including indicators of functioning that relates most closely to core impairments in autism, namely theory of mind.

### **8.3.2 Broader implications for the assessment of and research on children with autism**

Beyond its practical contribution to the improvements of SEN assessments in Singapore, the findings of this thesis offer some new insights into the assessment of children with autism in general.

Firstly, the research has provided some indication of the reliability and validity of the ICF for use with children with autism. As one of the key tools in World Health Organisation's Family of International Classification, the ICF was intended for use by all member states of the United Nation (WHO, 2001). The extent to which it can achieve the stated aims of being a common tool for use across cultures and all disability conditions rests on evidence of its validity. The validity of the ICF for individuals with physical disabilities has been established by previous research; the findings of the present thesis have provided some evidence of its validity for use with children with developmental disorders, namely autism. Future research might valuably investigate its applicability to other developmental disorders.

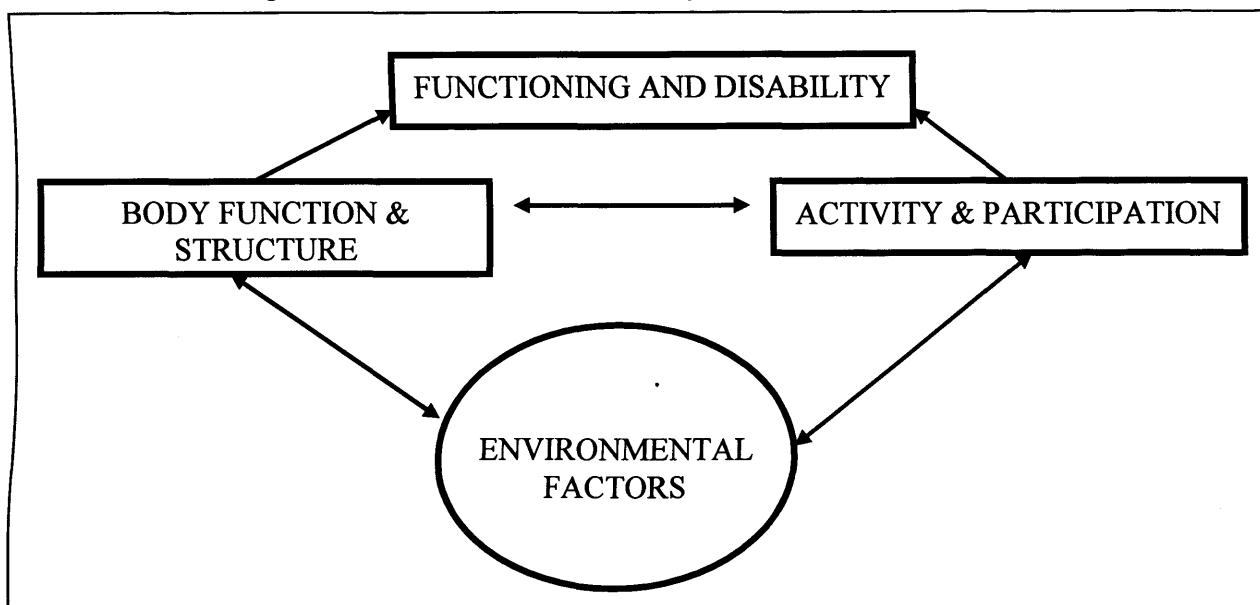
This thesis has also examined the use of dynamic assessments with children with autism. While the use of dynamic assessments (DA) with children with disabilities, e.g. Down Syndrome, hearing impairments and learning difficulties, has been studied; its use with children with autism has not been adequately examined and reported. The present study provided empirical data regarding the use of tests of cognitive modifiability for children with autism. The findings indicated that some of the claims made by the proponents of dynamic assessment methodology are unsupported; it does not appear that measures of cognitive modifiability can be simply used to replace information obtained from 'static' methods of assessments, e.g. IQ scores. More research would be needed to address concerns about the validity of dynamic assessment (DA), before it could be considered for use in high stakes SEN assessments. This issue extends beyond the applicability of DA measures for a particular population (i.e. autism), and would need to be addressed through a critical examination of the extent to which the claims regarding the utility and validity of DA methodologies can be substantiated and replicated.

The concept of and research on theory of mind has been well established within experimental psychology; and its utility as a concept for explaining the impairments in autism is well understood by practitioners. However, despite the abundance of experimental research that spanned over 20 years, there has been little work in developing a measure of theory of mind abilities for children with autism that can be used by practitioners in SEN assessments. The theory of mind battery that was developed for the present thesis and the evidence supporting its reliability and validity for use with children with autism could be a positive step in bridging this gap between research and practice in the assessment of children with autism. In order to establish more clearly the potential of the theory of mind measure as a practical assessment tool for psychologists working with children with autism, more research and development work would be needed, including the standardisation of the ToM Battery so that it can be considered for use with children across a wider age range.

#### 8.4 REVISITING THE INTERACTIONAL PERSPECTIVE ON SEN

The interactional perspective on SEN is a key conceptual framework that has guided the development of the present thesis. At the core of the concept is an understanding that children's needs and functioning are the result of an interaction of 'within-child' and environmental factors. This model of functioning and disability posits the view that the extent to which children with autism can participate adequately in life activities is directly influenced by impairments within the child (i.e. impairments in body function/structures) and environmental factors which may act as barriers or facilitators to functioning. (see Fig.8.4 below).

Fig 8.4. : ICF Model of Functioning & Disability



The indicators of SEN that were explored in this thesis, namely intelligence, theory of mind, central coherence, executive function and cognitive modifiability, represent aspects of 'Mental Functions' (i.e. one of the domains of Body Functions in the ICF model; see Table 8.4). It is therefore unsurprising that most<sup>36</sup> of these indicators were strongly correlated with children's SEN level.

Out of all the indicators investigated in this thesis, theory of mind emerged as the one that best reflects the SEN level and the ability of children with autism to

<sup>36</sup> Note: the lack of significant correlation for central coherence was probably due to the methodological limitation in the choice of instruments used to measure this construct; see discussions in section 5.7.3).

cope with mainstream schools. It could be argued that the relative prominence of theory of mind ability may be due to the fact that it is one aspect of mental functions that underpins the core impairments in autism, i.e. deficits in the ability to understand another person's perspectives or thoughts. Unlike general indicators of mental functioning, e.g. IQ and cognitive modifiability, impairments of theory of mind is unique to children with autism. Thus, while deficits in any of the other indicators would limit the child's participation in activities such as Learning and Applying Knowledge, Communication, and Self-Care (see Table 8.4); it can be hypothesized that for children with autism, impairments in theory of mind have an additional impact on the quality of Interpersonal Interactions and Relationships.

**Table 8.4 : Functional Domains in the ICF Components**

<b>BODY FUNCTION :</b>	<b>BODY STRUCTURE :</b>
<ul style="list-style-type: none"> <li>• Mental functions</li> <li>• Sensory Functions and Pain</li> <li>• Voice and Speech Functions</li> <li>• Functions of Cardiovascular, Haematological, Immunological and Respiratory Systems.</li> <li>• Functions of Digestive, Metabolic, Endocrine Systems</li> <li>• Functions of the Skin</li> </ul>	<ul style="list-style-type: none"> <li>• Structure of the Nervous System</li> <li>• The Eye, Ear and Related Structures</li> <li>• Structures Involved in Voice and Speech</li> <li>• Structure of the Cardiovascular, Immunological and Respiratory Systems</li> <li>• Structure Related to the Digestive, Metabolic and Endocrine Systems</li> <li>• Structure Related to Movement.</li> <li>• Skin and Related Structures</li> </ul>
<b>ACTIVITIES &amp; PARTICIPATION</b>	
<ul style="list-style-type: none"> <li>• Learning and Applying Knowledge</li> <li>• Communication</li> <li>• Self-Care &amp; Domestic Life</li> <li>• Interpersonal Interactions and Relationships</li> <li>• Major Life Ares</li> </ul>	
<b>ENVIRONMENTAL FACTORS (Barriers &amp; Facilitators)</b>	
<ul style="list-style-type: none"> <li>• Products and Technology</li> <li>• Natural Environment and Human-Made Changes to Environment</li> <li>• Support and Relationships</li> <li>• Attitudes</li> <li>• Services, Systems and Policies</li> </ul>	

It would be interesting to see if the significant link between theory of mind and SEN observed in this thesis holds true for children with other types of developmental disabilities, i.e. whether the findings of the thesis are unique to children with autism, or perhaps applicable to children with SEN in general.

While the present thesis has established a link between theory of mind in children with autism and their ability to cope with mainstream schools, the underlying causal factors are unexplored. Based on the ICF model, one possible

factor is the intermediatelying effect of the environment. The ICF model highlights the potential impact of environmental factors such as Products and Technology, Attitudes, Support and Relationships, and Services, Systems and Policies (See Table 8.4). Two environmental factors could be hypothesized to have a significant intermediatelying impact on the link between theory of mind and the ability of children with autism to integrate successfully with the demands of mainstream school, namely 'Attitudes, Support and Relationships', and 'Services, Systems and Policies'

### *Attitudes, Support and Relationships*

Children with autism with relatively intact theory of mind (or have ways of compensating for the deficit) would be better able to process information about others' perspective and show the behaviours that are perceived by others as reflecting the ability to cope or comply with the demands of the mainstream schools. Hence it could be argued that theory of mind has a direct impact on the attitudes of significant others (e.g. teachers, professionals, peers), specifically in terms of their willingness to support children with autism in a mainstream setting. Thus, while theory of mind affects children's ability to cope with the social demands of mainstream school, it also affects the perception and attitudes of others regarding the child's ability to succeed in mainstream environment. This perception and attitude, in turn, are potential barriers or facilitator to the child's functioning. This complex interaction between impairments in a mental function (i.e. theory of mind) and environmental factors (attitudes) is reflected in the ICF model by the use of the bi-directional causal arrows (see Fig. 8.4).

The findings of the present thesis were unable to illuminate the underlying causal patterns for the interaction between the child's theory of mind, and attitudes/perceptions of significant others, and the child's ability to cope with mainstream schools. To explore this issue, a longitudinal study could be considered to investigate the extent to which the theory of mind abilities of children with autism (at pre-school) are predictive of the attitudes and perceptions of teachers/professionals of their ability to cope with mainstream primary schools, and their long-term likelihood of being transferred to special schools.

### *Services, Systems and Policies*

From the ICF model it could be hypothesized that the strength of relationship between theory of mind abilities and the child's ability to cope with mainstream school is contingent on policies regarding the integration or inclusion of children with special needs. In the Singapore context, where the ability to cope with the demands of mainstream schools is one of the key pre-requisite for remaining in mainstream settings, there is perhaps greater clarity in the link between children who are perceived to be able to cope with mainstream schools and their theory of mind abilities. However, in other systems, where decisions regarding children's placement into mainstream settings are not contingent on his/her perceived coping skills, there may be less clarity on the extent to which theory of mind can distinguish between children with autism who can successfully integrate in mainstream settings from those that need special schooling. This issue can be investigated by comparing the link between theory of mind and SEN of children with autism across different educational systems (i.e. systems with different policies towards intergration).

## **8.5 CONCLUSION**

This thesis has established a clear link between theory of mind and the special educational needs of children with autism in Singapore. The interactionist perspective of SEN (i.e. the ICF model) has been useful in explaining the observed link between theory of mind and the special educational needs of children's with autism. It also highlighted some issues that were not addressed in the present thesis: firstly, to what extent is this link between theory of mind and SEN unique to children with autism? Secondly, how do environmental factors, such as attitudes and educational policies affect the link between the deficits in theory of mind of children with autism, and his/her ability to cope successfully in mainstream settings?

To address these issues, it would be worthwhile to replicate the studies which were conducted as part of this thesis with children with SEN with and without autism (e.g. children with other types of developmental disabilities), and in settings that uphold policies with a strong emphasis on the integration and inclusion of children with special needs, e.g. education systems in the UK, USA

or Australia. In addition, as discussed, a longitudinal predictive study on the theory of mind abilities in children with autism may uncover the intermediating effects of attitudinal and relationship factors that may act as environmental barriers/facilitators affecting the ability of children with autism to cope with mainstream schools in the long-term.

## APPENDIX A

## ICF CHECKLIST

### *Version 2.1a, Clinician Form*

## *for International Classification of Functioning, Disability and Health*

This is a checklist of major categories of the International Classification of Functioning, Disability and Health (ICF) of the World Health Organization. The ICF Checklist is a practical tool to elicit and record information on the functioning and disability of an individual. This information can be summarized for case records (for example, in clinical practice or social work). The checklist should be used along with the ICF or ICF Pocket version.

**H1.** When completing this checklist, use all information available. Please check those used:

[1] written records [2] primary respondent [3] other informants [4] direct observation

If medical and diagnostic information is not available it is suggested to complete appendix 1: Brief Health Information (p 9-10) which can be completed by the respondent.

**H2. Date**   /  /   **H3. Case ID**   , **H4. Participant No.**   , **FTC Site**     
Day Month Year CE or CS Case No. 1<sup>st</sup> or 2<sup>nd</sup> Evaluation Participant

## A. DEMOGRAPHIC INFORMATION

**A.1 NAME (optional) First \_\_\_\_\_** **FAMILY \_\_\_\_\_**

**A.2 SEX** (1)  Female (2)  Male

**A.3 DATE OF BIRTH**    /    /    (date/month/year)

#### **A.4 ADDRESS (optional)**

## **A.5 YEARS OF FORMAL EDUCATION**

**A.6 CURRENT MARITAL STATUS: (Check only one that is most applicable)**

(1) Never married [ ] (4) Divorced [ ]  
(2) Currently Married [ ] (5) Widowed [ ]  
(3) Separated [ ] (6) Cohabiting [ ]

**A.7 CURRENT OCCUPATION** *(Select the single best option)*

(1) Paid employment	[ ]	(6) Retired	[ ]
(2) Self-employed	[ ]	(7) Unemployed (health reason)	[ ]
(3) Non-paid work, such as volunteer/charity	[ ]	(8) Unemployed (other reason)	[ ]
(4) Student	[ ]	(9) Other	[ ]
(5) Keeping house/House-maker	[ ]	(please specify) _____	

**A.8 MEDICAL DIAGNOSIS of existing Main Health Conditions, if possible give ICD Codes.**

### 1. No Medical Condition exists

3. .... ICD code: \_\_\_\_\_

4. ....

ICD code: \_\_\_\_\_

5. A Health Condition (disease, disorder, injury ) exists, however its nature or diagnosis is not known

## **PART 1a: IMPAIRMENTS of BODY FUNCTIONS**

- Body functions are the physiological functions of body systems (including psychological functions).
- Impairments are problems in body function as a significant deviation or loss.

***Qualifier:***

### *Extent of impairments*

**0** No impairment, **1** Mild impairment, **2** Moderate impairment

**3 Severe impairment, 4 Complete impairment, 8 Not specified**

9 Not applicable

<i>List of Body Functions</i>	<i>Qualifier</i>
<b>b1. MENTAL FUNCTIONS</b>	
b1100 Consciousness ( <i>state</i> )	
b1101 Consciousness ( <i>continuity</i> )	
b1102 Consciousness ( <i>quality</i> )	
b1140 Orientation ( <i>time</i> )	
b1141 Orientation ( <i>place</i> )	
b1142 Orientation ( <i>person</i> )	
b117 Intellectual ( <i>incl. Retardation, dementia</i> )	
b122 Global psychosocial function	
b1260 Temperament and personality function ( <i>extraversion</i> )	
b1261 Temperament and personality function ( <i>agreeableness</i> )	
b1262 Temperament and personality function ( <i>conscientiousness</i> )	
b1263 Temperament and personality function, ( <i>psychic stability</i> )	
b1264 Temperament and personality function ( <i>openness to experience</i> )	
b1265 Temperament and personality function, ( <i>optimism</i> )	
b1266 Temperament and personality function ( <i>confidences</i> )	
b1267 Temperament and personality function ( <i>trustworthiness</i> )	
b1300 Energy and drive functions ( <i>energy</i> )	
b1301 Energy and drive functions ( <i>motivation</i> )	
b1302 Energy and drive functions ( <i>appetite</i> )	
b1303 Energy and drive functions ( <i>craving</i> )	
b1304 Energy and drive functions ( <i>impulse control</i> )	
b1340 Sleep ( <i>amount</i> )	
b1341 Sleep ( <i>onset</i> )	
b1342 Sleep ( <i>maintenance</i> )	
b1343 Sleep ( <i>quality of sleep</i> )	
b1400 Attention ( <i>sustaining</i> )	
b1401 Attention ( <i>shifting</i> )	
b1402 Attention ( <i>dividing</i> )	
b1403 Attention ( <i>sharing attention</i> )	
b144 Memory	
b1470 Psychomotor functions ( <i>control of psychomotor functions</i> )	
b1471 Psychomotor functions ( <i>quality of psychomotor functions</i> )	
b1520 Emotional functions ( <i>appropriateness of emotions</i> )	
b1521 Emotional functions ( <i>regulation of emotions</i> )	
b1522 Emotional functions ( <i>range of emotions</i> )	
b1560 Perceptual functions ( <i>auditory</i> )	
b1561 Perceptual functions ( <i>visual</i> )	
b1562 Perceptual functions ( <i>olfactory</i> )	
b1563 Perceptual functions ( <i>gustatory</i> )	
b1564 Perceptual functions ( <i>tactile</i> )	
b1565 Perceptual functions ( <i>visuospatial</i> )	
b1600 Thought functions ( <i>pace of thought</i> )	
b1601 Thought functions ( <i>form of thought</i> )	
b1602 Thought functions ( <i>content of thought</i> )	

<b>b1603</b> Thought functions ( <i>control of thought</i> )	
<b>b1640</b> Higher level cognitive functions ( <i>abstraction</i> )	
<b>b1641</b> Higher level cognitive functions ( <i>organization &amp; planning</i> )	
<b>b1642</b> Higher level cognitive functions ( <i>time management</i> )	
<b>b1643</b> Higher level cognitive functions ( <i>cognitive flexibility</i> )	
<b>b1644</b> Higher level cognitive functions ( <i>insight</i> )	
<b>b1645</b> Higher level cognitive functions ( <i>judgment</i> )	
<b>b1646</b> Higher level cognitive functions ( <i>problem-solving</i> )	
<b>b1670</b> Language ( <i>reception of language</i> )	
<b>b1671</b> Language ( <i>expression of language</i> )	
<b>b1672</b> Language ( <i>integration of language</i> )	
<b>b1720</b> Calculation functions ( <i>simple calculations</i> )	
<b>b1721</b> Calculation functions ( <i>complex calculations</i> )	
<b>b176</b> Complex movements	
<b>b180</b> Experience of self & time ( <i>incl self, body image and time</i> )	
<b>b2. SENSORY FUNCTIONS AND PAIN</b>	
<b>b210</b> Seeing	
<b>b230</b> Hearing	
<b>b235</b> Vestibular ( <i>incl. Balance functions</i> )	
<b>b250</b> Additional sensory function ( <i>taste</i> )	
<b>b250</b> Additional sensory function ( <i>smell</i> )	
<b>b250</b> Additional sensory function ( <i>proprioceptive</i> )	
<b>b250</b> Additional sensory function ( <i>touch</i> )	
<b>b250</b> Additional sensory function ( <i>sensing temperature/vibration etc</i> )	
<b>b280</b> Pain	
<b>b3. VOICE AND SPEECH FUNCTIONS</b>	
<b>b310</b> Voice	
<b>b320</b> Articulation	
<b>b330</b> Fluency and rhythm	
<b>b340</b> Alternative vocalisations	
<b>b4. FUNCTIONS OF THE CARDIOVASCULAR, HAEMATOLOGICAL, IMMUNOLOGICAL AND RESPIRATORY SYSTEMS</b>	
<b>b410</b> Heart	
<b>b430</b> Haematological ( <i>blood</i> )	
<b>b435</b> Immunological ( <i>allergies, hypersensitivity</i> )	
<b>b440</b> Respiration ( <i>breathing</i> )	
<b>b5. FUNCTIONS OF THE DIGESTIVE, METABOLIC AND ENDOCRINE SYSTEMS</b>	
<b>b515</b> Digestive	
<b>b525</b> Defecation	
<b>b555</b> Endocrine glands ( <i>hormonal changes</i> )	
<b>b6. GENITOURINARY AND REPRODUCTIVE FUNCTIONS</b>	
<b>b620</b> Urination functions	
<b>b640</b> Sexual functions	
<b>b7. NEUROMUSCULOSKELETAL AND MOVEMENT RELATED FUNCTIONS</b>	
<b>b710</b> Mobility of joint	
<b>b730</b> Muscle power	
<b>b750</b> Movements	
<b>b8. FUNCTIONS OF THE SKIN AND RELATED STRUCTURES</b>	
<b>ANY OTHER BODY FUNCTIONS</b>	
<b>b810</b> Skin	
<b>b850</b> Hair & nails	

### **Part 1 b: IMPAIRMENTS of BODY STRUCTURES**

- Body structures are anatomical parts of the body such as organs, limbs and their components.
- Impairments are problems in structure as a significant deviation or loss.

<b>First Qualifier: Extent of impairment</b>	<b>Second Qualifier: Nature of the change</b>
0 No impairment	0 No change in structure
1 Mild impairment	1 Total absence
2 Moderate impairment	2 Partial absence
3 Severe impairment	3 Additional part
4 Complete impairment	4 Aberrant dimensions
8 Not specified	5 Discontinuity
9 Not applicable	6 Deviating position
	7 Qualitative changes in structure, including accumulation of fluid
	8 Not specified
	9 Not applicable

1. NERVOUS SYSTEM	2. EYE, EAR AND RELATED STRUCTURES	3. STRUCTURES INVOLVED IN VOICE AND SPEECH
s1. STRUCTURE OF THE NERVOUS SYSTEM		
s2. THE EYE, EAR AND RELATED STRUCTURES		
s3. STRUCTURES INVOLVED IN VOICE AND SPEECH		
s4. STRUCTURE OF THE CARDIOVASCULAR, IMMUNOLOGICAL AND RESPIRATORY SYSTEMS		
s5. STRUCTURES RELATED TO THE DIGESTIVE, METABOLISM AND ENDOCRINE SYSTEMS		
s6. STRUCTURE RELATED TO GENITOURINARY AND REPRODUCTIVE SYSTEM		
s7. STRUCTURE RELATED TO MOVEMENT		
s8. SKIN AND RELATED STRUCTURES		
ANY OTHER BODY STRUCTURES		

## **PART 2: ACTIVITY LIMITATIONS & PARTICIPATION RESTRICTION**

- Activity is the execution of a task or action by an individual. Participation is involvement in a life situation.
- Activity limitations are difficulties an individual may have in executing activities. Participation restrictions are problems an individual may have in involvement in life situations.

The Performance qualifier describes what an individual does in his or her current environment. Because the current environment brings in the societal context, performance can also be understood as involvement in a life situation" or "the lived experience" of people in the actual context in which they live. This context includes the environmental factors – all aspects of the physical, social and attitudinal world that can be coded using the Environmental Factors.

Qualifier
0 No difficulty
1 Mild difficulty
2 Moderate difficulty
3 Severe difficulty
4 Complete difficulty
8 Not specified
9 Not applicable

Activity Limitations	Participation
<b>d1. LEARNING AND APPLYING KNOWLEDGE</b>	
d110 Watching	
d115 Listening	
d120 Purposeful sensing	
d130 Copying	
d135 Rehearsing	
d140 Learning to read	
d145 Learning to write	
d150 Learning to calculate (arithmetic)	
d155 Acquiring skills	
d160 Focusing attention	
d163 Thinking	
d166 Reading	
d170 Writing	
d175 Solving problems	
d177 Making decisions	
<b>d2. GENERAL TASKS AND DEMANDS</b>	
d210 Undertaking a single task	
d220 Undertaking multiple tasks	
d230 Daily routine	
d240 Stress / responsibility	
<b>d3. COMMUNICATION</b>	
d310 Communicating – receiving spoken messages	
d315 Communicating – receiving nonverbal messages	
d325 Communicating – receiving written messages	
d330 Communicating – producing spoken messages	
d335 Communicating – producing nonverbal messages	
d340 Communicating – producing written messages	
d350 Conversation	
d355 Discussion	
d360 Communication devices	
<b>d4. MOBILITY</b>	
d410 Changing position	

<b>d415</b> Maintaining position	
<b>d420</b> Transferring position	
<b>d430</b> Lifting and carrying objects	
<b>d435</b> Moving objects	
<b>d440</b> Fine hand use ( <i>picking up, grasping</i> )	
<b>d445</b> Hand and arm use	
<b>d450</b> Walking	
<b>d465</b> Moving around ( <i>in different location, using equipment e.g. wheelchair, skates, etc.</i> )	
<b>d470</b> Using transportation ( <i>car, bus, train, plane, etc.</i> )	
<b>d475</b> Driving ( <i>riding bicycle and motorbike, driving car, etc.</i> )	
<b>d5. SELF CARE</b>	
<b>d510</b> Washing oneself ( <i>bathing, drying, washing hands, etc.</i> )	
<b>d520</b> Caring for body parts ( <i>brushing teeth, shaving, grooming, etc.</i> )	
<b>d530</b> Toileting	
<b>d540</b> Dressing	
<b>d550</b> Eating	
<b>d560</b> Drinking	
<b>d570</b> Looking after one's health	
<b>d6. DOMESTIC LIFE</b>	
<b>d620</b> Acquisition of goods and services ( <i>shopping, etc.</i> )	
<b>d630</b> Preparation of meals ( <i>cooking etc.</i> )	
<b>d640</b> Doing housework ( <i>cleaning house, washing dishes laundry, ironing, etc.</i> )	
<b>d660</b> Assisting others	
<b>d7. INTERPERSONAL INTERACTIONS AND RELATIONSHIPS</b>	
<b>d7100</b> Basic interpersonal interactions ( <i>respect &amp; warmth</i> )	
<b>d7101</b> Basic interpersonal interactions ( <i>appreciation</i> )	
<b>d7102</b> Basic interpersonal interactions ( <i>tolerance</i> )	
<b>d7103</b> Basic interpersonal interactions ( <i>criticism</i> )	
<b>d7104</b> Basic interpersonal interactions ( <i>social cues</i> )	
<b>d7105</b> Basic interpersonal interactions ( <i>physical contact</i> )	
<b>d7200</b> Complex interpersonal interactions ( <i>forming relationships</i> )	
<b>d7201</b> Complex interpersonal interactions ( <i>terminating relationships</i> )	
<b>d7202</b> Complex interpersonal interactions ( <i>regulating relationships</i> )	
<b>d7203</b> Complex interpersonal interactions ( <i>maintaining social rules</i> )	
<b>d7204</b> Complex interpersonal interactions ( <i>maintaining social space</i> )	
<b>d7305</b> Relating with strangers	
<b>d740</b> Formal relationships	
<b>d750</b> Informal social relationships	
<b>d7601</b> Family relationships (child-parent)	
<b>d7602</b> Family relationships (sibling)	
<b>d7603</b> Family relationships (extended family)	
<b>d8. MAJOR LIFE AREAS</b>	
<b>d810</b> Informal education	
<b>d815</b> Preschool education	
<b>d820</b> School education	
<b>d9. COMMUNITY, SOCIAL AND CIVIC LIFE</b>	
<b>d910</b> Community Life	
<b>d920</b> Recreation and leisure	
<b>d930</b> Religion and spirituality	

### PART 3: ENVIRONMENTAL FACTORS

- Environmental factors make up the physical, social and attitudinal environment in which people live and conduct their lives.

*Qualifier in environment:*

*Barriers or facilitator*

0 No barriers  
 -1 Mild barriers  
 -2 Moderate barriers  
 -3 Severe barriers  
 -4 Complete barriers

0 No facilitator  
 +1 Mild facilitator  
 +2 Moderate facilitator  
 +3 Substantial facilitator  
 +4 Complete facilitator

<i>List of Environment</i>	<i>Qualifier</i> <i>Barriers or facilitator</i>
<b>e1. PRODUCTS AND TECHNOLOGY</b>	
e110 For personal consumption (food, medicines)	
e115 For personal use in daily living	
e120 For personal indoor and outdoor mobility and transportation	
e125 Products for communication	
e130 Products for education	
e140 Products for recreation, sport	
e145 Products for religion	
e150 Design, construction and building products and technology of buildings for public use	
e155 Design, construction and building products and technology of buildings for private use	
<b>e2. NATURAL ENVIRONMENT AND HUMAN MADE CHANGES TO ENVIRONMENT</b>	
e210 Physical geography	
e215 Population	
e220 Flora & fauna	
e225 Climate	
e230 Natural events (e.g. earthquake)	
e235 Human caused events (e.g. wars)	
e240 Light	
e245 Time related changes	
e250 Sound	
e255 Vibration	
e260 Air quality	
<b>e3. SUPPORT AND RELATIONSHIPS</b>	
e310 Immediate family	
e312 Extended family	
e320 Friends	
e325 Acquaintances, peers, colleagues, neighbours and community members	
e330 People in position of authority	
e340 Personal care providers and personal assistants	
e345 Strangers	
e355 Health professionals	
e360 Other professionals	
<b>e4. ATTITUDES</b>	
e410 Individual attitudes of immediate family members	
e410 Individual attitudes of extended family members	
e420 Individual attitudes of friends	
e410 Individual attitudes of acquaintances, peers, colleagues, neighbours and community members	
e430 Individual attitudes of people in position of authority	
e440 Individual attitudes of personal care providers and personal assistants	
e445 Individual attitudes of strangers	
e450 Individual attitudes of health professionals	
e455 Individual attitudes of other professionals	
e460 Societal attitudes	
e465 Social norms, practices and ideologies	
<b>E5. SERVICES, SYSTEMS AND POLICIES</b>	
e525 Housing services, systems and policies	
e535 Communication services, systems and policies	
e540 Transportation services, systems and policies	

e550 Legal services, systems and policies	
e570 Social security, services, systems and policies	
e575 General social support services, systems and policies	
e580 Health services, systems and policies	
e585 Education and training services, systems and policies	
e590 Labour and employment services, systems and policies	
<b>ANY OTHER ENVIRONMENTAL FACTORS</b>	

#### **PART 4: OTHER CONTEXTUAL INFORMATION**

**4.1** *Give a thumbnail sketch of the individual and any other relevant information.*

**4.2** *Include any Personal Factors as they impact on functioning (e.g. lifestyle, habits, social background, education, life events, race/ethnicity, sexual orientation and assets of the individual).*

## **BRIEF HEALTH INFORMATION**

**Self Report**

**Clinician Administered**

**X.1 Height:**   /  /   cm (or inches)

**X.2 Weight:**   /  /   kg (or pounds)

**X.3 Dominant Hand (prior to health condition):** Left  Right  Both hands equally

**X.4 How do you rate your physical health in the past month?**

Very good  Good  Moderate  Bad  Very bad

**X.5 How do you rate your mental and emotional health in the past month?**

Very good  Good  Moderate  Bad  Very bad

**X.6 Do you currently have any disease(s) or disorder(s) ?**

NO

YES

*If YES, please specify:* \_\_\_\_\_

**X.7 Did you ever have any significant injuries that had an impact on your level of functioning?**

NO

YES

*If YES, please specify* \_\_\_\_\_

**X.8 Have you been hospitalized in the last year?**

NO

YES

*If YES, please specify reason(s) and for how long?*

1. \_\_\_\_\_; \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_ days

2. \_\_\_\_\_; \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_ days

3. \_\_\_\_\_; \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_ days

**X.9 Are you taking any medication ( either prescribed or over the counter)?**

NO

YES

*If YES, please specify major medications*

1. \_\_\_\_\_

2. \_\_\_\_\_

3. \_\_\_\_\_

**X.10 Do you smoke?**

NO

YES

**X.11 Do you consume alcohol or drugs?**

NO

YES

*If YES, please specify average daily quantity*

Tobacco: \_\_\_\_\_

Alcohol: \_\_\_\_\_

Drugs: \_\_\_\_\_

**X.12** Do you use any assistive device such as glasses, hearing aid, wheelchair, etc.?

NO

YES

*If YES, please specify*

---

**X.13** Do you have any person assisting you with your self care, shopping or other daily activities?

NO

YES

*If YES, please specify person and assistance they provide*

---

**X.14** Are you receiving any kind of treatment for your health?

NO

YES

*If YES, please specify:*

---

**X.15** Additional significant information on your past and present health:

---

---

**X.16** IN THE PAST MONTH, have you cut back (i.e. reduced) your usual activities or work because of your *health condition*? (a disease, injury, emotional reasons or alcohol or drug use)

NO

YES If yes, how many days? \_\_\_\_\_

**X.17** IN THE PAST MONTH, have you been totally unable to carry out your usual activities or work because of your *health condition*? (a disease, injury, emotional reasons or alcohol or drug use)

NO

YES If yes, how many days? \_\_\_\_\_

## **GUIDELINES FOR THE USE OF ICF CHECKLIST VERSION 2.1A**

1. *This is a checklist of major categories of International Classification of Functioning, Disability and Health (ICF) of the World Health Organization. The ICF Checklist is a practical tool to elicit and record information on the functioning and disability of an individual. This information can be summarized for case records (for example, in clinical practice or social work).*
2. *This version (2.1a) is for use by a clinician, health or social care professional.*
3. *The checklist should be used along with the ICF full or short version which is scheduled for publication in September 2001. Until then the ICIDH-2 Final Draft, full version, WHO, 2001 will serve as reference document for the ICF checklist. The raters should familiarize themselves with the ICIDH-2 Final Draft by attending a brief educational programme or self-taught curriculum.*
4. *All information from written records, primary respondent, other informants and direct observation can be used to fill in the checklist. Please record all sources of information used on the first page.*
5. *Parts 1 to 3 should be filled in by writing the qualifier code against each of the function, structure, activity and participation term that shows some problem for the case being evaluated. Appropriate codes for the qualifiers are given on the relevant pages.*
6. *Comments can be made regarding any information that can serve as the additional qualifier or that is thought to be significant for the case being evaluated.*
7. *Part 4 (Environment) has both negative (barrier) and positive (facilitator) qualifier codes. For all positive qualifier codes, please use a plus (+) sign before the code.*
8. *The categories given in the checklist have been selected from the ICF and are not exhaustive. If you need to use a category that you do not find listed here, use the space at the end of each dimension to record these.*

## APPENDIX B

### DSM-IV AND ICD-10 DIAGNOSTIC CRITERIA FOR AUTISM

#### DSM-IV Criteria for Autistic Disorder<sup>37</sup>

- A. A total of six items from (1), (2) and (3), with at least two from (1), and one each from (2) and (3):
  - (1) Qualitative impairment in social interaction as manifested by at least two of the following:
    - (a) Marked impairment in the use of multiple nonverbal behaviour such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction;
    - (b) Failure to develop peer relationships appropriate to developmental level;
    - (c) Markedly impaired expression of pleasure in other people's happiness;
    - (d) Lack of social or emotional reciprocity.
  - (2) Qualitative impairments in communication as manifested by at least one of the following:
    - (a) Delay in, or total lack of, the development of spoken language (not accompanied by an attempt to compensate through alternative modes of communication such as gestures or mime);
    - (b) In individuals with adequate speech, marked impairment in the ability to initiate or sustain a conversation with others;
    - (c) Stereotyped and repetitive use of language or idiosyncratic language;
    - (d) Lack of varied spontaneous make-believe play or social imitative play appropriate to developmental level.
  - (3) Restricted repetitive and stereotyped patterns of behaviour, interest, and activities, a manifested by at least one of the following:
    - (a) Encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is normal either in intensity or focus;
    - (b) Apparently compulsive adherence to specific non-functional routines or rituals;
    - (c) Stereotyped and repetitive motor mannerisms (e.g. hand or finger flapping or twisting, or complex whole-body movements);
    - (d) Persistent preoccupation with parts of objects.
- B. Delays or abnormal functioning in at least one of the following areas with onset prior to age 3 years: (1) social interaction, (2) language as used in social communication, or (3) symbolic or imaginative play.
- C. Not better accounted for by Rett' Disorder or Childhood Disintegrative Disorder.

<sup>37</sup> Reproduced from Diagnostic and Statistical Manual of Mental Disorders, 4<sup>th</sup> Edition, (1994). Washington, DC: American Psychiatric Association, 70-71.

## ICD-10 Criteria for Autism<sup>38</sup>

### ***F84.0 Childhood Autism***

A. Abnormal or impaired development is evident before the age of 3 years in at least one of the following areas:

- (1) Receptive or expressive language is used in social communication;
- (2) The development of selective social attachments or of reciprocal social interaction;
- (3) Functional or symbolic play

B. A total of at least six symptoms from (1), (2) or (3) must be present, with at least two from (1) and at least one from each of (2) and (3):

- (1) Qualitative impairments in social interaction are manifested in at least two of the following areas:
  - (a) Failure adequately to use eye-to-eye contact, facial expression, body postures, and gestures to regulate social interaction;
  - (b) Failure to develop (in a manner appropriate to mental age, and despite ample opportunities) peer relationships that involve a mutual sharing of interests, activities, and emotions;
  - (c) Lack of socio-emotional reciprocity as shown by an impaired or deviant response to other people's emotions; or lack of modulation of behaviour according to social context; or a weak integration of social, emotional, and communicative behaviours;
  - (d) Lack of spontaneous seeking to share enjoyment, interests, or achievement with other people (e.g. a lack of showing, bringing, or pointing out to other people objects of interests to the individual).
- (2) Qualitative abnormalities communication as manifest in at least one of the following areas:
  - (a) Delay in, or total lack of, development of spoken language that is not accompanied by an attempt to compensate through the use of gestures or mime as an alternative mode of communication (often preceded by a lack of communicative babbling);
  - (b) Relative failure to initiate or sustain conversational interchange (at whatever level of language skill is preset), in which there is reciprocal responsiveness to the communications of the other persons;
  - (c) Stereotyped and repetitive use of language or idiosyncratic use of words or phrases;
  - (d) Lack of varied spontaneous make-believe play or (when young) social imitative play.
- (3) Restricted, repetitive, and stereotyped patterns of behaviour, interests, and activities are manifested in at least one of the following:
  - (a) An encompassing preoccupation with one or more stereotyped and restricted patterns of interest that are abnormal in content or focus; or one or more interests that are abnormal in their intensity and circumscribed nature though not in their content or focus;
  - (b) Apparently compulsive adherence to specific non-functional routines or rituals;
  - (c) Stereotyped and repetitive motor mannerisms that involve either hand or finger flapping or twisting, or complex whole-body movements;

<sup>38</sup> Reproduced from ICD-10 Classification of Mental and Behavioural Disorders – diagnostic criteria for research (1995). Geneva: World Health Organisation, 147-150.

- (d) Preoccupations with part-objects or non-functional elements of play materials (such as their odour, the feel of the surface, or the noise or vibration they generate).
- C. The clinical picture is not attributable to other varieties of pervasive developmental disorders; specific disorder of receptive language (F80.2); mental retardation (F70-F72) with some associated emotional or behavioural disorders; schizophrenia (F20) of unusually early onset; and Rett's Syndrome (F84.12).

***F84.1 Atypical Autism***

- A. Abnormal or impaired development is evident at or after the age of 3 years (criteria as for autism except for age of manifestation).
- B. There are qualitative abnormalities in reciprocal social interaction or in communication, or restrictive, repetitive, and stereotyped patterns of behaviour, interests and activities. (Criteria as for autism except that it is unnecessary to meet the criteria for a number of areas of abnormalities.)
- C. The disorder does not meet the criteria for autism (F84.0).

Autism may be typical in either age of onset (F84.10) or symptomatology (F84.11): the two types are differentiated with a fifth character for research purposes. Syndromes that are typical in both respects should be coded F84.12.

***F84.10 Atypicality in age of onset***

- A. The disorder does not meet the criterion A for autism (F84.0); that is abnormal or impaired development is evident only at or after age 3 years.
- B. The disorder meets criteria B ad C for autism (F84.0).

***F84.11 Atypicality in symptomatology***

- A. The disorder meets criterion A for autism (F84.0); that is, abnormal or impaired development is evident only at or after age 3 years.
- B. There are qualitative abnormalities in reciprocal social interactions or in communication, or restricted, repetitive and stereotyped patterns of behaviour, interests, and activities. Criteria as for autism except that it is unnecessary to meet the criteria for a number of areas of abnormality).
- C. The disorder meets criteria C for autism (F84.0).
- D. The disorder does not fully meet criteria B for autism (F84.0).

***F84.11 Atypicality in both age of onset and symptomatology***

- A. The disorder does not meet criterion A for autism (F84.0); that is, abnormal or impaired development is evident only at or after age 3 years.
- B. There are qualitative abnormalities in reciprocal social interactions or in communication, or restricted, repetitive and stereotyped patterns of behaviour, interests, and activities. Criteria as for autism except that it is unnecessary to meet the criteria for a number of areas of abnormality).
- C. The disorder meets criteria C for autism (F84.0).
- D. The disorder does not fully meet criteria B for autism (F84.0).

## APPENDIX C

### INVESTIGATOR-BASED INTERVIEW USED IN CONJUNCTION WITH THE ICF CHECKLIST

#### Description

This investigator-based interview protocol was developed for the present study, to be used in conjunction with the ICF checklist (see Appendix A). In this protocol, the definitions for the ICF codes (which are in grey) are extracted from the ICF handbook (WHO, 2001).

The questions and probes are developed for the present study to elicit parental report of the children's behaviour and/or functioning for each item. The rating scale (0 to 4) follows the framework of the ICF. The ICF recommends that each point in the rating scale be calibrated in different domains to population standards as percentiles. However, for items measuring the extent of impairments in young children, it can be argued that it would be more appropriate to calibrate the ratings based on the discrepancy between the child's current developmental attainments, and his/her chronological age.

---

#### b1 MENTAL FUNCTIONS

##### b110 Consciousness function

**General mental functions of the state of alertness, awareness, including clarity and continuity of wakeful state.**

*Inclusions: functions of the state of awareness and alertness, including clarity and continuity of wakeful state.*

*Exclusions: orientation function, energy & drive, sleep function.*

**b1100 State of consciousness:** Mental functions that when altered produced states such as clouding consciousness, stupor or coma.

**b1101 Continuity of consciousness:** Mental functions that produce sustained wakefulness, alertness and awareness, when disrupted, may produce trance etc

**b1108 Quality of consciousness:** Mental functions that when altered affect changes in the character of wakeful, alert and aware sentience e.g. drug altered states or delirium.

---

#### Probe & Rating Guidelines for b1100 to b1108

Has X ever lost consciousness? Does he<sup>39</sup> take any drugs or medication that affects his alertness or consciousness? If yes, how often has this problem occurred in the last month?

- 0 No impairment.
- 1 Mild impairment. Impairment present less than 25% of the time, e.g. 1-2 times a month.
- 2 Moderate impairment. Impairment present less than 50% of the time, e.g. 1-2 times a week.
- 3 Severe impairment. Impairment present more than 50% of the time, e.g. 3-4 times a week, and partially disrupting child's day-to-day life.
- 4 Complete impairment. Impairment present more than 95% of the time, e.g. daily, and totally disrupting child's day-to-day life.

---

<sup>39</sup> Purely for convenience, the pronoun 'he' is used throughout the interview protocol. During interview, preferable the child's first name is used during questioning/probing.

---

**b114 Orientation functions**

**General mental functions knowing and ascertaining one's relation to self, others, time and one's surroundings.**

*Inclusions: functions of orientations to time, place and person; orientation to self and others; disorientation to time, place & person. Exclusion: consciousness of function, attention function, memory function.*

---

**b1140 Orientation to time : Mental functions that produce awareness of day/month.**

**b1141 Orientation to place: Mental functions that produce awareness of one's location, e.g. immediate surrounding, town or country.**

**b1142 Orientation to person : Mental functions that produce awareness of own identity and individuals in immediate environment**

---

**Probe & Rating Guidelines for b1140 to 1442**

Is X aware of the day/month? Is he aware of where he is (e.g. home, school, outside)? Is he aware of the people he is with (e.g. with family, familiar people, unfamiliar people/strangers)? How often has the problem occur?

- 0 No impairment.
- 1 Mild impairment. Impairment present less than 25% of the time, e.g. 1-2 times a month.
- 2 Moderate impairment. Impairment present less than 50% of the time, e.g. 1-2 times a week.
- 3 Severe impairment. Impairment present more than 50% of the time, e.g. 3-4 times a week, and partially disrupting child's day-to-day life.
- 4 Complete impairment. Impairment present more than 95% of the time, e.g. daily, and totally disrupting child's day-to-day life.

---

**b117 Intellectual Function**

**General mental functions, required to understand & constructively integrate the various mental functions, including all cognitive functions & development over life-span.**

*Inclusion: functions of intellectual growth, retardation, dementia. Exclusion : memory function, thought function, higher-level cognitive function.*

**Probe & Rating Guidelines for b117**

Has X had IQ test in last 3 years? What was the IQ? If not done/ results unknown, code 8 (i.e. not specified).

- 0 No impairment. Average IQ or more (IQ 85 or more)
- 1 Mild impairment (IQ 70-84)
- 2 Moderate impairment (IQ 50-69)
- 3 Severe impairment (IQ 34 to 49)
- 4 Complete impairment (IQ 20-34)

---

**b122 Global psycho-social function**

**General mental functions required to understand and integrate mental functions that lead to formation of interpersonal skills needed to establish reciprocal social interaction, in terms of both meaning & purpose, e.g. such as in autism.**

**Rating Guidelines for b122**

Rate based on information from entire interview regarding level of communication and social interaction especially sections d3 Communication and d7 Interpersonal Interactions

- 1 Mild impairments. Mild impairments in communication & social interaction, eg can interact but inappropriate/unsuccessful.
- 2 Moderate impairment, e.g. interaction limited to familiar individuals only, on own terms.

3 Severe impairments. Severe communication impairments and interactions limited to requests for needs only.

4 Complete impairment. No communication/interaction. Behaviour limited to stereotyped, repetitive behaviours.

---

### b130 Energy and drive functions

**General mental functions of physiological & psychological mechanism that cause indiv to move towards satisfying needs and general goals in a persistent nature.**

**b1300 Energy level : Mental functions that produce vigor & stamina.**

**b1301 Motivation : Mental functions that produce incentive to act, conscious or unconscious force for action. NOTE: exclude cognitive /academic abilities.**

**b1302 Appetite : Mental functions that produce a natural longing or desire, esp. natural & recurring desire for food & drink.**

**b1303 Craving : Mental functions that produce the urge to consume substances, including substances that can be abused.**

**b1304 Impulse control : Mental functions that regulate & resist sudden intense urges to do something.**

---

#### Probe & Rating Guidelines for b1300 to 1304

(Energy) How is X's general energy level? Does he tire easily? (Motivation) Is X responsive when you ask him to do a task that you know he can do? (Appetite) How is X's appetite? Does he show interest in food? (Craving) Does X have any craving for a particular food/ type of food, e.g. sweets, soda? (Impulse control) Can X be told to wait?

If impairment exist in any of the areas pertaining to energy & drive functions, ask how often does the problem occur?

- 0 No impairment.
- 1 Mild impairment. Impairment present less than 25% of the time, e.g. 1-2 times a month.
- 2 Moderate impairment. Impairment present less than 50% of the time, e.g. 1-2 times a week.
- 3 Severe impairment. Impairment present more than 50% of the time, e.g. 3-4 times a week, and partially disrupting child's day-to-day life.
- 4 Complete impairment. Impairment present more than 95% of the time, e.g. daily, and totally disrupting child's day-to-day life.

---

### b126 Temperament & personality function

**General mental functions of constitutional disposition of the individual to react in a particular way to situations, including set of mental characteristics that makes the indiv. distinct from others.**

*Exclusion : intellectual function, energy & drive function, psycho-motor, emotional function.*

**b1260 Extraversion : Mental functions that produce a personal disposition that's outgoing, as contrasted to being shy, restricted, inhibited.**

**b12161 Agreeableness : Mental functions that produce a personal disposition that's being cooperative, amicable, accommodating as contrasted to being unfriendly, opposition & defiant**

**b12162 Conscientiousness : Mental functions that produce a personal disposition such as hard-working, methodical & scrupulous, in contrast to lazy, unreliable, irresponsible.**

**b12163 Psychic stability : Mental functions that produce a personal disposition that is even tempered, calm, composed, in contrast to irritable, worried, erratic & moody.**

**b12164 Openness to experience : mental functions that produce a personal disposition that is curious, imaginative, inquisitive & experience-seeking, in contrast to being stagnant, inattentive & emotionally inexpessive.**

**b12165 Optimism : Mental functions that produce a personal disposition that's cheerful, buoyant & hopeful, in contrast to being downhearted, gloomy & despairing.**

**b12166 Confidence : Mental functions that produce a personal disposition that's self-assured, bold & assertive, in contrast to being timid, insecure & self-effacing.**

---

---

**b12167 Trustworthiness : Mental functions that produce a personal disposition that's dependable & principled, as contrasted to being deceitful and anti-social.**

---

**Probe & Rating Guidelines for b1260 to b1265:**

(For each of the following personality traits, state the two extremes of a personality trait and ask parent which one describes X personality most of the time. For example : Does X show tendency to be shy & inhibited (negative trait), or, outgoing & sociable (positive trait)? If parent report that child tends to show negative trait, probe for frequency. (Note : Negative trait for each personality dimension listed are underlined).

- 0 No impairment. Child tends more towards the positive trait.
- 1 Mild impairment. Child shows negative traits less than 25% of the time (e.g. 1-2 times a month)
- 2 Moderate impairment. Child shows negative traits less than 50% of the time (e.g. 1-2 times a week)
- 3 Severe impairment. Child shows negative traits more than 50% of the time( e.g. 3-4 times a week).
- 4 Complete impairment. Child shows negative trait more than 95% of the time (e.g. daily).

---

**b134 Sleep Functions**

**Inclusions: functions of amount of sleeping, and onset, maintenance and quality of sleep; functions involving the sleep cycle, such as insomnia.**

**Exclusions : consciousness functions, energy & drive, attention function, psychomotor functions.**

**b1340 Amount of sleep : Mental functions involved in time spent in sleep.**

**b1341 Onset of sleep: Mental functions that produce transitions between wakefulness & sleep.**

**b1342 Maintenance of sleep : Mental functions that sustain the state of being asleep.**

**b1343 Quality of sleep : Mental functions produce material sleep leading to optimal physical and mental rest and relaxation.**

---

**Probe for b1340 to b1343 :**

Does X have any problems sleeping, e.g. getting to sleep, staying asleep? Is X a light sleeper? Does he suffer from insomnia? If yes, probe for frequency.

- 0 No impairment.
- 1 Mild impairment. Sleep difficulties present less than 25% of the time, e.g. 1-2 times a month.
- 2 Moderate impairment. Sleep difficulties present less than 50% of the time, e.g. 1-2 times a week.
- 3 Severe impairment. Sleep difficulties present more than 50% of the time, e.g. 3-4 times a week.
- 4 Complete impairment. Sleep difficulties present more than 95% of the time, e.g. daily.

---

**b144 Memory functions**

**Inclusions : functions of memory (ST & LT), immediate, recent & remote memory, memory span, retrieval of memory, functions used in learning, recalling. Exclusions : consciousness functions, orientation functions, energy & drive, thought functions function, higher-level cognitive functions, mental functions of language functions, perceptual functions, calculation functions**

**Probe & Rating Guidelines for b144**

How is X's memory? Does he have difficulties remembering things/new information (e.g. when learning new PECs cards or spelling list at school)? Is he good at remembering information related to his obsessions, e.g. names of dinosaurs? Does he remember where things are kept, or the places that he has been to?

- 0 No impairment/problem, including good memory.
- 1 Mild impairment. Able to remember/recall most things but require some assistance/repetition for learning new things/info (poor ST memory).
- 2 Moderate impairment. Adequate memory for info associated with interest, but otherwise poor short term memory.
- 3 Severe impairment. Able to remember/recall events, things in the environment but require much assistance/repetition for learning new things/info.
- 4 Complete impairment. Require much assistance/repetition for all types of info (eg events, things in the environment)

---

**b140 Attention functions**

*Exclusions : consciousness functions, energy & drive, sleep function, memory functions, psychomotor functions, perceptual functions*

**b1400 Sustaining attention : Mental function that produce concentration for the period of time required.**

---

**b1401 Shifting attention : Mental functions that permits refocusing concentration from one stimulus to another.**

**b1402 Dividing attention : Mental functions that permit focusing of one or more stimuli at the same time.**

**b1403 Sharing attention : Mental functions that permit focusing on the same stimulus by two or more people, such as child and caregiver both focusing on a toy.**

---

**Probe for b1400**

How easy is it to get X to move from one activity to another? Does this problem occur when he is engaged in activities that are his obsession, or does it also occur in most other activities (nb. Exclude obsessions)? How often does this problem occur?

**Probe for b1401**

How easy is it to get X to move from one activity to another? Does this problem occur when he is engaged in activities that are his obsession, or does it also occur in most other activities (nb. Exclude obsessions)? How often does this problem?

**Probe for b1402**

Can X attend to two things at one time? When do you see this problem, i.e. during what type of activities/contexts? Can he listen to radio while doing simple work, e.g. coloring, simple school work? What about tasks that require less concentration, e.g. eating & watching TV?

**Probe for b1403**

How often does X share his interest with you, e.g. brings a toy/object to show you? Are these limited to his 'obsessions' only? How often does X show difficulties in sharing attention?

**Rating Guidelines for b1400 to b1403**

- 0 No impairment.
- 1 Mild impairment. Difficulties occur less than 25% of the time, e.g. 1-2 times a month.
- 2 Moderate impairment. Difficulties occur less than 50% of the time, e.g. 1-2 times a week, or related to own interests/obsessions only.
- 3 Severe impairment. Difficulties occur more than 50% of the time, e.g. 3-4 times a week.
- 4 Severe impairment. Difficulties occur more than 95% of the time, e.g. daily.

---

**b147 Psychomotor functions**

*Specific mental functions of control over both motor & psychological events at body level.*

*Exclusions : consciousness functions, orientation functions, energy & drive, attention functions, intellectual functions, mental functions of sequencing complex movements.*

---

**1470 Psychomotor control : Mental functions that regulate the speed of beh or response time e.g. disruption of control producing psychomotor retardation (moving & speaking slowly, decrease in gesturing and spontaneity) or psychomotor excitement e.g. toe-tapping, handwringing, agitation or restlessness.**

**1471 Quality of psychomotor functions : Mental functions that produce non-verbal beh in the proper sequence & character of its sub-component parts, e.g. hand & eye coordination, or gait.**

---

**Probe & Rating Guidelines for b1470**

Does X show any motor stereotypies, or odd, involuntary physical movements such as handflapping, mid-line hand-wringing, tics etc. How often do these occur now?

- 0 No impairment.
- 1 Mild impairment. Difficulties occur less than 25% of the time, e.g. less than once a day.
- 2 Moderate impairment. Difficulties occur less than 50% of the time, e.g. 1-2 times a day.
- 3 Severe impairment. Difficulties occur more than 50% of the time, e.g. 3-4 times a day.

4 Severe impairment. Difficulties occur more than 95% of the time, e.g. more than 5 times a day.

**Probe for b1471**

Estimate extent of developmental delay (see continuum for gross motor skills & coordination below).

*Continuum for motor control and coordination (cf. DDST - S'pore)*

1:06yr	Throws ball indiscriminately
	Walk up with help
2:00yr	Throw ball fairly accurate
	Walk upstairs with feet together
3:00yr	Holds out hand to catch ball,
	Catches ball but clumsy
	Walk upstairs with alternating feet.
4:00yr	Catches ball fairly accurately
	Rides tricycle well
	Walks up and down stairs, alternating feet.
5:00yr	Climbs with agility (eg bars).
	Runs to catch a small ball

**Rating Guidelines for b1471**

- 0 No impairment. Gross motor control and coordination age-appropriate or developmental level 5 yrs.
- 1 Mild impairment. Less than 2 yrs delay.
- 2 Moderate impairment. 2 to 4 yrs delay.
- 3 Severe impairment. More than 4 yrs delay.
- 4 Complete impairment. More than 4 years delay and development is less than 1:06 yrs.

---

**b152 Emotional Functions**

**Specific mental functions related to the feeling and affective components of the processes of the mind.**

*Exclusions : consciousness functions, orientation functions, energy & drive, thought functions function, higher-level cognitive functions, mental functions of language functions, perceptual functions, calculation functions.*

---

**b1520 Appropriateness of emotions : Mental functions that produce congruence of feelings or affect with the situations, e.g. happiness at receiving good news.**

---

**b1521 Regulation of emotions : Mental functions that control the experience and display of affect.**

---

**b1522 Range of emotions : Mental functions that produce the spectrum of experience of arousal of affect or feelings such as love, hate, anxiousness, sorrow, joy, fear and anger.**

**Probe for b1520**

Has X ever shown emotions that are inconsistent with the context/situation, e.g. laughs at sad events, or gets angry or sad when given good news? How often does he show this?

**Probe for b1521**

Are the emotions he expressed tend to be 'exaggerated' or 'over-the-top'? If yes, how often does this occur? If no, is it the other extreme, i.e. he hardly express any emotions at all ?

**Probe for b1522**

Does he ever have problems show the full range of emotional expression, e.g. joy, happiness, reflective, puzzled, bored, jealous, angry, sad? Does he ever show a marked lack of emotion for long periods? How often has this occurred?

**Rating Guidelines for b1520 to b1522**

- 0 No impairment. Often shows appropriate levels/range of emotions.
- 1 Mild impairment. Problems present less than 25% of the time (once/twice a month)
- 2 Moderate impairment. Problems present less than 50% of the time (once/twice weekly)

---

3 Severe impairment. Problems present more than 50% of the time (3-4 times a week)  
 4 Complete impairment. Problems present more than 95% of the time daily.

---

### b156 Perceptual Functions

#### Specific mental functions of recognizing and interpreting sensory stimuli.

*Exclusions : consciousness functions, orientation functions, attention function, memory functions, mental functions of language, seeing & related functions, hearing & vestibular functions, additional sensory functions.*

b1560	Auditory perception : Mental functions involved in discriminating sounds, tones, pitches & other acoustic stimuli.
b1561	Visual perception : Mental functions involve in discriminating shape, size, color & other ocular stimuli.
b1562	Olfactory perception : Mental functions involved in distinguishing differences in smells.
b1563	Gustatory perception : Mental functions involved in distinguishing differences in taste, e.g. sweet, salty, sour & bitter stimuli, detected by tongue.
b1564	Tactile perception : Mental functions involved in distinguishing differences in texture, e.g. rough, smooth, detected by touch.
b1565	Visuospatial perception : Mental functions involved in distinguishing by sight the relative position of objects in the environment or in relation to oneself.

---

#### Probe & Rating Guidelines for b1560 to b1565

Does X have any problems in the following areas of perception? If yes, how/when was it diagnosed? What type of follow-up intervention did he need for this problem, as prescribed by doctors/therapists etc? NOTE: exclude hypersensitivities, which is coded elsewhere under b250-280. If structural damage is reported (4), probe for extent and nature of damage and code under 'Body Structure'.

- 0 No impairment.
- 1 Mild impairment. Impairment present less than 25% of the time (1-2 a month), with an intensity that the person can tolerate (i.e. do not require rehabilitation).
- 2 Moderate impairment. Impairment present less than 50% of the time (1-2 a week), which is interfering with child's day-to-day life (i.e. may require periodic/short term).
- 3 Severe impairment. Impairment present more than 50% of the time (3-4 times a week), which is partially disrupting child's day-to-day life, and requires major/frequent rehabilitation/intervention.
- 4 Complete impairment. Impairment present less than 95% of the time (daily), which is totally disrupting child's day-to-day life. If structural (anatomical) damage was reported, probe for extent & nature of damage and rate accordingly in next section 'Body Structure'.

---

### b160 Thought Functions

#### Specific mental functions related to ideational component of the mind.

*Exclusions : intellectual functions, higher-level cognitive functions, memory functions, mental functions of language, psychomotor functions, calculation functions.*

b1600	Pace of thought : Mental functions that govern speed of thinking process.
b1601	Form of thought : Mental functions that organize thinking process so as to its coherence or logic. Include - tangentiality and circumstantiality.
b1602	Content of thought : Mental functions consisting of the ideas that are present in the thinking process & what is being conceptualized. Include impairments of delusion, overvalued ideas and somatization.
b1603	Control of thought : Mental functions that provide volitional control of thinking and recognized as such by the person. Include impairments of rumination, obsession, thought broadcast and thought insertion.

---

### Probe for b1600

Does X have any problems in terms of thinking speed? When you give him a task or an instruction (which you know he can perform/understand), does he take a long time to respond? How often /in what contexts does he show this problem?

### Probe for b1601

When X tells you his ideas (e.g. when he tells you what happened in school, or gives information about his topic of interest), are his ideas coherent, logical? Does he go 'off-topic', i.e. suddenly talking about things which are unrelated? How often /in which contexts does he show this problem?

### Probe for b1602

Does X have any delusions? For example, he thinks he's someone else, e.g. delusions about being a Minister, King, etc.? How often does he show this problem?

### Probe for b1603

PROBE for b1603 : Does he have any obsessions? What are they? How often does he engage in these obsessive behaviour?

### Rating Guidelines of b1600 to b1603

- 0 No impairment
- 1 Mild impairment. Problem occurs but infrequently (eg 1-2 times a month)
- 2 Moderate impairment. Problem occurs occasionally (1-2 times a week), e.g. when faced with new situation/information.
- 3 Severe impairment. Problem occurs frequently (3-4 times a week), with new and familiar situation/information.
- 4 Complete impairment. Problem occurs daily, in all situation, with all information.
- 8 Parent report that child's communication is limited, so unable to gauge thought processes.

---

### b164 Higher level cognition functions

**Specific mental functions especially dependent on frontal lobe of brain, including complex goal-directed beh eg decision making, abstract thinking, planning and carrying out plans, mental flexibility & deciding which beh are appropriate under what circumstances, often called executive functions.**

*Exclusions : consciousness functions, orientation functions, attention function, memory functions, mental functions of language, seeing & related functions, hearing & vestibular functions.*

---

- b1640 Abstraction** : Mental functions of creating general ideas, qualities out of and distinct from, concrete realities, specific objects or actual instances.
- b1641 Organization & planning** : Mental functions of coordinating parts into a whole, of systematizing, the mental functions involved in developing a method of proceeding or acting.
- b1642 Time management** : Mental functions of ordering events in chronological sequence, allocating amount of time to events and activities.
- b1643 Cognitive flexibility** : Mental functions of changing strategies, or shifting mental states, esp. as involved in problem solving.
- b1644 Insight** : Mental functions of awareness & understanding of oneself & one's behaviour.
- b1645 Judgment** : Mental functions involved in discriminating between and evaluating different options, such as those involved in making an opinion.
- b1646 Problem-solving** : Mental functions of identifying, analyzing & integrating incongruent or conflicting information into a solution.

---

### Probe for b1640

Does X have problems understanding higher order concepts, e.g. that anger, sadness, happiness are all 'feelings'; shoes, shirt are 'clothes', and cars, buses are 'vehicles'? If yes, how often/in what contexts does he show this problem?

### Probe for b1641

When X is given a task to do, e.g. complete a jig-saw, draw a picture, does he do the task in a organized, step-by-step way? Does he tend to be haphazard in his approach, i.e. completing the task in an almost random way? If yes, how often/in what contexts does he show this problem?

**Probe for b1642**

When X is doing daily activities, does he allocate adequate time for it? Does he tend to spend too much or too little time on a task (nb: exclude obsessions)? How often/ in what contexts does he show this problem?

**Probe for b1643**

Does X tend to be rigid and fixed in the way he approach new tasks? For example. when X is given a new task/toy, is he able to adapt strategy that he has to suit the new task/toy? How often / in what contexts does he show this problem?

**Probe for b1644**

How aware is X of his own behaviour and how it affect others? If yes, give example of how he shows this awareness. e.g. showing remorse. If not, how often / in what contexts does he show this problem?

**Probe for b1645**

Is X able to evaluate two options, and make an opinion as to which one is the better choice/option? Is this limited to choices regarding his own needs? If yes, how often/in what contexts does he show this problem?

**Probe for b1646**

When faced with a problem situation, is he able to find a solution? In what context has he shown these skills? E.g. In schoolwork (problem sums), finding the best routes to a particular place; resolving conflicts between two people. If yes, how often/in what contexts does he show this problem?

**Rating Guidelines for b1640 to b1646**

- 0 No impairment
- 1 Mild problem. Problem present less than 25% of the time, e.g. 1-2 times a month, in contexts of unfamiliar setting/tasks.
- 2 Moderate problem. Problem present less than 50% of the time, e.g. 1-2 times a week, in contexts of unfamiliar setting/tasks.
- 3 Severe problem. Problem present more than 50% of the time, e.g. 3-4 times a week, in familiar and unfamiliar setting/tasks.
- 4 Severe problem. Problem present more than 95% of the time, e.g. daily, in familiar and unfamiliar setting/ tasks.

---

**b167 Mental functions of language****Specific mental functions of recognizing and using signs, symbols & other components of a language.**

*Exclusions : attention function, memory functions, mental functions of complex movements, higher-level cognitive functions, calculation functions, sensory functions and pains, voice & speech functions*

**b1670 Reception of language : Mental functions of decoding messages to obtain their meaning.**

**b1671 Expression of language : Specific mental functions necc to produce meaningful messages**

**b1672 Integrative language functions: Mental functions that organize semantic and symbolic meaning, grammatical structure and ideas for the production of messages in spoken, written or other forms of language.**

---

**Probe for b1670**

Based on parents' reports of child's behaviour on the continuum of receptive language, interviewer to estimate extent of delay in receptive language skills :

*Receptive Language Continuum (cf. S'pore Oracy Continuum).*

1yr - Respond to name only

1:3yr - Understands simple phrase in context, part learnt sequence, eg. Give mummy a kiss.

2yr - Knows meaning of some words

2:6yr- Follow instruction involving 2 named objects.

3yr - Understand a sequence of commands, 3 steps.

4yr - Understand instructions involving decisions.

**Probe for b1671**

Based on parents' reports of child's behaviour on the continuum of expressive language, interviewer to estimate extent of delay in receptive language skills :

*Expressive Language Continuum (cf. S'pore Oracy Continuum).*

1yr - Give names of people/object when asked.

1:8yr - Spontaneously says names of familiar objects.

2yr - Two word phrases "Want dinner"

3yr - Longer phrases with nouns & verbs, missing linking words.

3:6yr Talk in spontaneous sentences, present tense only.

4yr - Uses past, present future tense and complex grammar constructions.

#### **Rating Guidelines for b1670 to b1671**

- 0 No impairment. Age appropriate, or achieved development of 4 yrs.
- 1 Mild impairment. Less than 2 yrs delay
- 2 Moderate impairment. 2 to 4 yrs delay
- 3 Severe impairment. More than 4 yrs delay
- 4 Complete impairment. Not achieved 1:0 yr developmental level, e.g. no speech sounds, i.e. make noises without meaning, not babbling/baby noises.

#### **Probe & Rating Guidelines for b1672**

If child speech is at phrases/sentences level, does X use linking words such as 'and' 'then', 'but', 'because'. Does X use present/past tense consistently /accurately? How often does X show this problem?

- 0 No impairment. Able to use several linking sentences, with appropriate tenses.
- 1 Mild impairment. Able to use a few sentences. Some use of verbs and nouns, use simple tenses although at times incorrect.
- 2 Moderate impairment. Able to use a few sentences. Mainly names of object/people. No tense markers.
- 3 Severe impairment. Speech limited to two/three word phrases only.
- 4 Complete impairment. Child's speech is at one word level only less.
- 8 Unspecified. Child speech is at 2 yrs' old level or less.

---

#### **b172 Calculation functions**

**Specific mental functions of determining, approximation and manipulation of mathematical symbols and processes.**

*Exclusions : attention function, memory functions, mental functions of language functions.*

**b1720 Simple calculation : Mental functions of computing with numbers such as addition, subtraction, multiply & divide.**

**b1721 Complex calculations : Mental functions of translating word problems into mathematical formulas, translating mathematical formulas into arithmetic procedures, and other complex manipulations involving numbers.**

---

#### **Probe for b1720**

Based on parent's report of what child can do on numeracy continuum (below), interviewer to estimate delay in basic number skills:

*Numeracy Continuum (cf. S'pore Numeracy Continuum).*

3 yr Can count to 2 with meaning. Counting by rote up to 5.

4 yr Has one-to-one correspondence (can give 2 cups, 3 pencils). Count up to 5 with 1-2-1 correspondence.

5 yr Has number concept up to 10. Can classify objects by quantity (eg match 3 dolls with 3 ice-creams).

6 yr Can do simple addition.

7 yr Can do addition and subtraction of numbers up to 20.

8 yr Can do multiplication up to 5. Understand concept of division (e.g. can group objects into multiple sets of 5).

---

#### **Rating Guidelines for b1720**

- 0 No impairment. Age appropriate, or achieved development of 8 yrs.
- 1 Mild impairment. Less than 2 yrs delay
- 2 Moderate impairment. 2 to 4 yrs delay
- 3 Severe impairment. More than 4 yrs delay

4 Complete impairment. Not achieved 3:0 yr developmental level or less, e.g. no number concept.

**Probe & Rating Guidelines for b1721**

(If child is able to do simple addition/subtraction) Can X do problem sums (or come up with 'number sentences'). Give examples. What types of problem sums can X do independently, and which ones he has difficulty, e.g. with picture cues, concrete aids?

- 0 No impairment. Able to translate word problems into number sentences.
- 1 Mild impairment. Able to translate word problems into number sentences with picture cues.
- 2 Moderate impairment. Able to translate word problems into number sentences with concrete aids.
- 3 Severe impairment. Able to translate word problems into number sentences (with picture cue/concrete aids) for familiar problems only (i.e. sums that child has repeatedly been exposed to).
- 4 Complete impairment. No understanding of numbers/ developmentally less than 6yrs.
- 8 Not specified. Child is less than 6 yrs old.

---

**b176 Mental functions of sequencing complex movements**

**Specific mental functions of sequencing and coordinating complex purposeful movements.**

*Inclusions : psychomotor functions, higher-level cognitive functions, neuroskeletal & movement related functions.*

**Probe & Rating Guidelines for b176**

Can X follow sequence of movement, e.g. in dance or PE?

- 0 Child able to follow sequence of actions in PE / dance with good coordination.
- 1 Child is able to follow sequence of action in PE/dance, but is a little clumsy.
- 2 Child able to follow two-step physical actions.
- 3 Child able to follow one-step physical action.
- 4 Child unable to coordinate own movement, e.g. severe delay in gross motor skills/coordination.

---

**b180 Mental functions of awareness of self, identify and body image.**

**Specific mental functions related to awareness of one's identity, body, position in the reality of one's environment & time.**

*Inclusions : Functions of experience of self and body image.*

**Probe & Rating Guidelines for b180**

Does X tend to be confused about his identity, or body shape, e.g. he thinks he is younger or older than he actually is? Does he ever feel that his body is too fat or too thin? If yes, how often does this occur?

- 0 No impairment
- 1 Mild impairment. Problem observed infrequently (once/twice a month)
- 2 Moderate impairment. Problem observed occasionally (once/twice weekly)
- 3 Severe impairment. Problem observed frequently (3-4 times a week), partially disrupting child's life.
- 4 Complete impairment. Problem observed daily, and totally disrupting child's day-to-day life.
- 8 Not specified. Parent reported that child's communication too impaired to gauge awareness level.

---

**b2 SENSORY FUNCTIONS AND PAIN**

**b210 Seeing Functions**

**Sensory functions relating to sensing presence of light and sensing form, size, shape, and color of visual stimuli.**

**b215 Functions of structures adjoining the eye**

**Functions of structures in and around the eye that facilitate seeing functions.**

**b230 Hearing functions**

**Sensory functions relating to sensing the presence of sounds and discriminating the location, pitch, loudness and quality of sounds.**

**b235 Vestibular functions**

**Sensory functions of the inner ear related to position, balance and movement. (e.g. dizziness, vertigo)**

**b250 Additional sensory functions : Taste functions**

**Sensory functions of sensing qualities of bitterness, sweetness, sourness and saltiness.**

**b255 Additional sensory functions : Smell functions**

**Sensory functions of sensing odors and smell.**

**b260 Additional sensory functions : Proprioceptive functions**

**Sensory functions of sensing relative position of body parts.**

**b265 Additional sensory functions : Touch functions**

**Sensory functions of sensing surfaces and their texture or quality.**

**b270 Additional sensory functions : Sensory functions related to temperature and other stimuli**

**Sensory functions of sensing temperature, vibration, pressure (superficial or deep), burning sensation or noxious stimuli (including sensitivity to these sense).**

**b280 Sensation of Pain**

**Sensation of unpleasant feeling indicating potential or actual damage to some body structure.**

#### **Probe & Rating Guidelines for for b210 to b280**

Does X show any problems in these sensory functions, for e.g., any hypersensitivity / hypo sensitivity? Give examples. Was the problems diagnosed & by whom? What were the recommended interventions, for example, does he need therapy to overcome this? What type and for how long?

- 0 No impairment.
- 1 Mild impairment. Impairment present less than 25% of the time (1-2 a month), with an intensity that the person can tolerate (i.e. do not require rehabilitation).
- 2 Moderate impairment. Impairment present less than 50% of the time (1-2 a week), which is interfering with child's day-to-day life (i.e. may require periodic/short term).
- 3 Severe impairment. Impairment present more than 50% of the time (3-4 times a week), which is partially disrupting child's day-to-day life, and requires major/frequent rehabilitation/intervention.
- 4 Complete impairment. Impairment present less than 95% of the time (daily), which is totally disrupting child's day-to-day life. If structural (anatomical) damage was reported, probe for extent & nature of damage and rate accordingly in next section 'Body Structure'.

### **b3 VOICE AND SPEECH FUNCTIONS**

**b310 Voice functions**

**Functions of the production of various sounds by the passage of air through the larynx.**

*Inclusions : functions of production and quality of voice; functions of phonation, pitch, loudness and other qualities of voice. Exclusion : Mental functions of language, articulation functions.*

**b320 Articulation functions**

**Functions of the production of speech sounds, including enunciation, articulation of phonemes, spastic, ataxic, flaccid dysarthria, anarthria.**

*Inclusions : functions of enunciation, articulation of phonemes. Exclusion : Mental functions of language, voice functions*

**b330 Fluency and rhythm of speech functions**

**Functions of the production of flow and tempo of speech.**

*Inclusions : functions of fluency, rhythm, speed and melody of speech; impairments such as stuttering, stammering, bradyphasia, tachphasia. Exclusion : Mental functions of language, voice & articulation functions.*

**b340 Alternative vocalizations functions**

---

**Functions of the production other manners of vocalizations, including humming, crying aloud, screaming, babbling**

*Inclusions : functions of the production of notes and range of sounds, such as in singing, chanting, babbling and humming. Exclusion : Mental functions of language, voice functions, articulation functions, fluency and rhythm of speech functions.*

---

**Probe & Rating Guidelines for b310**

PROBE for b310 : When X speaks, is his voice clearly audible? Does it tend to be too soft or too loud? If problem exists, what is the extent of the problem?

- 0 No impairment
- 1 Mild impairment. Less than 25%, e.g. no impairment in voice/sound production but slight impairment in quality (eg in pitch, voice sometimes too loud/soft).
- 2 Moderate impairment. Less than 50%, e.g. no impairment in voice/sound production and moderate impairment in quality (eg in pitch, voice often too loud/soft).
- 3 Severe impairment. More than 50%, e.g. some impairment in voice/sound production and severe impairment in quality (eg in pitch, voice often too loud/soft).
- 4 Complete impairment. More than 95%, e.g. child does not produce speech/voice sounds (e.g. mute).

**Probe & Rating Guidelines for b320**

When X speaks or says a word, how easy is it to understand what he is saying? Would unfamiliar people understand what he is saying?

- 0 No impairment
- 1 Mild impairment. Less than 25%, e.g. only some words/sounds mispronounced/are unclear, but intelligibility unaffected.
- 2 Moderate impairment. Less than 50%, e.g. speech is easily understood by family but unfamiliar people would find it difficult.
- 3 Severe impairment. More than 50%, e.g. at times, even family members find it difficult to understand the words spoken.
- 4 Complete impairment. More than 95%, e.g. speech is meaningless echolalia only/no speech)

**Probe & Rating Guidelines for b330**

When X speaks, does he use a flat or monotonous tone? Do the phrases tend to have the same tone?

- 0 No impairment
- 1 Mild impairment. Less than 25%, e.g. speech is flat or monotonous but not a major problem (i.e. some phrases have the same tone)
- 2 Moderate impairment. Less than 50%, e.g. flat or monotonous speech is a marked feature, i.e. most phrases have the same tone)
- 3 Severe impairment. More than 50%, e.g. speed and fluency affected, such as marked stammering, marked echolalia.
- 4 Complete impairment. More than 95%, e.g. speech is meaningless echolalia only/no speech.

---

**b4 FUNCTIONS OF THE CARDIOVASCULAR, HAEMATOLOGICAL, IMMUNOLOGICAL & RESPIRATORY SYSTEMS**

**b410 Functions of the cardiovascular system**

*Functions of pumping blood in adequate or required amounts and pressure, and transporting blood throughout the body.*

---

**b430 Functions of the hematological and immunological systems**

*Functions of blood production & functions of body related to protection against foreign substances.*

---

---

**b440 Functions of the respiratory systems**

*Functions of inhaling air into the lungs, exchange of gases between air and blood and exhaling air.*

**b510 Functions related to digestive system**

*Inclusion : Ingestion functions, digestive function, assimilation function, defecation function, weight maintenance function.*

**b540 Functions related to metabolism and endocrine systems**

*Inclusion : General metabolic functions, water, mineral and electrolyte balance functions, thermoregulatory function, endocrine gland function*

**b610 Functions related to filtration, collection & discharge of urine.**

*Inclusion : Urinary excretory functions, urination functions, sensations associated with urinary functions.*

**b640 Genital & reproductive functions**

*Inclusion : Sexual function, menstruation functions, procreation function, sensation associated with genital and reproductive function*

**b710 Functions of the joint and bones**

*Inclusion : Mobility of joint functions, stability of joint functions, mobility of bone functions.*

**b730 Muscle functions**

*Inclusion : Muscle power functions, muscle tone functions, muscle endurance functions.*

**b750 Movement functions**

*Inclusion : Motor reflex functions , involuntary movement reaction functions, control of voluntary movement & involuntary movement functions (including tics, coprolalia, motor stereotypes)*

**b810 Functions of the skin**

*Inclusion : Protective, problems in repair functions of the skin.*

**b850 Functions of hair and nails**

*Inclusion : Loss of hair, pigmentation of hair; growth, pigmentation & quality of nails.*

**Probe & Rating Guidelines for b410 to b850**

Has X been diagnosed to have any problems in these areas of functioning. Who made the diagnosis? To what extent is the function impaired? How much rehabilitation is required/recommended?

- 0 No impairment.
- 1 Mild impairment. Impairment present less than 25% of the time (1-2 a month), with an intensity that the person can tolerate (i.e. do not require rehabilitation).
- 2 Moderate impairment. Impairment present less than 50% of the time (1-2 a week), which is interfering with child's day-to-day life (i.e. may require periodic/short term).
- 3 Severe impairment. Impairment present more than 50% of the time (3-4 times a week), which is partially disrupting child's day-to-day life, and requires major/frequent rehabilitation/intervention.
- 4 Complete impairment. Impairment present less than 95% of the time (daily), which is totally disrupting child's day-to-day life. If structural (anatomical) damage was reported, probe for extent & nature of damage and rate accordingly in next section 'Body Structure'.

---

**d1 PURPOSEFUL SENSORY EXPERIENCE****d110 Watching**

*Using the sense of seeing intentionally to experience visual stimuli, e.g. watching tv, sporting event or children playing.*

**d115 Listening**

*Using the sense of hearing intentionally to experience auditory stimuli, e.g. listening to radio, music or lecture.*

**d120 Purposeful sensing**

**Using body's other basic senses intentionally to experience stimuli, e.g. touching and feeling textures, tasting sweets, smelling flowers.**

**Probe & Rating Guidelines for d110 to d120**

Does X have any difficulty carrying out this activity? How often (or in what contexts) does he experience difficulties? Does he need any assistance, e.g. by other people, devices, or modification of the task/activity?

- 0 No difficulty. Engages in the activity independently.
- 1 Mild difficulty. Problems present less than 25% of the time, e.g. child not able to engage independently in some activity (i.e. if not directed/led by others).
- 2 Moderate difficulty. Problems present less than 50% of the time, e.g. child not able to engage independently other than a few/selected visual/auditory stimuli only, for e.g., watches only 1 TV program, listen to 1 song repeatedly.
- 3 Severe difficulty. Problems present more than 50% of the time, for e.g. child engages in activity but for sensory stimulation only, i.e. not purposeful, e.g. 'watching' TV for sounds and light movements.
- 4 Complete difficulty. Problem present more than 95% of the time. Child unable to engage in activity at all.

**d1 BASIC LEARNING**

**d130 Copying**

**Imitating or mimicking as a basic component of learning, e.g. copying a gesture, sound or letters of alphabet.**

**Probe & Rating Guidelines for d130**

Can X copy / imitate a gesture, an action or copy alphabet independently? Does he imitate actions spontaneously? Does he need any verbal prompting or physical assistance/assistance? How often/in what contexts are these problems present?

- 0 No difficulty. Child copies / imitate spontaneously.
- 1 Mild difficulty. Problem present less than 25% of the time, e.g. child imitates independently, if directed/led by others (i.e. not spontaneous).
- 2 Moderate difficulty. Problem present less than 50% of the time, e.g. child needs verbal prompting/guidance to perform imitation/copying tasks.
- 3 Severe difficulty. Problems present more than 50% of the time, e.g. child needs physical prompting and much / extended repetition.
- 4 Complete difficulty. Problems present more than 90% of the time, e.g. child engages in repetitive or self-stimulatory behaviour only.

**d135 Rehearsing**

**Repeating a sequence of events or symbols as a basic component of learning such as counting by tens or practicing the recitation of poems.**

**Probe & Rating Guidelines for d135**

Can rehearse the alphabet, poems, numbers by ones/tens independently? Does he do this spontaneously when learning new information? Does he need any verbal prompting, physical assistance or extended repetition? How often/ in what contexts does X have difficulties doing this activity?

- 0 No difficulty. Child recites , rehearses spontaneously.
- 1 Mild difficulty. Problems present less than 25% of the time, for e.g. child rehearses if directed/led by others.
- 2 Moderate difficulty. Problems present less than 50% of the time, for e.g. child needs some verbal prompting, and some repetition.
- 3 Severe difficulty. Problems present more than 50% of the time, for e.g. child needs verbal/physical prompting and much / extended repetition.
- 4 Complete difficulty. Problems present more than 95% of the time, e.g. child engages in repetitive or self-stimulatory behaviour only, or no speech.

**d140 Learning to read**

**Developing competency to read written material with fluency and accuracy, sounding out words and letters, understanding words and phrases.**

**Probe for d140**

What is X's current reading skills (estimate using Reading Continuum below)?

*Reading Continuum (cf. S'pore Reading Continuum)*

- 3:06yr Can recognise own name.
- 4:00yr Can match words to pictures
- 4:06yr Can recognise up to 10 familiar words.
- 5:06yr Can read simple first reading books.
- 7:00yr Can read books for children aged abt 7yr (K2 level)

**Rating Guidelines for d140**

- 0 No difficulty. Child able to read words/phrases with meaning/understanding, appropriate to age level or has attained 7 yrs' developmental level.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and reading level less than 3, e.g. no word recognition.

**d145 Learning to write**

**Developing competency to produce symbols that represent sounds, words and phrases in order to convey meaning.**

**Probe for d145**

What is X's current writing skills (estimate using Writing Continuum below)?

*Writing Continuum (cf. S'pore Writing Continuum)*

- 4:00yr Can write some letters by copying
- 4:06yr Can write simple words by copying
- 5:00yr Can write some letters without copying
- 5:03yr Can write simple words without copying
- 6:00yr Can write 12 or more words without copying
- 9:08yr Can write a short letter - not copied.

**Rating Guidelines for d145**

- 0 No difficulty. Child able to write words/phrases with meaning/understanding, appropriate to age level or has attained 9 yrs' developmental level.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and writing level less than at 4 yrs' level, e.g. cannot copy write.

**d150 Learning to calculate**

**Developing competency to manipulate numbers and perform simple and complex math operations, e.g. using math signs for additions, subtraction and applying correct math operation to a problem.**

**Probe for d145**

What is 'Xs level of numeracy skills (estimate using Numeracy continuum)? How much assistance did he require?

*Numeracy Continuum (cf. S'pore Numeracy Continuum)*

- 3 yrs Can count to 2 with meaning. Counting by rote up to 5.
- 4 yrs Has one-to-one correspondence (can give 2 cups, 3 pencils). Count up to 5 with 1-2-1 correspondence.

5 yrs Has number concept up to 10. Can classify objects by quantity (eg match 3 dolls with 3 ice-creams).

6 yrs Can do simple addition.

7 yrs Can do addition and subtraction of numbers up to 20.

8 yrs Can do multiplication up to 5. Understand concept of division (e.g. can group objects into multiple sets of 5).

**Rating Guidelines for d145**

- 0 No difficulty. Child numeracy skills age appropriate, or has attained 8 yrs' developmental level.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and writing level less than 3 yrs' old level, e.g. no number concept.

**d155 Acquiring skills (Basic)**

**Developing basic competencies in integrated sets of actions or tasks so as to initiate and follow through with acquisition of skills, such as manipulating tools or playing games like board games.**

**Probe & Rating Guidelines for d155**

Does X have difficulties learning a new skill/game at home, e.g. a board game/toy? How much help does he need? Can he learn to follow the game / use a tool by simple demonstration? Does he need much repetition or the task/game to be simplified? How often / in what contexts does X experience difficulties carrying out in this activity? NOTE : Exclude academic skills.

- 0 No difficulty. Child able to acquire skills through simple demonstration and/ or imitation.
- 1 Mild difficulty. Problems present less than 25% of the time, e.g. child requires tasks to be broken into simpler steps.
- 2 Moderate difficulty. Problems present less than 50% of the time, e.g. child needs steps to be broken down/simplified and some repetition.
- 3 Severe difficulty. Problems present more than 50% of the time, e.g. child needs much physical prompting and extended repetition.
- 4 Complete difficulty. Problems present more than 95% of the time, e.g. child has not acquired skills to use any simple tools or games (e.g. cutlery, leggo).

**d1 APPLYING KNOWLEDGE**

**d160 Focusing attention**

**Intentionally focusing on specific stimuli e.g. by filtering out distracting noises.**

**Probe & Rating Guidelines for d160**

Is X easily distracted? How much assistance (verbal or physical) does he need to stay on task for 15 mins? How often does he show this difficulty?

- 0 No difficulty. Can attend for abt 15 mins on an activity (achieved by 5 yrs)
- 1 Mild difficulty. Problem present less than 25% of the time.
- 2 Moderate difficulty. Problem present less than 50% of the time.
- 3 Severe difficulty. Problem present more than 50% of the time.
- 4 Complete difficulty. Problem present more than 95% of the time, e.g. fleeting attention, even with physical/verbal prompting.

**d163 Thinking**

**Formulating and manipulating ideas, concepts and images, whether goal-oriented or not, either alone or with others, such as creating fiction, playing with ideas, brainstorming, speculating or reflecting.**

**Probe & Rating Guidelines for d163**

Can X come up with his own story, or ideas? Are these related to his interest topics/obsessions? Can he come up with a story for a given topic (e.g. in composition exercises in school). Can he do this on his own (independently) or does he need

assistance? How often does he face difficulties engaging in such activities and how much assistance does he need?

- 0 No difficulty. Can come up with ideas / stories spontaneously.
- 1 Mild difficulty. Problems present less than 25% of the time, e.g. can come up with ideas independently if topic / story is related to his interest area.
- 2 Moderate difficulty. Problems present less than 50% of the time, e.g. can come up with ideas if led by others, e.g. questions as prompts.
- 3 Severe difficulty. Problems present more than 50% of the time, e.g. with much guidance, able to repeat stories heard from others, i.e. imitate /reproduce ideas.
- 4 Complete difficulty. Problems present more than 95% of the time, e.g. not able to come up with ideas, even with much prompting.
- 8 Not specified. Not enough communication to gauge ability to think.

---

#### d166 Reading

**Perform activities involved in the comprehension and interpretation of written language (e.g. books, instructions, newspapers), of purpose of obtaining info or general knowledge.**

#### Probe & Rating Guidelines for d166

Does X have difficulties reading? Does he understand / comprehend most of what he reads? How often do the problem occurs?

- 0 No difficulty. Able to read words, phrases with understanding.
- 1 Mild difficulty. Problems present 25% of the time, e.g. age appropriate reading but comprehension slightly delayed (1-2yrs).
- 2 Moderate difficulty. Problems present less than 50% of the time, e.g. 2 yr to 3 yrs delay in reading and comprehension.
- 3 Severe difficulty. Problems present more than 50% of the time, e.g. marked delay in reading and comprehension (4yrs or >).
- 4 Complete impairment. Problems present more than 95% of the time, e.g. no reading skills.

---

#### d170 Writing

**Using or producing symbols or language to convey information e.g. producing written record of events or ideas or drafting a letter.**

#### Probe & Rating Guidelines for d166

Does X have difficulties reading? Does he understand / comprehend most of what he reads? How often do the problem occurs?

- 0 No difficulty. Able to write words, phrases with understanding.
- 1 Mild difficulty. Problems present 25% of the time, e.g. age appropriate writing but comprehension slightly delayed (1-2yrs).
- 2 Moderate difficulty. Problems present less than 50% of the time, e.g. 2 yr to 3 yrs delay in writing and comprehension.
- 3 Severe difficulty. Problems present more than 50% of the time, e.g. marked delay in writing and comprehension (4yrs or >).
- 4 Complete impairment. Problems present more than 95% of the time, e.g. no writing skills.

---

#### d172 Calculating

**Performing computations by applying mathematical principles to solve problems that are described in words and producing or displaying the results, e.g. computing sums of 3 numbers.**

#### Probe & Rating Guidelines for d172

Does X have difficulties doing problem sums? What types of problem sums can X do independently, and which ones he has difficulty?

- 0 No difficulty. Able to translate word problems into number sentences (achieved by 6 yrs).

---

- 1 Mild difficulty. Problems present less than 25% of the time, e.g. able to translate word problems into number sentences if given picture cues.
- 2 Moderate difficulty. Problems present less than 50% of the time, e.g. able to translate word problems into number sentences with picture AND concrete aids.
- 3 Severe difficulty. Problems present more than 50% of the time, e.g. able to translate word problems into number sentences (with picture cue/concrete aids) for familiar problems only (i.e. sums that child has repeatedly shown).
- 4 Complete difficulty. Problems present more than 95% of the time, e.g. no understanding of numbers.

---

**d175 Problem solving**

**Finding solutions to questions or situations by identifying and analyzing issues, developing options & solutions, evaluating effects of solutions, and executing a chosen solution, e.g. in resolving disputes between two people.**

**Probe & Rating Guidelines for d175**

When faced with a problem situation, is he able to find a solution? How often/in what contexts does he face difficulties in problem solving? E.g. In schoolwork (problem sums), resolving conflicts between two people.

- 0 No difficulty. Able to problem solve in a variety of tasks/situations.
- 1 Mild difficulty. Problems present less than 25% of the time, e.g. can problem solve in most areas, but often encounter problems in one area (e.g. social/interpersonal conflicts).
- 2 Moderate difficulty. Problems present less than 50% of the time, e.g. can problem-solve if issues/topic is related to his interests, and not able to solve social/interaction conflicts?
- 3 Severe difficulty. Problems present more than 50% of the time, e.g. can provide solutions to familiar problems which is related to child's interest area. Not able to solve new/novel problems.
- 4 Complete difficulty. Problems present less than 25% of the time, e.g. no problem solving skills, able to make choices (basic needs) for self only.
- 8 Not specified. Parent report that child has too little communication to rate.

---

**d177 Making decision**

**Making a choice among options, implementing the choice, and evaluating the effects of the choice, such as selecting and purchasing a specific item, or deciding to undertake one task from among several tasks that need to be done.**

**Probe & Rating Guidelines for d177**

Does X have difficulties making choices/decisions for himself? How often/in what contexts does he face these difficulties?

- 0 No difficulty. Can make choices for self on many areas.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. can make choices for others on various areas, but often on own terms, e.g. chooses presents for sibling/parent that would be more suitable for himself.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. can make choices/decisions only if issue/topic is familiar or related to his interests, e.g. bus routes, vehicles).
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. able to make choices (basic needs) for self only.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. not able to make choices for daily needs.

---

**d2 GENERAL TASK DEMANDS**

**d210 Undertaking a single task**

**Carrying out a simple or complex coordinated actions related to the mental and physical components of a single task, such as initiating a task, organising time, space and materials, pacing task performance, and carrying out, completing and sustaining a task.**

*Inclusions : undertaking a single task independently or in groups. Exclusions : acquiring skills, solving problems, making decisions, undertaking multiple tasks.*

---

### Probe & Rating Guidelines for d210

How much difficulty does X have in carrying out a single task, for example when asks to draw a picture. How often/ in what contexts (group or one-to-one contexts) does he face the difficulties?

- 0 No difficulty. Can undertake a single task independently in group situation (based on group instructions)
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. can undertake a single task with individual instruction (1-2-1).
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. can undertake a single task (1-2-1 context) with verbal prompting.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. can undertake a single task with verbal and physical prompting/assistance.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. engages in repetitive, self-stimulatory behaviour only.

---

### d220 Undertaking multiple tasks

**Carrying out simple and coordinated actions as components of multiple, integrated and complex tasks in sequence or simultaneously.**

***Inclusions : undertaking and completing multiple tasks independently and in groups. Exclusions : acquiring skills, solving problems, making decisions, undertaking single tasks***

### Probe & Rating Guidelines for d220

How much difficulty does X have when doing multiple tasks e.g., making a birthday card (involves drawing, cutting, pasting). How often / in what contexts (group; one-to-one contexts) does he face these difficulties?

- 0 No difficulty. Can undertake multiple tasks independently in group situation (based on group instructions).
- 1 Achieved by 6 yrs.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. can undertake multiple tasks with individual instruction (1-2-1)
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. can undertake multiple tasks (1-2-1 context) with verbal prompting.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. can undertake multiple tasks with verbal and physical prompting/assistance.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. can only do single tasks with physical prompts.
- 8 Not specified. Child is less than 6yrs.

---

### d230 Carrying out daily routine

**Carrying out simple and coordinated actions in order to plan, manage and complete requirements of day-to-day procedures or duties.**

***Inclusions : managing and completing daily routines. Exclusions : undertaking multiple tasks.***

### Probe & Rating Guidelines for d230

Does X have difficulties managing daily routines independently? For e.g. getting ready for school, following routines in school. How often / in what contexts (e.g. how much assistance does he need) does he face these difficulties?

- 0 No difficulty. Can manage daily routine independently (appropriate to age).
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. can manage daily routine but sometimes need reminders (achieved by 6 yrs).
- 2 Moderate difficulty. Difficulties present less than 25% of the time, e.g. can start/initiate daily routine (may need reminders) but needs help to complete them.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. needs prompting to start and complete daily routine.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. completely reliant on others to initiate and complete daily routine.
- 8 Not specified. Child is less than 6yrs.

---

**d240 Handling stress and other psychological demands**

**Carrying out coordinated actions to manage and control the psychological demands required to carry out tasks demanding significant responsibilities and involving stress.**

*Inclusions : handling responsibilities; handling stress and crisis.*

**Probe & Rating Guidelines for d240**

Does X have difficulties handling stress? Does he gets stress easily? How often / in what contexts does he face these difficulties? Are these 'real' or perceived stressors? How does X react to stress, e.g. demands placed by daily responsibilities in school and during school exam?

- 0 No difficulties. Able to handle stress (appropriate to age).
- 1 Mild difficulties. Difficulties present less than 25% of the time, e.g. able to handle responsibilities but some problems in dealing with stressful situations (e.g. gets stressed/anxious easily).
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. able to handle responsibilities but problems in dealing with stressful situations (reacts badly under perceived stress, e.g. 'shuts down', temper tantrums).
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. problem handling responsibilities and stressful situations.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. not aware of responsibilities/stress and unable to control own reactions to stress. Problems in handling even daily routines (d230=3 or 4)
- 8 Not specified. Child is less than 6yrs.

**d3 COMMUNICATION****d310 Communicating - receiving spoken messages**

**Comprehending literal and implied meaning of messages spoken in language, e.g. understanding that a statement asserts a fact or is an idiomatic expression.**

**Probe & Rating Guidelines for d310**

Check b167 for delay in receptive lang., if any. If no delay, asks if X's understanding tends to be literal.

- 0 No difficulty. No delay in receptive language (see b167) and able to understand non-literal/inferential language use.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. no delay in receptive language (see b167), but understanding is literal.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. ability to understand (receptive language, b167) shows 2 yrs' delay.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. ability to understand (receptive language, b167) shows 3 yrs' delay.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. ability to understand (receptive language, b167) shows 4 yrs' delay or more.

**d315 Communicating - receiving non-verbal messages**

**Comprehending literal and implied meaning of messages conveyed by gestures, symbols and drawings e.g. realizing that a child is tired when she rubs her eyes.**

**Probe & Rating Guidelines for d210**

Does X have difficulties understandings non-verbal messages? What are the range/type of non-verbal gestures that child understands? (Give examples of imperative, descriptive and emotional gestures, and social cues).

- 0 No difficulty. Understands a wide range of gestures (imperative, descriptive) and some non-verbal cues (rubs eyes to mean 'tired', looks at watch to mean 'bored').
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. understands a wide range of gestures (descriptive & imperative), but not social cues.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. understands simple imperative gestures, e.g. "keep quiet", "go away", and "Come here".

- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. beginning to understand nods and shaking head to mean 'yes' and 'no'.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. does not understand simple gestures.

---

**d325 Communicating - receiving written messages**

**Comprehending literal and implied meaning of messages conveyed through written language e.g. following events in newspaper.**

**Probe & Rating Guidelines for d325**

Check b167 for delay in reading levels, if any. If no delay, asks if X's has difficulty with inferential questions in reading comprehension tasks in school.

- 0 No difficulty. No delay and able to understand non-literal/inferential language use.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. no delay in reading but understanding is literal.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. delay of 2 yrs in reading skills.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. delay of 3 yrs in reading.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. delay of 4 yrs or more in reading.
- 8 Not specified. Child under 6yrs.

---

**d330 Speaking**

**Producing words, phrases and longer passages in spoken messages in both literal and implied meaning e.g. expressing a fact or telling a story.**

**Probe & Rating Guidelines for d330**

Check b167 for delay in expressive lang., if any. If no delay, asks if X's use of lang. tends to be literal.

- 0 No difficulty. No delay and able to use non-literal/inferential language use.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. no delay in expressive language but use of words is literal.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. delay of 2 yrs in expressive language skills.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. delay of 3 yrs in expressive language.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. delay of 4 yrs or more in expressive language.

---

**d335 Producing non-verbal messages**

**Using gestures, symbols and drawings to convey messages.**

**Probe & Rating Guidelines for d335**

Does X have difficulties using non-verbal gestures? What are the range/type of non-verbal gestures that X uses, if any?

Give example of descriptive, imperative gestures.

- 0 No difficulty. Uses a wide range of gestures (imperative, descriptive) and some non-verbal cues (rubs eyes to mean 'tired', looks at watch to mean 'bored').
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. uses a wide range of gestures (descriptive & imperative), but not social cues.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. uses simple imperative gestures, e.g. "keep quiet", "go away", "Come here".
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. beginning to use nods and shaking head to mean 'yes' and 'no'. Points to objects from a distance.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. does not use gestures, rarely use pointing.

---

**d345 Producing Written Messages**

**Producing literal and implied meaning of messages that are conveyed through written language, e.g. writing a letter to a friend.**

**Probe & Rating Guidelines for d325**

Check b170 for delay in writing skills. If no delay, asks if X's has used words in a non-literal way in composition tasks in school or letters (e.g. use metaphors).

- 0 No difficulty. No delay and able to use non-literal/inferential language.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. no delay in writing but use of words is literal.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. delay of 2 yrs in writing skills.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. delay of 3 yrs in writing.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no writing skills.
- 8 Not specified. Child under 6yrs.

**d350 Conversation**

**Starting, sustaining and ending an interchange of thoughts and ideas carried out by means of spoken, written, sign or other forms of language, with one or more people one knows or who are strangers.**

**Probe & Rating Guidelines for d350**

Does X have difficulties engaging in conversations? How often/ in what contexts doe he face these difficulties? During conversations, does X often break rules of conversation? Whom does he converse with - family members or unfamiliar adults? Does he talk beyond his topic of interests?

- 0 No difficulty. Child is able to engage in conversations (beyond topic of own interest), with familiar adults (e.g. family, teachers) and others (e.g. visitors, strangers)
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child is able to engage in conversations (beyond topic of own interest), but sometimes violate rules of conversations (e.g. abruptly ending a conversation, abruptly changing from one topic to another without appropriate conversational signals/markers).
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child is able to engage in conversations with familiar people and strangers but only on topic of his interests.
- 3 Severe difficulty. Difficulties present more than 25% of the time, e.g. child participates in conversations with familiar people only, but on his own terms only. E.g. asks questions but not interested in the answers.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. not enough language to participate in conversation. Communicate needs only.

**d355 Discussion**

**Starting, sustaining and ending an examination of a subject matter, with arguments for and against, or debate carried out by means of spoken, written or other forms of language, with one or more people one knows or who are strangers, in formal or casual settings.**

**Probe & Rating Guidelines for d355**

Does X have difficulties participating in discussions? How often / in what contexts does he face these difficulties, e.g. with familiar/unfamiliar people? Can he participate in discussions beyond his interest topic?

- 0 No difficulty. Child is able to engage in discussions (self & other chosen topics) with familiar adults (e.g. family, teachers) and others (e.g. visitors, strangers).
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child is able to engage in discussions with familiar adults and others, but on self-chosen topic only.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child is able to engage in discussions with familiar people and strangers but only on topic of his interests.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. child participates in discussions with familiar people only, and only on topic of his interests, and on his own terms.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. child does not have enough language to participate in discussions, and able to communicate needs only.

8 Not specified. Child under 6 yrs.

**d360 Using communication devices and techniques**

Using devices, techniques and other means for the purpose of communicating, e.g. calling a friend on the telephone, sending emails.

**Probe & Rating Guidelines for d360**

Does X use the telephone? How / when does he use it and with whom?

- 0 No difficulty. Child is able to use phone to communicate with familiar and unfamiliar adults/people.
- 0 Achieved by 4yr.
- 1 Mild difficulty. Child able to use the phone, but only with familiar people.
- 2 Moderate difficulty. Child uses in a limited way, for e.g. able to answer calls and pass messages but does not use the phone to contact others.
- 3 Severe difficulty. Child uses the devices imitating other's behaviour/ learnt responses, i.e. not communicative, e.g. picks up the phone when it rings and says "hello", but not beyond that.
- 4 Complete difficulty. Not enough language. Communicates needs only.

**d4 MOBILITY**

**d410 Changing basic body positions**

Getting into and out of body position and moving from one location to another, e.g. getting up on a chair to lie down on a bed, and getting into out of position of kneeling or squatting.

**d415 Maintaining body position**

Staying in the same body position as required e.g. remaining seated or remaining standing for school or work.

**d420 Transferring oneself**

Moving from one surface to another, e.g. sliding along a bench or moving from a bed to a chair without changing position.

**d430 Lifting and carrying objects**

Raising up an object or taking something from one place or another, e.g. as when lifting a cup or carrying a child from one room to another.

**d435 Moving objects with lower extremities (legs and feet)**

Performing coordinated actions aimed at moving a object by using legs and feet, e.g. kicking a ball or pushing pedals.

**d440 Fine hand use**

Performing the coordinated actions of handling objects, picking up, manipulating and releasing them using one's hand, fingers, thumb, such as required to lift coins off a table or dial a knob.

**d445 Hand and arm use**

Performing the coordinated actions required to move objects or manipulate them by using hands, arms, such as when turning door handles or throwing or catching an object.

**d450 Walking**

Moving along a surface on foot, step by step, so that one foot is always on the ground, e.g. when strolling, sauntering, walking forwards, backwards or sideways.

**d455 Moving around**

Moving the whole body from one place to another by means other than walking such as climbing or a rock or running down a street, skipping, scampering, jumping, or running around obstacles.

**d460 Moving around in different locations**

Walking and moving around in various places and situations, such as walking between rooms in a house, within a building or down a street of a town.

**d465 Moving around using equipment**

**Moving the whole body from place to place, on any surface or space, by using specific devices designed to facilitate moving around or create other ways of moving around, e.g. wheelchair, walker, skates, ski, scuba equipment.**

**d470 Moving around using transportation**

**Using transportation to move around as passenger, such as being in a driven car or on a bus, rickshaw, private or public taxi, bus, train, subway, boat or aircraft.**

**Probe & Rating Guidelines for d410 to d470**

Does X have any difficulty in moving about, such as (specify each aspect of mobility, giving examples). (If difficulty is observed) Does he show this difficulty in all contexts (familiar & unfamiliar places?) What is the extent of difficulty and how much assistance does X needs to perform the activity adequately?

- 0 No difficulty. Child can perform the task independently.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child performs task independently, but slight problems in coordination.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child is able to perform tasks in familiar places (e.g. home) but not in unfamiliar surroundings.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. child requires some assistance in mobility in all contexts, e.g. needs supervision, verbal prompting.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. child relies on physical assistance from others/assistive devices for mobility.

---

**d2 SELF-CARE**

**d510 Washing self**

**Washing and drying one's whole body, or body parts, using water and appropriate cleaning and drying material and methods, e.g. bathing showering, washing hands & feet, drying with a towel.**

**Probe for d510**

Using developmental continuum below, estimate delay in washing self.

*Developmental Milestones (cf. S'pore Adaptive Beh Scales).*

- 3 yr Need physical help in washing body and cleaning body parts (e.g. face, hair).
- 5 yr Able to wash whole body independently but need some physical help in cleaning body parts (e.g. hair).
- 6 yr Able to wash and dry whole body independently but need some supervision in cleaning body parts (e.g. hair). Achieved by 6 yrs.
- 8 yr Able to wash and dry whole body and body parts (face, hair) independently

**Rating Guidelines for d510**

- 0 No difficulty. Age appropriate skills.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and less than 3 yrs' developmental level.

---

**d520 Caring for body parts**

**Looking after those parts of the body, such as skin, face, teeth, nails, scalp and genitals that require more than washing and drying.**

**Probe for d520**

Using developmental continuum below, estimate delay in caring for body parts.

*Developmental Milestones (cf. S'pore Adaptive Beh. Scales).*

- 4 yr Severe difficulty. Need physical help to care of body parts, but aware of need for cleanliness.
- 6 yr Needs supervision to care for body parts, and some / occasional physical help needed.

8 yr Able to care for body parts but occasionally needs some supervision (not physical help).

12 yr Able to care for body parts (cut nails, clean teeth) independently.

#### **Rating Guidelines for d520**

- 0 No difficulty. Age appropriate skills.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and less than 4 yrs' developmental level.

---

#### **d530 Toileting**

**Planning and carrying out the elimination of human waste (urination, defecation) and cleaning oneself afterwards.**

#### **Probe for d530**

Using developmental continuum below, estimate delay in washing self.

*Developmental Milestones (cf. Spore Adaptive Beh Scales)*

3 yr Dry in the day but needs diapers at night / occasional accidents at night. But able to indicate need.

5 yr Able to indicate need, and prepare self for urination / defecation (e.g. take clothes off. Choose appropriate place); but need physical help in cleaning afterwards.

6 yr Able to indicate need, and prepare self for urination / defecation (e.g. take clothes off, choose appropriate place); but need supervision/verbal prompts (not physical help) in cleaning afterwards.

8 yr Able to indicate need, prepare and clean self afterwards.

#### **Rating Guidelines for d530**

- 0 No difficulty. Age appropriate skills.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and less than 3 yrs' developmental level.

---

#### **d540 Dressing**

**Carrying out the coordinated actions and tasks of putting on and taking off clothes and footwear in sequence and in keeping with climatic and social conditions, e.g. putting on, adjusting and removing shirts, skirts, blouses, pants, undergarments, saris, kimono, tights, hats, gloves, coats, boots, sandals, slippers.**

*Inclusions: putting on or taking off clothes and footwear and choosing appropriate clothing.*

#### **Probe for d540**

Using developmental continuum below, estimate delay in washing self.

*Developmental Milestones (cf. Spore Adaptive Beh Scales).*

3 yr Able take off clothing, but needs help putting on clothes the right way and help in fastenings and footwear.

4 yr Able to put on and take off clothing, but needs help in fastenings and footwear.

8 yr Able to put on and take off clothing and footwear. Needs help to choose appropriate clothing.

10 yr Able to put on and take off clothing and footwear, able to choose appropriate clothing.

#### **Rating Guidelines for d540**

- 0 No difficulty. Age appropriate skills.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and less than 3 yrs' developmental level.

**d550 Eating**

**Carrying out the coordinated actions and tasks of eating food that has been served, bringing it mouth and consuming it in a culturally acceptable ways, cutting or breaking down food into pieces, opening bottles and cans, using eating implements, having meals, feasting or dining.**

**Probe for d550**

Using developmental continuum below, estimate delay in eating.

*Developmental Milestones (cf. S'pore Adaptive Beh Scales).*

- 1:6 yr Always has to be fed.
- 2 yr Feeds self with spoon, but messy and may need help with some food (e.g. soup).
- 5 yr Feeds self with fork or spoon.
- 8 yr Feeds self with fork and spoon (or chopsticks) but may need help with difficult food (e.g. meat chunks, fish with bones).
- 12 yr Able to help self at the table with all types of food independently.

**Rating Guidelines for d550**

- 0 No difficulty. Age appropriate skills.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and less than 1 yrs' developmental level.

**d560 Drinking**

**Taking hold of a drink, bringing it to the mouth, and consuming the drink in a culturally acceptable way, mixing, stirring and pouring liquids through straw or drinking running water such as from tap; feeding from the breast.**

**Probe for d560**

Using developmental continuum below, estimate delay in drinking.

*Developmental Milestones (cf. DDST - S'pore).*

- <8mth Needs special drinking container (e.g. bottle or feeding cup)
- 1 yr Drinks from cup with help.
- 1:6 yr Holds own cup but some spilling.
- 2 yr Holds own cup without spilling.
- 4yr Able to pour drink from jug/tap without help

**Rating Guidelines for d560**

- 0 No difficulty. Age appropriate skills.
- 1 Mild difficulty. Less than 2 yrs delay.
- 2 Moderate difficulty. Between 2 to 4 yrs delay.
- 3 Severe difficulty. More than 4 yrs delay.
- 4 Complete difficulty. More than 4 years delay and less than 1 yrs' developmental level.

**d6 DOMESTIC LIFE****d620 Acquisition of goods and services**

**Selecting procuring and transporting all goods and services required of daily living, including shopping and gathering daily necessities. Code 8 if under 6 yr.**

**Probe & Rating Guidelines for d620**

Does X have difficulties buying food from the school canteen/shops? How often and in what contexts e.g., unfamiliar settings, does he face these difficulties? Code 8 if child is under 6 yrs old.

- 0 No difficulty. Child can buy food and purchase goods from shops independently.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child able to perform the task with

supervision in specific contexts only e.g. school.

- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child able to perform the task with supervisions and in specific contexts only e.g. school canteen / school bookshop.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. with some physical assistance, child able to perform task briefly in familiar settings (e.g. point to food choice and give money).
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. child is completely reliant on physical assistance from others acquisition of goods.

---

#### **d630 Preparing meals**

**Planning, organizing, cooking, serving simple meals for oneself. Code 8 if under 6 yrs.**

##### **Probe & Rating Guidelines for d630**

Can X prepare simple meals for himself, e.g. make a drink, butter toast for breakfast, instant noodles? How much help does he need?

- 0 No difficulty. Child can perform the activity independently and through own initiative.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child able to perform the task with some supervision.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child able to perform the task with supervision in specific contexts only e.g. school during cooking lessons.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. child able to perform task briefly (in specific contexts only) and with some physical assistance, for e.g. during cooking lesson in school with close prompting from teacher.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. Child is completely reliant on physical assistance from others all food preparation.

---

#### **d640 Doing housework**

**Managing household by cleaning the house, washing clothes, using household appliances, storing food and disposing of garbage, such as by sweeping, mopping, washing counters, walls and other surfaces.**

##### **Probe & Rating Guidelines for d640**

Can X assist in doing household chores, e.g. sweeping, making beds, clearing table after meals? Code 8 if under 6yrs.

- 0 No difficulty. Child can perform the activity independently, as part of an assigned responsibility or through own initiative.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child able to perform the task with supervision.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child able to perform the task with supervision in specific contexts only e.g. in school during 'home economics' / 'daily living skills' lessons.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. child able to perform task in specific contexts only and with physical assistance, for e.g. during life skills lessons in school.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. child is not able to do any aspect of housework.

---

#### **d660 Assisting others**

**Assisting household members and others with their learning, communicating, self-care, movement, within or outside; being concerned about the well-being of household members and others.**

##### **Probe & Rating Guidelines for d510**

Does X help any other members of the family in any way? For example, walk/fetch younger sibling to/from school, assist sibling in homework or self care, assist parent to purchase of weekly groceries, assist grandparents? Code 8 if under 6yrs.

- 0 No difficulty. Child provides assistance to others as part of an assigned responsibility/ through own initiative.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child provides assistance of instructed to do so by parents/others.

---

2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child able to provide assistance if instructed and supervised by others.

3 Severe difficulty. Difficulties present more than 50% of the time, e.g. child able to provide assistance briefly, and close supervision and verbal prompts.

4 Complete difficulty. Difficulties present more than 95% of the time, e.g. child does not provide assistance to others.

---

## d7 INTERPERSONAL INTERACTIONS & RELATIONSHIPS

### d710 Basic interpersonal interactions

---

d7100 Show respect and warmth in relationship: Showing and responding to consideration and esteem, in a contextually and socially appropriate manner.

---

d7101 Appreciation in relationship: Showing and responding to satisfaction and gratitude in a contextually and socially appropriate manner.

---

d7102 Tolerance in relationship: showing and responding to understanding and acceptance of behaviour, in a contextually and socially appropriate manner.

---

d7103 Criticism in relationship: providing and responding to implicit and explicit differences of opinion or disagreement, in a contextually and socially appropriate manner.

---

d7104 Social cues in relationship : Giving and reacting appropriately to signs and hints that occur in social interactions.

---

d7105 Physical contact in relationship : making and responding to bodily contact with others in a contextually and socially appropriate manner.

---

#### Probe & Rating Guidelines for d7100

Does X have difficulties showing affection spontaneously, e.g. says 'I Love You', give hugs to you/other family members? Does he greet or show pleasure when you come home from work? How often / on what contexts does he have face difficulties?

- 0 No difficulty. Able to show warmth/respect spontaneously, in appropriate contexts, in familiar and unfamiliar settings.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. occasionally shows respect / warmth spontaneously, in familiar settings only.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. shows respect / warmth as a learnt response with familiar people, in familiar settings.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. if prompted, would sometimes show respect / warmth as a learnt response, e.g. responds appropriately to 'Give mummy a hug/kiss'.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no social interaction, or relationship based on communicating own needs only.

#### Probe & Rating Guidelines for d7100

Does X have difficulties showing appreciation, e.g. saying 'thank you' or showing gratitude in other ways (e.g. hugs) when you give him something that he likes? How often / in what contexts?

- 0 No difficulty. Able to show appreciation spontaneously, in appropriate contexts.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. occasionally shows appreciation spontaneously, but only in familiar contexts/with familiar people.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. shows appreciation as a learnt response.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. If prompted, able to show appreciation as a learnt response, e.g. responds appropriately to cue 'Say Thank You'.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no social interaction, or relationship based on communicating own needs only.

#### Probe & Rating Guidelines for d7102

Does X get angry or unhappy when others behave differently from him, or don't agree with his choice / opinion? How does he react?

- 0 No difficulty. Able to show tolerance in appropriate contexts.

- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. shows tolerance if others' behaviour have been explained.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. occasionally gets angry if others behave or think differently from him/herself.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. often gets angry if others behave or think differently.
- 4 Complete difficulty. Difficulties present more than 50% of the time, e.g. no social interaction, or relationship based on communicating own needs only.

#### **Probe & Rating Guidelines for d7103**

Does X have difficulty accepting criticism? How does he react when he loses in games? How does he react when he gets answers wrong at school?

- 0 No difficulty. Able to accept criticism and losing in appropriate contexts.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. beginning to accept if he loses a game or if things are broken/imperfect.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. cannot accept if he gets things wrong, or loses a game, or things are 'imperfect'. Gets angry or annoyed.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. marked intolerance (e.g. throws tantrum) if he gets things wrong, or loses a game, or things are 'imperfect'.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no social interaction, relationship based on communicating own needs only. No concept of losing/winning.

#### **Probe & Rating Guidelines for d7104**

Does X have difficulties in reading social cues, e.g. knows (through non-verbal cues, tone of voice) when others are bored, impatient or angry with him? Can he understand/respond to more complex ones, e.g. looking at watch to show impatience?

- 0 No difficulty. Able to understand and react to appropriate social cues (simple & complex).
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. able to react to simple social cues, e.g. knows when others are angry / unhappy through tone and gestures.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. able to react to basic/simple social cues as part of learnt response, e.g. greets and extends handshakes when introduced to others / receiving visitors to the house.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. reaction to basic social cues (as part of learnt response reaction) often inconsistent/sometime inappropriate, e.g. shakes hand too hard when receiving visitors, or only when prompted and cued by parents/others.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no social interaction. No awareness of social consequence.

#### **Probe & Rating Guidelines for d7105**

Does X have difficulties with physical contact, e.g. does he enjoy being cuddled, or hug you in return? Does he get angry when you try and hold him?

- 0 No difficulty. Enjoys cuddles and reciprocates (appropriately)
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. enjoys cuddles and reciprocates, but at times inappropriately, e.g. hugs friends/acquaintances, or hugs too hard.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. enjoys cuddles, 'allows' self to be hugged but does not reciprocate.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. reacts badly to physical contact, e.g. screams, gets angry, reacts as if he's been hurt.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no social interaction, no awareness of social consequence.

#### **Probe & Rating Guidelines for d7105**

Does X have difficulties with physical contact, e.g. does he enjoy being cuddled, or hug you in return? Does he get angry when you try and hold him?

- 0 No difficulties.

- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. enjoys cuddles and reciprocates, but at times inappropriately, e.g. hugs friends/acquaintances, or hugs too hard.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. enjoys cuddles, 'allows' self to be hugged but does not reciprocate.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. reacts badly to physical contact, e.g. screams, gets angry, reacts as if he's been hurt.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no social interaction, no awareness of social consequence.

---

**d720 Complex interpersonal interactions**

**d7200 Forming relationships:** beginning and maintaining interactions with others for a short or long period of time, such as introducing oneself, finding and establishing friendship and professional relationship, starting a relationship that may be permanent, romantic intimate.

**d7201 Terminating relationship:** bringing relationships to a close in a contextually and socially appropriate manner, e.g. by ending temporary relationship at the end of a visit, ending long-term relationship with a friend who has moved to a new town.

**d7202 Regulating behaviours within an interaction:** Regulating emotions and impulses, verbal aggression in interactions with others, in a contextually and socially appropriate ways.

**d7203 Interacting according to social rules:** Acting independently in social interactions and complying with social conventions governing one's role, positions or social status in interacting with others.

**d7204 Maintaining social space:** Being aware of and maintaining a distance between oneself and others that is contextually, socially and culturally appropriate.

---

**Probe for d7200**

Does X have difficulties initiating interactions with peers in school? When X is at the playground, does he have problems initiating play with other children? If under 6yrs, code 8.

**Probe for d7201**

Does X gets upset when close peers/cousins leave after a visit to your place?

**Probe for d7202**

Does X behave aggressively to others, physically or verbally?

**Probe for d7203**

Does X have difficulties when interacting independently in social situations? Does he tend to behave more like a younger child, or an adult?

**Probe for d7204**

Does X have difficulty maintaining social distance? Does he tend to sit/stand too close to others?

**Rating Guidelines for d7200 to 7204**

- 0 No difficulty. Child able to show adequate interpersonal skills in familiar and unfamiliar contexts.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child shows difficulty when interacting with unfamiliar people/contexts.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child shows difficulty when interacting in unstructured situations, but OK in context where social roles are clear and structured, e.g. school, social skills group.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. has difficulty in most social interaction contexts, can manage basic interpersonal interactions only (if d710 is 2 or 3, code 3 for d720)
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. interaction/relationship with family members limited to expression of own needs.

---

**d730 Relating with strangers**

**Engaging in temporary contacts and links with strangers for specific purposes, e.g. when asking for directions or making public purchase.**

---

### Probe & Rating Guidelines for d730

Does X have difficulties relating to strangers, e.g. shopkeepers? Can he pay for things in the neighborhood shop, if supervised?

- 0 No difficulty. Able to relate to strangers in all contexts, independently.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. able to relate to strangers in familiar settings only, e.g. when visitors come to house, purchasing things from neighborhood shop.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. able to relate to strangers in familiar settings only, and only if supervised, e.g. given verbal prompts.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. only able to relate to strangers in a learnt manner, greets visitors, people in neighbourhood, when prompted by parent.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no interaction with strangers, ignores/avoids unfamiliar people.
- 8 Not specified. Child under 6 yrs old.

---

### d740 Formal relationships

**Creating and maintaining specific relationships in formal settings, e.g. with employers, professionals, including persons in authority (e.g. school principals).**

### Probe & Rating Guidelines for d740

Does X have a special friend? What is quality of the friendship? Does he take on a passive role, or is the friendship based on sharing of obsession/interest topic? Does he have any concept of friendship?

- 0 No difficulty. Has special friends or friends, and friendship is reciprocal.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. has special friend but friendship is based on sharing of same obsessional interest topic/toys.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. has friends in structured situations only, e.g. assigned buddy in school, social skills groups, and takes on purely passive role.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. no friends but has concept of friendship. But unable / unsuccessful in making/maintaining friends.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no concept of friendship, no special friends.

---

### d760 Family relationships

**d7600 Creating and maintaining kinship relationships, such as with members of nuclear family, extended family, foster and adopted family and step-relationships, more distant relationships e.g. cousins or legal guardians.**

**d7601 Child-parent relationship : Creating and maintaining relationships with one's parent, e.g. a young child obeying his/her parents.**

**d7602 Sibling relationship : Creating and maintaining a brotherly or sisterly relationship with a person.**

**d7203 Extended family relationship : Creating and maintaining a family relationship with one's extended family, e.g. cousins, uncles, aunts and grandparents.**

### Probe & Rating Guidelines for d7600 to d7203

How is X's relationship with these family members? Does he take on a passive role? Does he show concern for them? Does he show preference for family members over other visitors?

- 0 No difficulty. Child shows reciprocal social relationship with parent/sibling/extended family.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. child takes on a mainly passive role in relationship with parent/sibling/extended family.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. child's relationship with parent/sibling/extended family if often one-sided or on X's own terms only.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. child is aware of and shows preference for parent/sibling/extended family members as carers. No interactions.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no preference for family members

---

---

**d8 MAJOR LIFE AREAS****d810 Informal education**

Learning at home or in some other non-institutional setting, e.g. learning crafts and other skills from parents or family members or home schooling.

**d815 Preschool education**

Learning at an initial level of organized instruction, designed primarily to introduce the child to school-type environment and prepare for compulsory education.

**d820 School education**

Gaining admission to school, education, engaging in all school-related responsibilities and privileges, and learning course materials, subjects and other curriculum requirements in a primary or secondary education program, including attending school regularly, working cooperatively with other students, taking direction from teachers, completing assignments and tasks.

**Probe & Rating Guidelines for d810 to d820**

Does X have difficulty engaging in learning activities at home, e.g. academic skills, or learning to use tools (e.g. bicycle, computers, cooking) ? Does X have any difficulties participating in activities in preschool or school?

- 0 No difficulty. Able to engage in all activities, appropriate to age/ability level.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. excluded from a few activities (e.g. exclusion from PE for safety reasons, MT exemptions), mainly due to ability factors.
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. excluded from some activities due to ability and behavioural limitations, for e.g. not able to complete work assignments/tasks.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. excluded from most activities (not able to participate/complete activity/tasks for most areas).
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. does not attend school/suspended.

---

**d9 COMMUNITY, SOCIAL AND CIVIC LIFE****d910 Community life**

Engaging in all aspects of community social life, e.g. engaging in charitable organization, service clubs (achieved by 12 yrs, if younger, code 9).

**d915 Recreation and leisure**

Engaging in any form of play, recreational or leisure activity, e.g. informal or organized play or sport, programs of amusement, e.g. cinemas, theatres; engaging in crafts or hobbies, sightseeing, tourism, sightseeing for pleasure.

**d920 Religion ad spirituality**

Engaging in spiritual or religious activity, organizations or practices for self-fulfillment, establishing connections with divine power, e.g. involved in attending church, temple or mosque for religious purposes. (If under 6yrs, code 8)

**Probe & Rating Guidelines for d910 to d920**

Does X engage in this activity? Is he a passive participant or actively enjoys the experience? What are the main difficulties that he encounters?

- 0 No difficulty. Able to engage successfully and enjoy all appropriate activities.
- 1 Mild difficulty. Difficulties present less than 25% of the time, e.g. joins in and participates in some activities, although not always successful
- 2 Moderate difficulty. Difficulties present less than 50% of the time, e.g. participates in some activities only, but mainly as passive participant.
- 3 Severe difficulty. Difficulties present more than 50% of the time, e.g. joins in one/two activities only, but does not participate.
- 4 Complete difficulty. Difficulties present more than 95% of the time, e.g. no activities outside of home/school contexts.
- 9 Not relevant. Child's family does not engage in these activities, e.g. no religious affiliation.

**e1 PRODUCTS AND TECHNOLOGY**

**e110 Products and substances for personal consumption (Food/Diet & Drugs).**

**Any natural or man-made object or substance gathered, processed or manufactured for ingestion, including food and drugs.**

**Probe for e110**

Does he consume any special food, diet or drugs to help him alleviate/reduce his problems? What are these and what are they for? How much has these drugs/food substances reduced his problems? How often does he consume them? Does he have any problems linked to diet/drugs? How much disruption does this problem caused? How often do they occur?

**Rating Guidelines for e110 (Facilitator)**

- 0 No facilitator. Additional consumption mainly food supplements to increase general health e.g. fish oil, vitamins
- 1 Slight/mild facilitator. Child relies on special food/diet/drugs less than 25% of the time (one or two times a month).
- 2 Moderate facilitator. Child relies on special food/diet/drugs less than 50% of the time (once a week).
- 3 Strong facilitator. Child relies on special food/diet/drugs more than 50% of the time (3-4 times a week), and has reduced child's problems partially (50%).
- 4 Complete facilitator. Child relies on special food/diet/drugs more than 95% of the time (daily), and has reduced child's problems significantly (more than 75%).

**Rating Guidelines for e110 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Problem associated with food/diet/drugs encountered less than 25% of the time (one to two times a month).
- 2 Moderate barrier. Problem associated with food/diet/drugs encountered less than 50% of the time (once a week).
- 3 Strong barrier. Problem associated with food/diet/drugs encountered more than 50% of the time (3-4 times a week), and partially disrupting child's life (50%).
- 4 Complete barrier. Problem associated with food/diet/drugs encountered more than 95% of the time (daily), and totally disrupting child's day-to-day life.

**e125 Products and technology for communication**

**Equipment, products and technologies used by people in activities of sending and receiving information, including those adapted or specially designed, located in, or on or near the person using them.**

**Probe for e120**

Does he use any devices or special materials to communicate? E.g. PEC cards, picture key-boards, visual cue cards? Do you use any of these materials when communicating with him? If so, in how often and in what contexts? Are these used in the school or in the home or only when teaching new skills?

**Rating Guidelines for e125 (Facilitator)**

- 0 No use of any products/adaptive devices/technology for communication, uses speech only.
- 1 Slight/mild adaptation. Child use communication devices less than 25% of the time (i.e. once/twice a month, for e.g., PEC cards when introduced to new environment/routines).
- 2 Moderate adaptation. Child uses adapted communication devices less than 50% of the time (once weekly), e.g. in most weekly sessions with ST in school.
- 3 Strong adaptation. Child communicates using adapted communication more than 50% of the time (2-3 times a week), e.g. for learning tasks in most activities in school.
- 4 Complete adaptation. Child communicates using adapted communication devices more than 95% of the time (daily), e.g. in most settings/occasions.

**Rating Guidelines for e125 (Barrier)**

- 0 No barriers/difficulties in using communication products and technology.

- 1 Mild barrier. Difficulties in using communication products experienced less than 25% of the time (1-2 times a month).
- 2 Moderate barrier. Difficulties in using communication products experienced less than 50% of the time (1-2 times a week).
- 3 Strong barrier. Difficulties in using communication products experienced more than 50% of the time (3-4 times a week).
- 4 Complete barrier. Difficulties in using communication products experienced more than 95% of the time (daily), and causing total disruption to child's day-to-day life.

---

**e130 Products and technology for use in education.**

**Equipment, products, processes, methods, technology used for acquisition of knowledge, expertise or skill, including those adapted or specially designed. Inclusion: books, manuals, toys, computer hardware & software.**

**Probe for e130**

Confirm child's current school, curriculum used in school. In his group, does he use the same materials as other in the class? To what extent are the materials used (e.g. topics covered, materials/syllabus, software) tailored to his interests/ability levels? Is he receiving educational support outside school (e.g. tuition / enrichment classes). What materials/methods are used in these classes? Are they adapted to meet his needs?

**Rating Guidelines for e130 (Facilitator)**

- 0 No use of any adaptive devices/products/technology in education.
- 1 Slight/mild adaptation. Child use adaptive technology/products in education less than 25% of the time (i.e. once/twice a month, for e.g., PEC cards when introduced to new environment/routines).
- 2 Moderate adaptation. Child uses adapted technology/products devices less than 50% of the time (once weekly), e.g. in weekly sessions with support teacher in school.
- 3 Strong adaptation. Child uses adapted technology/products more than 50% of the time (2-3 times a week), e.g. for learning tasks in most activities in school.
- 4 Complete adaptation. Child communicates using adapted communication devices more than 95% of the time (daily), e.g. in most settings/occasions.

**Rating Guidelines for e130 (Barrier)**

- 0 No barriers/difficulties in using products and technology in education.
- 1 Mild barrier. Difficulties in using products/technology in education experienced less than 25% of the time (1-2 times a month).
- 2 Moderate barrier. Difficulties in using products/technology in education experienced less than 50% of the time (1-2 times a week).
- 3 Strong barrier. Difficulties in using products/technology in education experienced more than 50% of the time (3-4 times a week).
- 4 Complete barrier. Difficulties in using products/technology experienced more than 95% of the time (daily).

---

**e140 Products and technology for culture, recreation and sport.**

**Equipment, products and technology used for the conduct and enhancement of cultural, recreational, and sporting activities, including those adapted or specially designed.**

**Probe for e140**

Confirm the type of sports/recreation activities child participates. Does he use any special equipment/materials that helped in to participate better in these activities.

**Rating Guidelines for e140 (Facilitator)**

- 0 No adaptation or does not participate in such activities.
- 1 Slight/mild adaptation. Child use adaptive products for recreation less than 25% of the time.
- 2 Moderate adaptation. Child uses adapted products devices less than 50% of the time.
- 3 Strong adaptation. Child uses adapted products more than 50% of the time, e.g. consistently in 1 setting, e.g. school OR home.

4 Complete adaptation. Child uses adaptive products more than 95% of the time (daily), e.g. in most settings/occasions, at home and school.

**Rating Guidelines for e140 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Difficulties in using products experienced less than 25% of the time. Difficulties occur but only occasionally.
- 2 Moderate barrier. Difficulties in using products experienced less than 50% of the time. Difficulties occur occasionally in one setting (e.g. for some recreational activities).
- 3 Strong barrier. Difficulties in using products/technology for recreation experienced more than 50% of the time. E.g. difficulties occur all the time in one setting, i.e. home OR school.
- 4 Complete barrier. Difficulties in using products/technology experienced more than 95% of the time, difficulties occur all the time in all settings, i.e. at home and school.

---

**e145 Products and technology for the practice of religion or spirituality.**

**Equipment, products and technology, unique or mass-produced, that are given or take on a symbolic meaning in the context of the practice of religion or spirituality, including those adapted or specially designed.**

**Probe for e145**

Confirm the types of religious activities that child participates, if any. Does he use any special materials that helped in to participate better in these activities?

**Rating Guidelines for e145 (Facilitator)**

- 0 No adaptation needed.
- 1 Slight/mild adaptation. Child use adaptive technology/products less than 25% of the time.
- 2 Moderate adaptation. Child uses adapted technology/products less than 50% of the time.
- 3 Strong adaptation. Child uses adapted technology/products more than 50% of the time.
- 4 Complete adaptation. Child uses adaptive products/technology more than 95% of the time.

**Rating Guidelines for e145 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Difficulties in using products/technology experienced less than 25% of the time. Difficulties occur but only occasionally.
- 2 Moderate barrier. Difficulties in using products/technology for recreation experienced less than 50% of the time, e.g. for some recreational activities.
- 3 Strong barrier. Difficulties in using products/technology experienced more than 50% of the time. E.g. difficulties occur all the time in one setting, i.e. home OR school.
- 4 Complete barrier. Difficulties in using products/technology experienced more than 95% of the time, difficulties occur all the time in all settings, i.e. at home and school.

---

**e150 Design, construction and building products & technology of building for public use**

**Products and technology that constitute and individual's indoor and outdoor human-made environment that is planned, designed and constructed for public use.**

**Probe for e150**

Does he have problems participating in activities in public buildings, e.g. shopping centers, school? Have any adaptation been made school buildings that has helped him to move about independently? E.g. visual signs/cues for toilets, classrooms, canteen, his seat in classroom. Is the classroom organized differently to accommodate his needs, e.g. more structure in organization of materials / activities areas.

**Rating Guidelines for e150 (Facilitator)**

- 0 No adaptation needed.
- 1 Slight/mild adaptation. Child use adaptive technology/products less than 25% of the time.
- 2 Moderate adaptation. Child uses adapted technology/products less than 50% of the time.
- 3 Strong adaptation. Child uses adapted technology/products more than 50% of the time.

4 Complete adaptation. Child uses adaptive products/technology more than 95% of the time (daily), e.g. in most occasions.

**Rating Guidelines for e150 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Difficulties in using products/technology experienced less than 25% of the time. Difficulties occur but only occasionally.
- 2 Moderate barrier. Difficulties in using products/technology for recreation experienced less than 50% of the time. Difficulties occur occasionally in one setting (e.g. for some recreational activities).
- 3 Strong barrier. Difficulties in using products/technology experienced more than 50% of the time. E.g. difficulties occur all the time in one setting, i.e. home OR school.
- 4 Complete barrier. Difficulties in using products/technology experienced more than 95% of the time, difficulties occur all the time in all settings, i.e. at home and school.

---

**e155 Products & technology that constitute an individual's indoor and outdoor human-made environment that is designed and constructed for private use (e.g. home).**

**Probe for e155**

Have any adaptation been made to the home to help him? E.g. Organization of furniture, special rooms, visual signs/cues, lighting, temperature control. Does he face any particular difficulties associated with the design/construction of the home?

**Rating Guidelines for e155 (Facilitator)**

- 0 No adaptation needed.
- 1 Slight/mild adaptation. Child uses adaptive products less than 25% of the time.
- 2 Moderate adaptation. Child uses adapted products less than 50% of the time.
- 3 Strong adaptation. Child uses adapted products more than 50% of the time.
- 4 Complete adaptation. Child uses adaptive products more than 95% of the time (daily), e.g. in most occasions.

**Rating Guidelines for e155 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Difficulties in using products experienced less than 25% of the time. Difficulties occur but only occasionally.
- 2 Moderate barrier. Difficulties in using products less than 50% of the time.
- 3 Strong barrier. Difficulties in using products more than 50% of the time.
- 4 Complete barrier. Difficulties in using products more than 95% of the time, difficulties occur all the time in all settings.

---

**e2 NATURAL ENVIRONMENT AND HUMAN MADE CHANGES**

**e210 Physical geography.**

**Features of landforms and bodies of water.**

**e215 Population**

**Groups of people living in a given environment who share same pattern of environmental adaptation.**

**e220 Flora and fauna**

**Plants and animals**

**e225 Climate**

**Meteorological events, such as weather, including : temperature, humidity, atmospheric pressure, precipitation, wind, seasonal variation.**

**e230 Natural events**

**Geographical and atmospheric changes that cause disruption in an individual's physical environment, occurring regularly or irregularly, e.g. earthquakes, tornadoes, lightning, violent weather conditions.**

**e235 Human caused events**

	<b>Alterations or disturbances in natural environment, caused by humans, that may result in the disruption of people's day to day lives including events or conditions links to wars e.g. displacement of people, destruction of social infrastructure, homes, lands, water &amp; air pollution.</b>
e240	<b>Light</b> Electromagnetic radiation by which things are made visible by either sunlight or artificial lighting (e.g. candles, oil, paraffin lamps, fires and electricity), and which may provide useful or distracting information about the world.
e245	<b>Time-related changes</b> Natural, regular or predictable temporal changes, e.g. night/day cycles, lunar calendar.
e250	<b>Sound</b> A phenomenon that is or may be heard, such as banging, ringing, thumping, singing, whistling or buzzing, in any volume, timbre or tone, and that may provide useful or distracting information about the world. Inclusion: sound intensity and quality.
e255	<b>Vibration</b> Regular or irregular to and fro motion of an object or an individual caused by physical disturbance, such as shaking, quivering, quick jerky movements of things, explosion.
e260	<b>Air quality</b> Characteristic of the atmosphere (outside buildings) which may provide useful or distracting information about the world.
<b>Probe and Rating Guidelines for e210 to e260</b>	
Any aspect of the physical geography, climate, or urban environment in this country that is causing any concern for him? How has this limited his participation in life activities here?	
<ol style="list-style-type: none"> <li>0 Problems in this aspect of the environment, although present, does not pose any barrier.</li> <li>1 Barriers may occur, but infrequently, e.g. 1-2 times a month.</li> <li>2 Barriers occasionally occur in one life area, e.g. 1-2 times a week.</li> <li>3 Barriers frequently occur in one life activity, e.g. mobility (3-4 times a week)</li> <li>4 This aspect pose significant barriers to child in several life activities, e.g. daily</li> </ol>	

### e3 SUPPORT AND RELATIONSHIPS

*Note: This section refers to individuals that provide direct support to the child. Focus is on the quality and accessibility of the support.*

e310	<b>Immediate family</b>
e315	<b>Extended family</b>
e320	<b>Friends</b>
e325	<b>Acquaintances, peers, colleagues, community members</b>
e330	<b>People in positions of authority, e.g. teachers, principals.</b>
e340	<b>Personal care providers and personal assistants. Include in-class support, full-time maids, esp those that employed mainly for the child.</b>
e345	<b>Strangers. Include temporary teachers who do not know the child, neighborhood shopkeeper.</b>
e355	<b>Health professionals, including all service providers within health systems, including speech therapist, docs, nurses, medical social workers.</b>

**e360 Other professionals, e.g. private tutor, teachers, psychologists, counselors, guidance teachers.**

**Probe and Rating Guidelines for e310 to e360**

Who are the people directly involved in giving him support. For each, how frequent and intensive is the support? Are there any difficulties in getting the level of support that he needs from these people? (Nb. difficulties can be due to limited availability, high cost, lack of trained personnel).

**Rating Guidelines for e310 to e360 (Facilitator)**

- 0 No facilitator. Child does not rely direct support from these individuals.
- 1 Mild facilitator. Child receives support from these individuals less than 25% of the time (e.g. once a month).
- 2 Moderate facilitator. Child receives support from these individuals less than 50% of the time (e.g. 1-2 times a week).
- 3 Strong facilitator. Child receives support from these individuals less than 50% of the time (e.g. 1-2 times a week).
- 4 Complete support. Support by trained individuals and is consistent, easily accessible and intensive (3X per week or more).

**Rating Guidelines for e310 to e360 (Barrier)**

- 0 No barrier. No difficulties receiving support from these individuals.
- 1 Mild barrier. Child faces difficulties in getting support from these individuals less than 25% of the time.
- 2 Moderate barrier. Child faces difficulties in getting support from these individuals less than 50% of the time.
- 3 Strong barrier. Child faces difficulties in getting support from these individuals more than 50% of the time.
- 4 Complete barrier. Child faces difficulties in getting support from these individuals more than 95% of the time.

---

**E4 ATTITUDES (Note: Attitudes of other people, i.e. not the child's own attitudes/beliefs)**

**Probe for e410 to e460**

How have the opinions and beliefs of others abt X affected their behaviour towards him? Are the attitudes helpful or has it caused problems? How often has X encountered such attitude? How widespread is the attitude (e.g. who shares the attitude)? Note: if child has no extended family, no personal care providers or no contact with authority figures/health professionals etc, code 9 (i.e. irrelevant).

**e410 Individual attitudes of immediate family**

**Rating Guidelines for e410 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time, for example in a few family members only.
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time, for example in some family members.
- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time, for example in most family members.
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time, for example in all family members.

**Rating Guidelines for e410 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time, e.g. in a few family members only.
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time, e.g. in some family members.
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time, e.g. in most family members.
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time, e.g. in all family members.

---

**e415 Individual attitudes of extended family.**

**Rating Guidelines for e415 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time, for example in a few extended family members only, e.g. some adults.
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time, for example in some extended family members, most adults.
- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time, for example in most extended family members (adults & children).
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time, for example in all extended family members.

**Rating Guidelines for e415 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time, for example in a few extended family members only, e.g. some peers.
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time, for example in some extended family members, e.g. most children/peers.
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time, for example in most extended family members (adults and peers).
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time, for example in all extended family members.

---

**e420 Individual attitudes of friends**

**Rating Guidelines for e420 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time, i.e. in a few of child's friends, e.g. friends in social-skills training/circle of friends group in school.
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time, e.g. friends in class.
- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time, for example most of friends in school.
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time, for example most friends in and outside school.

**Rating Guidelines for e420 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time, for example in a few friends in class/school.
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time, for example some friends in class/school.
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time, for example in most friends in school.
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time, for example in all friends in and outside school.

---

**e425 Individual attitudes of acquaintances, peers, community members**

**Rating Guidelines for e425 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time, i.e. among a few of child's acquaintances in familiar environment, e.g. neighborhood.
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time, i.e. among most of child's acquaintances in familiar environment, e.g. neighborhood.

- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time, i.e. all of child's acquaintances in familiar environment, e.g. neighborhood.
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time, in familiar and unfamiliar environments.

**Rating Guidelines for e425 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time, for example in a few acquaintances in unfamiliar environments..
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time, i.e. among a few of child's acquaintances in familiar environment, e.g. neighborhood.
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time, i.e. among most of child's acquaintances in familiar environment, e.g. neighborhood.
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time, i.e. among all of child's acquaintances in familiar and unfamiliar environments.

**e430 Individual attitudes of people in positions of authority, e.g. principals.**

**Rating Guidelines for e430 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time (once a week).
- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time (daily)

**Rating Guidelines for e430 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time (once a week).
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time (daily).

**e440 Individual attitudes of personal care providers and personal assistants (e.g. in-class support, nannies, full-time maid and others).**

**Rating Guidelines for e440 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time (once a week).
- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time (daily)

**Rating Guidelines for e440 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time (once a week).
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time (daily).

**e445 Individual attitudes of strangers**

**Rating Guidelines for e445 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time (once a week).

- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time (daily)

**Rating Guidelines for e445 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time (once a week).
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time (daily).

---

**e450 Individual attitudes of health professionals**

**Rating Guidelines for e450 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time (once a week).
- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time (daily)

**Rating Guidelines for e450 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time (once a week).
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time (daily).

---

**e455 Other professionals, e.g. tutors, private teachers, psychologists, counselors, outreach advisors.**

**Rating Guidelines for e455 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time (once a week).
- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time (daily)

**Rating Guidelines for e455 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate barrier. Negative attitudes encountered less than 50% of the time (once a week).
- 3 Strong barrier. Negative attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete barrier. Negative attitudes encountered more than 95% of the time (daily).

---

**e460 Societal attitudes, general / specific opinions and beliefs held by people about ASD or special needs. Probe for examples of direct experience that reflect the attitude.**

**Rating Guidelines for e460 (Facilitator)**

- 0 No facilitator.
- 1 Mild facilitator. Positive attitudes encountered less than 25% of the time (1-2 times a month).
- 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time (once a week).
- 3 Strong facilitator. Positive attitudes encountered more than 50% of the time (2-3 times a week).
- 4 Complete facilitator. Positive attitudes encountered more than 95% of the time (daily)

**Rating Guidelines for e460 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative attitudes encountered less than 25% of the time (1-2 times a month).

---

2 Moderate barrier. Negative attitudes encountered less than 50% of the time (once a week).  
 3 Strong barrier. Negative attitudes encountered more than 50% of the time (2-3 times a week).  
 4 Complete barrier. Negative attitudes encountered more than 95% of the time (daily).

---

**e465 Social norms, practices, and ideologies**

**Rating Guidelines for e465 (Facilitator)**

0 No facilitator.  
 1 Mild facilitator. Positive attitudes encountered less than 25% of the time (1-2 times a month).  
 2 Moderate facilitator. Positive attitudes encountered less than 50% of the time (once a week).  
 3 Strong facilitator. Positive attitudes encountered more than 50% of the time (2-3 times a week).  
 4 Complete facilitator. Positive attitudes encountered more than 95% of the time (daily)

**Rating Guidelines for e465 (Barrier)**

0 No barrier.  
 1 Mild barrier. Negative attitudes encountered less than 25% of the time (1-2 times a month).  
 2 Moderate barrier. Negative attitudes encountered less than 50% of the time (once a week).  
 3 Strong barrier. Negative attitudes encountered more than 50% of the time (2-3 times a week).  
 4 Complete barrier. Negative attitudes encountered more than 95% of the time (daily).

**e5 SERVICE SYSTEMS AND POLICIES**

**e580 Health services, systems and policies**

**Probe and Rating Guidelines for e580**

Based on X's experiences, to what extent have the health services, systems and policies has been able to meet his needs? Have any aspects of the health services been accommodated for meet his special needs?

**Rating Guidelines for e580 (Facilitator)**

0 No facilitator.  
 1 Mild facilitator. Positive experiences with health services than 25% of the time (e.g. in one or two aspects of health services).  
 2 Moderate facilitator. Positive experiences with health services encountered less than 50% of the time (e.g. most aspects of one health services).  
 3 Strong facilitator. Positive experiences with health services more than 50% of the time (e.g. most aspects in health services, such as assessment, intervention).  
 4 Complete facilitator. Positive experiences with health services encountered more than 95% of the time.

**Rating Guidelines for e580 (Barrier)**

0 No barrier.  
 1 Mild barrier. Negative experiences with health services encountered less than 25% of the time.  
 2 Moderate barrier. Negative experiences with health services encountered less than 50% of the time.  
 3 Strong barrier. Negative experiences with health services encountered more than 50% of the time.  
 4 Complete barrier. Negative experiences with health services encountered more than 95% of the time.

---

**e585 Education and training services, systems and policies**

**Probe and Rating Guidelines for e585**

Based on X's experiences, to what extent have the education services, systems and policies has been able to meet his needs? Have any aspects of the education services/system been accommodated for meet his special needs?

**Rating Guidelines for e585 (Facilitator)**

0 No facilitator.  
 1 Mild facilitator. Positive experiences with education services than 25% of the time (e.g. in one or two aspects of education services)  
 2 Moderate facilitator. Positive experiences with education services encountered less than 50% of the time (e.g. most aspects of one education services)

- 3 Strong facilitator. Positive experiences with education services more than 50% of the time (e.g. most aspects in education services, such as assessment, intervention)
- 4 Complete facilitator. Positive experiences with education services encountered more than 95% of the time.

**Rating Guidelines for e585 (Barrier)**

- 0 No barrier.
- 1 Mild barrier. Negative experiences with education services encountered less than 25% of the time.
- 2 Moderate barrier. Negative experiences with education services encountered less than 50% of the time.
- 3 Strong barrier. Negative experiences with education services encountered more than 50% of the time.
- 4 Complete barrier. Negative experiences with education services encountered more than 95% of the time.

---

END

## APPENDIX D

### STUDY 1 : SAMPLE INVITATION LETTER AND PARENTAL INFORMED CONSENT

To Parents,  
Thru'  
(Name of Principal & school/institution)

Dear parent,

#### Re : Invitation to Participate in Research on Autistic Spectrum Disorder

I am a Chartered Educational Psychologist from the Ministry of Education and currently conducting a research on the assessment of special educational needs of children with autistic spectrum disorder (ASD) in Singapore. The aim of this study is to identify the needs of children with ASD in Singapore schools. This research is undertaken as part of a PhD programme in Psychology at University College London<sup>40</sup>, and is funded by the Ministry of Education.

I would like to invite you and your child to participate in this study. Your participation would involve completion of a short (10 min) checklist, an in-depth interview with the researcher on your child's functioning and participation in daily activities (about 1.5 hours) and analysis of your child's medical/psychological reports (if any). Findings from the study would be very useful to assist schools and parents in identifying areas of need and planning interventions for the child. All information about individual children will be kept strictly confidential. Parents could have access to feedback on the findings for their child, if requested.

The outcomes of this study would go a long way towards improving our understanding of the needs of children with ASD in Singapore. I would greatly appreciate it if you would accept this invitation by completing the enclosed consent form. If you need further clarifications, please feel free to contact me at this e-mail address :

s.aljunied@ucl.ac.uk (or telephone

Thank you very much.

Yours sincerely,

Mariam Aljunied (C.Psychol)  
Postgraduate Student, University College London

<sup>40</sup> This study is approved by the University College London Committee on the Ethics of Non-NHS Human Research.

Name of child :

School :

Class/Age :

Name of parent(s):

Address :

Contact	: Tel (Home)	Tel (Office)
	Tel (HP)	E-mail

✓ I accept the invitation to participate in the research. I am happy for the researcher to contact me to provide more details on the study.

Name of parent

Signature: \_\_\_\_\_  
Date :

*Please return the completed form to (name of contact person in school/institution).*

### **Informed Consent Form**

(This form to be completed independently by the parent)

Title of Project :

Validity and Reliability of the International Classification of Disability, Health and Functioning Children with Autism in Singapore.

	YES	NO
I have read the invitation letter, which gives background information to the study.		
I have had a discussion with Ms Aljunied about the nature and purpose of the project, and had the opportunity to ask questions.		
I understand that a video recording of my interview with Ms Aljunied may be made, and that these recordings will be used solely for the purpose of training of psychologists.		
I understand that my involvement in the study is voluntary and that I can withdraw my participation at any time.		
I understand that following the interview, I can request for feedback on the results about my child.		
I understand that the same information could be given to my child's school, with my consent.		

Signed : ..... Date: .....

Name in capitals : .....

## APPENDIX E

### STUDY 1 : TRAINING FOR ICF INTER-RATER STUDY

#### Introduction

As part of the study on the reliability and validity of the International Classification of Functioning and Health, or ICF (WHO, 2001), inter-rater reliability of the ICF checklist was evaluated. This was done by measuring the percentage of agreement in the ratings for individual children obtained from 2 sources: 1) ICF ratings given by the researcher (first rater) based on an interviews with the child's parent, and 2) ICF ratings given by a second rater, based on video/audio-taped interviews conducted by the researcher.

#### Aims

To ensure that both raters share a common understanding and knowledge of the ICF, training on the use on the ICF checklist was conducted. The objectives of the training were:

1. To understand background and key components of the ICF;
2. To understand concept underlying the items in the ICF checklist;
3. To understand the probing techniques and rating criteria used in the study; and
4. To achieve a 65% agreement between the ratings given by the researcher and the 2<sup>nd</sup> rater.

#### Participants

The researcher, who is also the first rater for the interviews, conducted the training. Both the researcher and the 2<sup>nd</sup> rater are qualified educational psychologists, with experience in conducting clinical interviews with parents of children with learning disabilities (including autism). In addition, both are competent speakers of 'Singapore English'<sup>41</sup> which is the main language used in the interviews.

#### Method

The training included the following activities :

- explanation and discussion using power-point presentation and the ICF manual;
- demonstration (via video) of the use of ICF in the present study; and
- joint-rating sessions, based on video-taped interviews.

Details of the activities for the 2-day training is presented in Table E1.

#### Evaluation

At the end of Day 2 of training, kappa agreement was calculated based on the joint-rating for a final case (i.e. case viewed on final session on the 2<sup>nd</sup> day). The overall inter-rater agreement for this case between the two raters was 0.84, which is considered to be 'excellent' (Landis & Kock, 1977).

<sup>41</sup> 'Singapore English' is a variety of English spoken by the majority of people in Singapore, which often includes language use and vocabulary which are taken from other languages used in the country, e.g. Mandarin and Malay (Foley, 1998).

Table E1 : Outline and Details for Inter-rater Training (Study 1)

TIME	ACTIVITIES	MATERIALS
<b>DAY 1</b>		
9 am – 12 pm	<ul style="list-style-type: none"> <li>- Introduction &amp; background to study of ICF inter-rater reliability</li> <li>- Overview of the ICF framework &amp; ICF checklists <ul style="list-style-type: none"> <li>o Rationale, background</li> <li>o Theoretical model</li> <li>o Structure</li> <li>o Components in checklists</li> <li>o Judgment criteria.</li> <li>o Probing techniques.</li> </ul> </li> </ul>	ICF Manual Interview protocol
12 – 1pm 1 – 4 pm	<ul style="list-style-type: none"> <li>- BREAK</li> <li>- 'Body Function' &amp; 'Body Structure' components in ICF checklist <ul style="list-style-type: none"> <li>o Description of items</li> <li>o Probe questions used</li> <li>o Judgment/criteria used</li> </ul> </li> <li>- Video demonstration<sup>42</sup> interviews on items for 'Body Functions' and 'Body Structure'.</li> <li>- Discussion and clarifications.</li> </ul>	
<b>DAY 2</b>		
9 – 12 pm  12 – 1 pm 1 – 4 pm	<ul style="list-style-type: none"> <li>- 'Activities and Participation' components in ICF checklist <ul style="list-style-type: none"> <li>o description of items</li> <li>o probe questions used</li> <li>o judgment criteria used</li> </ul> </li> <li>- Video demonstration interviews on items for 'Activities &amp; Participation'.</li> <li>- 'Environmental Factors' components in ICF checklist <ul style="list-style-type: none"> <li>o description of items</li> <li>o probe questions used</li> <li>o judgment criteria used</li> </ul> </li> <li>- Video demonstration interviews on items for 'Environment'</li> <li>- Discussion and clarification.</li> <li>- BREAK</li> <li>- Joint-rating of a final case (fresh case not seen before in training sessions)</li> <li>- Discussion of ratings and clarification of discrepancies.</li> </ul>	Videotaped interview.

<sup>42</sup> Note: during the demonstration, 2<sup>nd</sup> rater will note her ratings of the relevant items demonstrated in the videotaped interviews. In the discussion that follows, discrepancies in the ratings will be discussed and clarified.

## APPENDIX F

### STUDY 1 : KAPPA VALUES (INTER-RATER AGREEMENT) FOR INDIVIDUAL ICF ITEMS

Items for Body Functions	Kappa values <sup>H</sup>
<b>b1. MENTAL FUNCTIONS</b>	
b1140 Orientation ( <i>time</i> )	<b>1.00</b>
b1142 Orientation ( <i>person</i> )	<b>1.00</b>
b117 Intellectual ( <i>incl. Retardation, dementia</i> )	<b>0.72</b>
b122 Global psychosocial function	<b>0.65</b>
b1260 Temperament and personality function ( <i>extraversion</i> )	<b>1.00</b>
b1261 Temperament and personality function ( <i>agreeableness</i> )	<b>0.80</b>
b1262 Temperament and personality function ( <i>conscientiousness</i> )	<b>0.61</b>
b1263 Temperament and personality function, ( <i>psychic stability</i> )	<b>1.00</b>
b1264 Temperament and personality function ( <i>openness to experience</i> )	<b>0.75</b>
b1266 Temperament and personality function ( <i>confidences</i> )	<b>0.61</b>
b1267 Temperament and personality function ( <i>trustworthiness</i> )	<b>1.00</b>
b1301 Energy and drive functions ( <i>motivation</i> )	<b>0.45</b>
b1302 Energy and drive functions ( <i>appetite</i> )	<b>0.60</b>
b1303 Energy and drive functions ( <i>craving</i> )	<b>0.62</b>
b1304 Energy and drive functions ( <i>impulse control</i> )	<b>1.00</b>
b1340 Sleep ( <i>amount</i> )	<b>0.46</b>
b1341 Sleep ( <i>onset</i> )	<b>0.46</b>
b1342 Sleep ( <i>maintenance</i> )	<b>0.70</b>
b1400 Attention ( <i>sustaining</i> )	<b>0.43</b>
b1401 Attention ( <i>shifting</i> )	<b>1.00</b>
b1402 Attention ( <i>dividing</i> )	<b>0.85</b>
b1403 Attention ( <i>sharing attention</i> )	<b>0.66</b>
b144 Memory	<b>1.00</b>
b1470 Psychomotor functions ( <i>control of psychomotor functions</i> )	<b>0.60</b>
b1471 Psychomotor functions ( <i>quality of psychomotor functions</i> )	<b>0.53</b>
b1520 Emotional functions ( <i>appropriateness of emotions</i> )	<b>1.00</b>
b1521 Emotional functions ( <i>regulation of emotions</i> )	<b>1.00</b>
b1522 Emotional functions ( <i>range of emotions</i> )	<b>0.64</b>
b1600 Thought functions ( <i>pace of thought</i> )	<b>1.00</b>
b1601 Thought functions ( <i>form of thought</i> )	<b>1.00</b>
b1602 Thought functions ( <i>content of thought</i> )	<b>1.00</b>
b1603 Thought functions ( <i>control of thought</i> )	<b>0.59</b>
b1640 Higher level cognitive functions ( <i>abstraction</i> )	<b>1.00</b>
b1641 Higher level cognitive functions ( <i>organization &amp; planning</i> )	<b>0.81</b>
b1642 Higher level cognitive functions ( <i>time management</i> )	<b>0.75</b>
b1643 Higher level cognitive functions ( <i>cognitive flexibility</i> )	<b>1.00</b>
b1645 Higher level cognitive functions ( <i>judgment</i> )	<b>0.48</b>
b1646 Higher level cognitive functions ( <i>problem-solving</i> )	<b>0.50</b>
b1670 Language ( <i>reception of language</i> )	<b>0.61</b>
b1671 Language ( <i>expression of language</i> )	<b>0.53</b>
b1672 Language ( <i>integration of language</i> )	<b>0.50</b>
b1720 Calculation functions(simple calculations)	<b>0.83</b>
b1721 Calculation functions(complex calculations)	<b>0.65</b>
b176 Complex movements	<b>0.82</b>
<b>b2. SENSORY FUNCTIONS AND PAIN</b>	
b230 Hearing	<b>0.70</b>
b235 Vestibular ( <i>incl. Balance functions</i> )	<b>0.45</b>
b250 Additional sensory function ( <i>taste</i> )	<b>0.36</b>

<sup>H</sup> Kappa values were calculated for items where the variance of ratings was greater than zero.

b250 Additional sensory function ( <i>touch</i> )	0.40
b250 Additional sensory function ( <i>sensing temperature/vibration etc</i> )	1.00
b280 Pain	0.61
<b>b3. VOICE AND SPEECH FUNCTIONS</b>	
b310 Voice	0.81
b320 Articulation	1.00
b330 Fluency and rhythm	0.83
b340 Alternative vocalisations	1.00
<b>b4. FUNCTIONS OF THE CARDIOVASCULAR, HAEMATOLOGICAL, IMMUNOLOGICAL AND RESPIRATORY SYSTEMS</b>	
<b>b5. FUNCTIONS OF THE DIGESTIVE, METABOLIC AND ENDOCRINE SYSTEMS</b>	
<b>b6. GENITOURINARY AND REPRODUCTIVE FUNCTIONS</b>	
<b>b7. NEUROMUSCULOSKELETAL AND MOVEMENT RELATED FUNCTIONS</b>	
<b>b8. FUNCTIONS OF THE SKIN AND RELATED STRUCTURES</b>	
<i>Items for Activity &amp; Participation</i>	
<b>d1. LEARNING AND APPLYING KNOWLEDGE</b>	
d110 Watching	1.00
d115 Listening	1.00
d135 Rehearsing	1.00
d140 Learning to read	0.83
d145 Learning to write	0.70
d150 Learning to calculate ( <i>arithmetic</i> )	0.71
d155 Acquiring skills	0.75
d160 Focusing attention	0.57
d163 Thinking	0.62
d166 Reading	0.54
d170 Writing	0.42
d175 Solving problems	0.40
d177 Making decisions	0.36
<b>d2. GENERAL TASKS AND DEMANDS</b>	
d210 Undertaking a single task	0.81
d220 Undertaking multiple tasks	0.59
d230 Daily routine	0.81
d240 Stress / responsibility	0.80
<b>d3. COMMUNICATION</b>	
d310 Communicating – receiving spoken messages	0.34
d315 Communicating – receiving nonverbal messages	0.65
d325 Communicating – receiving written messages	0.52
d330 Communicating – producing spoken messages	0.52
d335 Communicating – producing nonverbal messages	0.48
d340 Communicating – producing written messages	0.53
d350 Conversation	0.69
d355 Discussion	0.55
d360 Communication devices	1.00
<b>d4. MOBILITY</b>	
<b>d5. SELF CARE</b>	
d510 Washing oneself ( <i>bathing, drying, washing hands, etc</i> )	0.68
d520 Caring for body parts ( <i>brushing teeth, shaving, grooming, etc.</i> )	0.39
d530 Toileting	0.67
d540 Dressing	0.50
d550 Eating	0.65
d560 Drinking	1.00
<b>d6. DOMESTIC LIFE</b>	
d620 Acquisition of goods and services ( <i>shopping, etc.</i> )	1.00

<b>d630</b> Preparation of meals ( <i>cooking etc.</i> )	<b>0.82</b>
<b>d640</b> Doing housework ( <i>cleaning house, washing dishes laundry, ironing, etc.</i> )	<b>0.63</b>
<b>d660</b> Assisting others	<b>0.42</b>
<b>d7. INTERPERSONAL INTERACTIONS AND RELATIONSHIPS</b>	
<b>d7100</b> Basic interpersonal interactions ( <i>respect &amp; warmth</i> )	<b>0.61</b>
<b>d7101</b> Basic interpersonal interactions ( <i>appreciation</i> )	<b>0.78</b>
<b>d7102</b> Basic interpersonal interactions ( <i>tolerance</i> )	<b>0.82</b>
<b>d7103</b> Basic interpersonal interactions ( <i>criticism</i> )	<b>0.86</b>
<b>d7104</b> Basic interpersonal interactions ( <i>social cues</i> )	<b>0.70</b>
<b>d7105</b> Basic interpersonal interactions, ( <i>physical contact</i> )	<b>0.40</b>
<b>d7200</b> Complex interpersonal interactions ( <i>forming relationships</i> )	<b>0.84</b>
<b>d7201</b> Complex interpersonal interactions ( <i>terminating relationships</i> )	<b>0.66</b>
<b>d7202</b> Complex interpersonal interactions ( <i>regulating relationships</i> )	<b>0.66</b>
<b>d7203</b> Complex interpersonal interactions, ( <i>maintaining social rules</i> )	<b>0.84</b>
<b>d7204</b> Complex interpersonal interactions, ( <i>maintaining social space</i> )	<b>0.67</b>
<b>d7305</b> Relating with strangers	<b>1.00</b>
<b>d740</b> Formal relationships	<b>1.00</b>
<b>d750</b> Informal social relationships	<b>0.82</b>
<b>d7601</b> Family relationships (child-parent)	<b>1.00</b>
<b>d7602</b> Family relationships (sibling)	<b>1.00</b>
<b>d7603</b> Family relationships (extended family)	<b>0.85</b>
<b>d8. MAJOR LIFE AREAS</b>	
<b>d810</b> Informal education	<b>0.62</b>
<b>d815</b> Preschool education	<b>0.40</b>
<b>d820</b> School education	<b>1.00</b>
<b>d9. COMMUNITY, SOCIAL AND CIVIC LIFE</b>	
<i>Items for Environmental Factors</i>	
<b>e1. PRODUCTS AND TECHNOLOGY</b>	
<b>e110</b> For personal consumption ( <i>food, medicines</i> )	<b>0.64</b>
<b>e125</b> Products for communication	<b>0.60</b>
<b>e130</b> Products for education	<b>0.47</b>
<b>e155</b> Design, construction and building products and technology of buildings for private use	<b>0.89</b>
<b>e2. NATURAL ENVIRONMENT AND HUMAN MADE CHANGES TO THE ENVIRONMENT</b>	
<b>e3. SUPPORT AND RELATIONSHIPS</b>	
<b>e312</b> Extended family	<b>0.74</b>
<b>e320</b> Friends	<b>0.62</b>
<b>e325</b> Acquaintances, peers, colleagues, neighbours and community members	<b>0.70</b>
<b>e355</b> Health professionals	<b>0.60</b>
<b>e360</b> Other professionals	<b>0.64</b>
<b>e4. ATTITUDES</b>	
<b>e410</b> Individual attitudes of immediate family members	<b>0.73</b>
<b>e410</b> Individual attitudes of extended family members	<b>0.81</b>
<b>e420</b> Individual attitudes of friends	<b>1.00</b>
<b>e410</b> Individual attitudes of acquaintances, peers, colleagues, neighbours and community members	<b>0.42</b>
<b>e430</b> Individual attitudes of people in position of authority	<b>0.22</b>
<b>e440</b> Individual attitudes of personal care providers and personal assistants	<b>1.00</b>
<b>e445</b> Individual attitudes of strangers	<b>0.62</b>
<b>e450</b> Individual attitudes of health professionals	<b>0.44</b>
<b>e455</b> Individual attitudes of other professionals	<b>0.70</b>
<b>e460</b> Societal attitudes	<b>0.64</b>
<b>e465</b> Social norms, practices and ideologies	<b>0.69</b>
<b>e5. SERVICES, SYSTEMS AND POLICIES</b>	
<b>e580</b> Health services, systems and policies	<b>0.61</b>
<b>e585</b> Education and training services, systems and policies	<b>0.52</b>

## APPENDIX G

### KEY FEATURES OF THE BEHAVIOURAL ASSESSMENT OF DYSEXECUTIVE SYNDROME FOR CHILDREN (BADS-C) SUBTESTS

<i>Description of Subtests</i>	<i>Information recorded</i>	<i>Scoring criteria</i>
<b>Playing Card test</b>  The aim of task is to see children's ability to change an established pattern of responding.  The children are presented with a set of 21 non-picture playing cards, that are turned over one at a time. Children are required to say 'yes' or 'no' to each card, based on a given rule. The child is tested twice, and with each presentation of the cards, a different rule is given: <u>Rule 1</u> : say 'yes' if card is red, 'no' if its black <u>Rule 2</u> : say 'yes' if card is the same colour as the one before, 'no' if its different.	Child's responses to task with Rule 1 & 2.  Time taken to complete each round/ presentation.  Number of errors made in applying Rule 1 and Rule 2.	Scores awarded for: <ul style="list-style-type: none"> <li>Number of errors made in application of Rule 2.</li> </ul>
<b>Water test</b>  The aim of the task is to see the children's ability to identify and implement a plan of action in order to solve a given problem.  The problem presented in this task requires physical manipulation of a variety of materials, e.g. water, plastic tube, screw top and a cork. A small cork is placed inside a long, thin glass tube, and placed adjacent to a beaker that is half filled with water. The children are asked to get the cork out of the tube, using the materials given, but with some rules/conditions, e.g. the children are not allowed to touch the lid with their fingers.  The apparatus is set up such that in order for the children to solve the given problem, a five-step solution is required. During the test, at stipulated time intervals, if the children do not progress from one solution step to another, the tester will prompt by the next appropriate steps.	What the child did to solve the problem.  Each time the child performed any of the 5 steps solution independently.  Number of times the child perseverated, i.e. repeats previous / inefficient response 3 or more times.  Time taken to complete the task.	Scores awarded for <ul style="list-style-type: none"> <li>Completing each of the 5 steps independently.</li> </ul> Scores deducted if <ul style="list-style-type: none"> <li>The child had broken any rules in each step</li> <li>The child had perseverated during the test.</li> </ul>
<b>Key Search test</b>  The aim of this test is to see the children's ability to plan an efficient, systematic, implementable plan of actions.  The children are presented with a piece of A4 square size paper with a 100mm square in the middle and a black dot 50mm below the square. The children are asked to imagine that the square is a	Child's actions demonstrating understanding of task requirements, e.g. searching file within field boundary, point of entry, making continuous line.  Whether or not the search was systematic, planned or efficient,	Scores awarded for <ul style="list-style-type: none"> <li>Understanding task requirement;</li> <li>Producing systematic, plan,</li> <li>Producing an</li> </ul>

<p>large field in which they have lost their keys. Starting from the dot, the children are asked to draw a line to show how they would search the field for the lost keys.</p>	<p>and implementable.</p>	<p>implementable search.</p>
<b>Zoo Map tests 1 &amp; 2</b>		
<p>The aim of the test is to see the children's ability to plan a route, under specified rules/conditions.</p> <p>In both tests, the children are presented with a map of a zoo and asked to plan a route in order to visit six out of 12 possible locations in the zoo. The rules/conditions specified are for example, the children can only take 1 ride on the 'camels' and use 'white paths' in the map only once.</p>	<p>Sequence of places visited by child.</p> <p>Errors made by child, e.g. violation of rule/.constraints, failure to make continuous line, cutting across the grass, number of inappropriate places visited.</p> <p>Time spent on planning and total time taken to complete the tasks.</p> <p>Observations of child's comments and pauses during testing.</p>	<p>Scores awarded for:</p> <ul style="list-style-type: none"> <li>Visiting the right place in the right sequence.</li> </ul> <p>Scores deducted for :</p> <ul style="list-style-type: none"> <li>Number of times the rules are broken.</li> </ul>
<p>Zoo map test 1 is more demanding than test 2, as test 1 is open-ended, where little structure is provided. Hence, the child would need to figure out the order in which each designated place must be visited so that no rules/conditions are violated. In test 2, the order of the places the child must visit is specified, hence, test 2 places a relatively a lower demand on the child's planning skills.</p>		
<b>Six Part test</b>		
<p>The aim of this test is to see the children's ability to plan, schedule task and self-monitor performance.</p> <p>Children are given 3 colour coded tasks to do : green tasks (simple arithmetic); blue tasks (picture naming) and red tasks (sorting). Each of the tasks has 2 parts, i.e. part 1 &amp; part 2, so in total, there are six parts to the test.</p> <p>The children's task is to attempt each of the six parts within a time limit of 5 minutes. They are not required to complete all of the items (impossible to do), but rules/conditions are given, i.e. the children cannot move from one part of one colour to another part of the same colour consecutively.</p>	<p>Order of parts attempted by child.</p> <p>Time spent on each of the six parts.</p> <p>Number of times rules/conditions are broken.</p> <p>Total time for completion of tasks.</p> <p>Whether or not the child is aware that s/he has broken the rules</p>	<p>Scores awarded for:</p> <ul style="list-style-type: none"> <li>Each part attempted</li> <li>Use of clear strategy</li> </ul> <p>Deduction of scores for:</p> <ul style="list-style-type: none"> <li>Each task in which the rule was broken</li> <li>If child returns to any part of the test 3 times or more (i.e. perseverating)</li> </ul>

## APPENDIX H

### FIELD-TEST OF THE BEHAVIOURAL ASSESSMENT FOR DYSEXECUTIVE SYNDROME FOR CHILDREN (BADS-C)

#### Objectives

The aim of this field-test is to evaluate the extent to which:

1. The task and language demands in the BADS-C are appropriate for the target sample in the present thesis, i.e. children with autism, age 8 to 12 years old.
2. There is adequate consistency (i.e. internal reliability) between children's scores in individual subtests and their total (overall) scores in the BADS-C.
3. The BADS-C scores of children with autism reflect greater impairments in executive function compared with the normative sample, as predicted by the executive function hypothesis for autism (Ozonoff, 1985) (i.e. criterion validity).

#### Methods

##### Participants

10 children, aged 8 to 12 years (mean age 9.9 yrs), with pre-diagnosis of autism participated in the field-test. The children were selected from the database of an on-going study by White et al. (in prep.), based on their ages. The children were from two special schools, with non-verbal IQ range of between 90 and 125 (mean IQ = 109.22; SD = 14.5); and language ability range of between 73 and 123 (mean = 92.3; SD = 13.5).

##### Measures

The full BADS-C (6 subtests) was used (see description of the subtests in Appendix G). Estimates of children's general non-verbal ability were based on the Standards Progressive Matrices or SPM (Raven, 1957); and estimates of language ability were based on the British Picture Vocabulary Scale, or BPVS (Dunn et al., 1982).

##### Procedure

All 10 children were tested individually using the BADS-C by the researcher in a quiet room in their respective schools. Testing time was about 30 minutes per child and standardised procedures were adhered to. The children's scores for the estimated non-verbal IQ (SPM scores) and estimates of language ability (BPVS scores) were obtained from White et al.'s database.

#### Results

Children's BADS-C raw scores were converted to scaled scores (range 1 to 19, mean = 10) and standard scores (mean = 100, SD = 15) based on their respective age and IQ. The results are presented based on the 3 stated aims of the field-test (see 'Objectives').

##### Task and language demands

The first aim was to see if the task and language demands in the test protocols and instructions were appropriate for present thesis, i.e. children with autism, age 8 to 12 years old. Table H1 shows the age and standard scores for language, non-verbal IQ and BADS-C for individual participants in the field-test.

Table H1: Profile and Standard Scores for Individual Participants in BADS-C Field-Test

Participant	Age (in years)	Non-Verbal IQ (Std Score)	Language (Std Score)	BADS-C Overall Score (Std Score & Classification)
Child 1	12	112	94	80 Low Average
Child 2	12	91	79	74 Borderline
Child 3	11	122	87	74 Borderline
Child 4	10	108	83	62 Impaired
Child 5	8	125	123	62 Impaired
Child 6	10	123	94	58 Impaired
Child 7	9	119	79	56 Impaired
Child 8	11	90	98	54 Impaired
Child 9	8	93	75	Not computable
Child 10	8	108	73	Not computable

Out of the 10 children, 2 were unable to respond adequately to the task demands, and consequently, no BADS-C scores could be computed (see Table H1, Child 9 & 10). It can be seen that although the two children had estimated non-verbal IQ that were within the average ranges (i.e. 93 and 108 respectively), their language scores were the lowest in the group (standard score of 75 and 73 respectively, i.e. well below average).

During testing, it was noted that Child 9 displayed frequent repetitive physical movements (motor stereotypies), such as swaying from side to side and hand flapping. In addition, frequent echolalia was observed, which affected the child's ability to respond to some of the task. For example, in subtest 1, i.e. Playing Cards, he answered 'Yes' to all questions and said 'Yes' repetitively throughout the subsequent subtests<sup>43</sup>.

The results suggest that for children with autism, the BADS-C may not be suitable for children with very low verbal comprehension skills (standard score 75 or less) and children showing frequent occurrences of repetitive stereotyped behaviours, such as motor stereotypies and echolalia.

#### *Internal reliability*

The second aim was to see if the internal consistency between children's scores on individual subtests and their overall scores in the BAS-C battery was adequate. Cronbach alpha was used to evaluate the internal reliability, and as shown in Table H2, the alpha value for overall BADS-C Scaled Score is 0.60, which is considered as indicating adequate reliability (Field, 2000).

<sup>43</sup> Note : In the Playing Card test, the child is asked to say 'yes' or 'no' each time a card is presented. For details see Appendix G.

Table.H2 : Alpha Values for Overall Scaled Scores (BADS-C Field-test)

Test	Mean	Std Error	Min	Max	Internal Reliability (Cronbach Alpha)
<b>Overall Scaled Score</b>	<b>43.50</b>	<b>1.71</b>	<b>38</b>	<b>51</b>	<b>0.60*</b>
<b>Subtests</b>					<i>Alpha value if subtest item is deleted from Overall Scaled Score</i>
Playing card test	5.38	1.50	1	14	0.58
Water test*	7.50	1.34	4	14	0.47*
Key search test	8.00	1.16	3	14	0.56
Zoo map test 1*	7.12	1.07	4	14	0.45*
Zoo map test 2	10.25	1.49	1	13	0.62
Six part test*	5.25	0.56	3	8	0.52*

\*Alpha values for Overall Scaled Score fell below the 0.6 for 'adequate' reliability when these subtests were removed.

However, when individual subtests were removed from the Overall Scaled Score, some the reliability indices vary considerably. For example, the alpha values for Overall Scaled Score dropped below the 0.60 (adequate) level when children's scores on the Water test, Zoo Map Test 1 and Six Part test were individually removed from the overall BADS-C score (see Table H2). This suggests that these 3 subtests may have had greater influence on the reliability estimates for the overall scale. The BADS-C manual did not provide internal reliability estimates based on the norm sample and as such, it has not been possible to evaluate the extent to which the reliability index obtained in the field-test is comparable to the test reliability that was based on the larger normative sample.

#### *Criterion validity*

The third aim was to see if the BADS-C scores of children with autism would reflect greater impairments in executive function. Previous research have indicated that children with autism exhibit impaired functioning in tests of executive function (Ozonoff, 1985; Hill, 2004). One criterion that could reflect the criterion validity of the BADS-C is extent to which the scores of children with autism in the field-test sample is lower than the children in the normative sample.

For this analysis, two comparisons were made: one based on the children Overall BADS-C standards scores (mean =100; SD = 15), and another based on subtests scaled scores.

**Overall BADS-C Standard Scores :** From the present field-test, out of the 8 children who were able to respond adequately to the test demands, 5 children (or 63%) were classified as having 'Impaired' executive function based on their BADS-C scores, 2 children (25%) showed 'Borderline' impairments in executive function, and 1 child had 'Low Average' scores (see Table H3). These impairments appear not be related children's non-verbal learning abilities as all the children had at least average non-verbal IQ. The results suggest that, the children with autism in the field-test showed greater impairments in executive functioning, as reflected by their BADS-C Overall Scores.

**Table H3: Individual Scores and Summary of Children's Executive Function (BADS-C Field-test)**

Participant	BADS-C Overall Score (Std Score & Classification)		Est. IQ (Std score)	Language (Std score)
Child 1	80	Low Average	112	94
Child 2	74	Borderline	91	79
Child 3	74	Borderline	122	87
Child 4	62	Impaired	108	83
Child 5	62	Impaired	125	123
Child 6	58	Impaired	123	94
Child 7	56	Impaired	119	79
Child 8	54	Impaired	90	98
<b>Summary</b>				
Classification for Executive Function			No. of children	Percentage
Low average			1	12.5%
Borderline			2	25%
Impaired			5	62.5%
	Total			100%

**BADS-C Scaled Scores** : Comparison of the scaled scores of children with autism in the field-test with the children in the control group from the BADS-C norm sample suggests that the children with autism have markedly lower overall Scaled Scores, i.e. about 14 points lower than the mean scaled scores for controls (see Table H4). Differences in scores for individual subtests however were varied: for example, the children with autism had lower scaled scores for all except for Zoo Map Test 2, where the scores of the autism group were slightly higher than controls.

It should be noted that these differences were not tested for statistical significance. However, coupled with the results discussed earlier, i.e. a large majority of the children in the field-test showed impaired executive functioning (see Table H3), the results suggest that the BADS-C appears to be able to reflect the impairments in executive function in children with autism. This is consistent with the executive dysfunction hypothesis in autism (Ozonoff, 1985).

**Table H4: Comparison Of Mean Scaled Scores For Children With Autism And Controls (BADS-C Field-test)**

	CONTROLS (from BADS-C norm sample) n = 259	AUTISM (from field-test sample) n = 8
Playing card test	10.02	5.38
Water test	9.97	7.50
Key search test	11.38	8.00
Zoo map test 1	10.00	7.12
Zoo map test 2	9.93	10.25
Six part test	10.04	5.25
<b>Overall Scaled Score</b>	<b>61.28</b>	<b>43.50</b>

## **Conclusion**

Results from the field-test indicate that:

- The BADS-C is suitable for children with autism, except those with very low language skills (i.e. standard score 75 or less), and children showing frequent occurrences of repetitive stereotyped behaviours, such as motor stereotypies and echolalia.
- The internal reliability for the overall BADS-C Scaled Score is adequate, although individual subtests may have different impact on the internal reliability of the overall scale.
- A large majority of the children with autism in the sample (63%) showed significant impairments in executive function as measured by the BADS-C. The mean scaled scores for children with autism in the field-test was lower than the controls by about 14 points. This suggests that the BADS-C was able to reflect the overall impairments in executive function that is often seen in children with autism.

## **Use of the BADS-C in the present thesis**

Results of the field-test provide some support for the use of BADS-C as a measure of executive function of children with autism, aged 8 to 12 years. However, difficulties are expected with children with very low language skills and severe motor stereotypies. The overall BADS-C Scale Score, which is based on the sum of all subtests scores, appears to have a greater reliability and validity as an indicator of executive functioning compared with individual subtest scores.

## APPENDIX I

### FIELD-TEST OF THE CHILDREN'S EMBEDDED FIGURES TEST (CEFT)

#### Aims

The aim of the field-test was to gauge the suitability of the CEFT with a small sample of children which is representative of the participants in study 2, i.e. children with autism, aged 8 to 12 years old. The field test focused on following aspects:

- The suitability of the language and task demands in the test for the target group;
- The appropriateness of the difficulty level of the test items; and
- Identifying suitable outcome measures for the CEFT.

#### Methods

##### *Participants*

4 children, aged 9:0 to 11:02 years old, with pre-diagnosis of autism participated in the field-test. The children from a special school in the UK which had agreed to participate in the study and were selected based on their age. Five children were initially selected, but one child dropped out from the study due to illness. The children's estimated nonverbal IQ, based on the Standard Progressive Matrices or SPM (Raven, 1957) range from 75 to 125; while their language abilities range from 65 to 109, based on the British Picture Vocabulary Scale, or BPVS (Dunn et al, 1982). See Table I1 below:

Table I1 : Profile of children in CEFT field-test

Child	AGE	Language Std. Score*	Est. Nonverbal IQ Std Score*
A	9:07 yrs	109	125
B	11:02 yrs	78	85
C	11:02 yrs	65	75
D	9:00 yrs	73	77

\*Std. Score Mean = 100; SD = 15

##### *Procedure*

The children were tested individually using the CEFT by the researcher. The children were first tested with the CEFT, then the BPVS and the SPM. Testing was conducted in a quiet room in their school. Testing time per child was about 20 minutes.

#### Results

Individual children's responses on the CEFT were used to obtain 2 scores, i.e. Total CEFT, which was the total number of items answered correctly by each child, and secondly, the CEFT Test-Age which was the age-equivalent score.

##### *Task and language demands*

All four children passed the discrimination and practice items in the test and were able to cope with the instructions and task demands of the test. As shown in Table I2, Child A, who has the highest non-verbal IQ and language skills in the sample, also obtained the highest CEFT score.

**Table I2 : Children's Individual CEFT Scores (CEFT Field-test)**

Child	AGE	Language	Est. Nonverbal IQ	Total CEFT Max. = 24	CEFT Test Age Max. = 16 yrs
A	9:07 yrs	109	125	15	9-10 yrs
B	11:02 yrs	78	85	11	7-8 yrs
C	11:02 yrs	65	75	10	7-8 yrs
D	9:00 yrs	73	77	6	7-8 yrs

To further explore the relationship (if any) between non-verbal IQ and language skills in children's performance on the CEFT, a non-parametric correlation analysis was carried out (See Table I3). The analysis showed a strong correlation between Total CEFT scores and children's language skills, based on BPVS scores ( $r = 0.80$ ); and a moderate correlation with children's estimated nonverbal IQ ( $r = 0.40$ ). These correlations however, did not reach statistically significance level, probably due to the very small number of cases involved, i.e.  $n = 4$ .

**Table I3: Non-parametric Correlation between Language, Nonverbal IQ and CEFT Scores (Spearman's rho) (CEFT Field-test)**

	Language	Nonverbal IQ	Total CEFT
Language	1.00	-	-
Nonverbal IQ	0.20	1.00	-
Total CEFT	0.80	0.40	1.00

#### *Difficulty level*

To gauge whether the difficulty level of the CEFT was appropriate for children in the study, children's scores were analysed for possible ceiling or floor effects. From Table I2, it can be seen that Child A obtained the highest score (15), which was well below the maximum/ceiling score of 24. At the same time, the lowest score obtained was 6 (Child D), which was well above the minimum/basal score. This suggests that for the group of children in the study, there were no floor or ceiling effects.

#### *Outcome measure*

Two measures were obtained, i.e. CEFT Test Score (i.e. raw scores) and CEFT Test-Age (i.e. age equivalent scores). Child A, who had the highest raw score, also had the highest CEFT test age, i.e. 9 to 10 yrs (see Table I2). However, although Child B, C and D had the same CEFT Test-Age score (i.e. 7 to 8 yrs), the corresponding raw scores (Total CEFT) were different for all three children. This indicated that the children's raw scores (i.e. Total CEFT) reflected a finer differentiation of children's performance than the age-equivalent score (i.e. CEFT Test-Age).

#### **Conclusion**

The results of the field-test provided some support for the use of CEFT as a measure of central coherence in children with autism in study 2: the children in the sample were able to cope with the task and language demands of the test; and there were no ceiling or floor effects. The CEFT total score appeared to be a more appropriate outcome measure than CEFT age-equivalent score. However, when used with children across a wider age range, possible age effects may need to be accounted for in the analyses of raw scores.

## APPENDIX J

### DESCRIPTION OF THE SIX SUBTESTS IN THE COGNITIVE MODIFIABILITY BATTERY (CMB)

Subtest	Target Skill	No.	Abstraction level (i.e. difficulty level)
1. Seriation	Ordering of colored blocks according to one or more rules presented in sequential order. Example of rules : progression of height (e.g. high to low); colours (i.e. red-green-blue-yellow) and numbers.	4	Low abstraction level. Items progress from ordering with one rule to a combination of multiple rules.
2. Reproduction of patterns	Copying of patterns of coloured blocks.	9	Low abstraction level. Items progress from simple patterns of 1 dimension (e.g. colour), to more complex patterns with variations along 3 dimensions (colour, height & position).
3. Analogies*	Completion of three dimensional pattern based on analogical transformation, e.g. if $A \rightarrow B, C \rightarrow ?$ . Transformations are based on changes in colour, height and position of blocks in the window plates.	23	Moderate to high abstraction level. Items progress from transformations based on 1 dimension, e.g. change in color, to transformations based on multiple dimension, e.g. change in colour and height and position of blocks. More difficult (higher-order) items are also included, involving double transformation, e.g. if $A \rightarrow B, A \rightarrow C$ , then $C \rightarrow ?$
4. Sequences	Completion of three dimensional patterns based on systematic progression, e.g. if $A \rightarrow B \rightarrow C \rightarrow ?$ .	16	Low to moderate abstraction level. Items progress from progressive sequences involving 1 dimension (e.g. progression in size), to progression involving multiple dimensions, e.g. progression in size, position and number of blocks.
5. Memory (Note: This is the only subtest which is administered without mediation)	Recall of position, i.e. position of 'open' window in the wooden plate (see description of Materials)	9	Low to moderate abstraction level Items progress from simple recall, i.e. where the position of stimulus plate is identical to the targeted recall, to more difficult items where the position of the stimulus plate is rotated 90 & 180 degrees from the targeted response/recall.
6. Mental rotation	Matching of patterns in rotations of 45, 90 and 135 degrees.	6	Low to moderate abstraction level. Items progress from 45-degree rotation of symmetrical patterns, to 90 and 145-degree rotations of asymmetrical patterns.

\* The Analogies subtest is used in the present thesis (Study 2 & 3).

## APPENDIX K

### FIELD-TEST OF THE COGNITIVE MODIFIABILITY BATTERY (CMB) ANALOGIES SUBTEST

#### Aims

The aims of the field-test were:

1. To see if the task and language demands in the test protocols and instructions are appropriate for participants in study 2, i.e. children with autism, aged 8 to 12 years old;
2. To see if the difficulty level of the items in the test was appropriate for the intended participants. This was felt necessary as the CMB was originally developed for children of slightly lower age range, i.e. kindergarten to Grade 3;
3. To identify the appropriate scoring criteria for use in the present study. The CMB has two alternative criteria, namely the 'all or none', and 'partial credit' scoring criteria.

#### Methods

##### Participants

5 children, aged 9:0 to 11:02 years, with pre-diagnosis of autism participated in the field-test. The children were from a special school in UK and were selected based on their age. The children's estimated IQ (based on the Standard Progressive Matrices) range between 75 to 125, while their language abilities range between 65 to 109 (based on the British Picture Vocabulary Scale) (See Table K1). During the testing period, one child fell ill and was on long medical leave. Hence full profile data was only available for 4 children

Table K1 : Profile of Children in CMB Field-test

Child No.	Age	Language	Est. IQ	Time for Teaching Phase
A	9:07 yrs	109	125	30mins
B	11:02 yrs	65	85	80 mins
C	11:02 yrs	78	75	45 mins
D	9:00 yrs	73	77	30 mins
E	9:11 yrs	.	.	70 mins

##### Procedure

All 5 children were tested individually using the CMB Analogies subtest by the researcher, who had undergone training in the use of the CMB conducted by Tzuriel in Singapore in 1999. In addition, prior to the field-testing the researcher re-familiarized herself with the administration and mediation procedures by reviewing a video recording of Tzuriel's administration of the CMB Analogies.

The children were first tested with the CMB Analogies, then the BPVS and followed by the SPM. Testing was conducted in a quiet room in the school. Testing time for the pre

and post-teaching phases of the CMB was 30 mins (i.e. 15 mins for each phase) and varied between 30 to 80 mins per child for the teaching phase.

For this field-test, prior to pre-teaching phase, children's understanding of the 4 dimensions in the test were checked, namely colour (green, red, blue, yellow); height (big, small); number (1 to 4); and position (top and bottom). This is important to ensure that they had the pre-requisite language for the Analogies tasks. The check is achieved by asking a series of pre-requisite questions which were incorporated into the scripted administration procedures.

## Results

Children's CMB Analogies responses were scored using both Method 1 (All or None) and Method 2 (Partial Credit):

Table K2 : Individual Children's Results by Methods 1 & 2 Scoring Criteria (CMB Field-test)

Child	Method 1 (All or None; Max = 23)			Method 2 (Partial Credit; Max = 92)		
	Pre 1	Post 1	Gains 1	Pre 2	Post 2	Gains 2
A	18	19	1	82	87	5
B	13	14	1	64	73	9
C	9	12	3	58	74	16
D	2	7	5	45	59	14
E	1	0	-1	34	40	6

### *Task and language demands*

All 5 children responded correctly to all the pre-requisite questions, indicating that they have the pre-requisite skills for the Analogies subtest. In addition, all of the children managed to perform adequately in at least one of the Pre-Teaching items (the lowest Pre-teaching score was obtained by Child E, who had total scores of 1 (Method 1) and 34 (Method 2). This indicated that all of the children were able to cope with the language and task demands.

### *Item difficulty*

None of the child reached the maximum score for Pre-Teaching and Post-teaching Phases, as the highest Post-Teaching score was obtained by Child A with 19 (using Method 1) and 87 (using Method 2) – See Table K2. Thus, based on the field-test, there appears to be no ceiling effect in the children's Post-Teaching scores. However, it should be noted that Child A's Post-Teaching score is close to maximum score. In the sample of participants in study 2, where children's age may extend up to 12 yrs, there could be some children who may reach the maximum score.

### *Scoring criteria*

The field-test data also indicates that scores based on Method 2 (i.e. partial credit) appears to reflect a finer gradation in the children's performance. For example, the lowest score (i.e. 1) using Method 1 was obtained by Child E. His corresponding score based on

Method 2 is 34, which is far above the minimum/basal score of zero. Similarly, although he showed a negative Gain score using Method 1, his Gain score using the 'partial credit' (Method 2) criteria indicates an increase of 6 points (See Table K2). Coupled with the findings of Tzuriel (2000) that reliability coefficients were higher for 'partial credit' (Method 2) than for the 'all or none' (Method 1) scoring criteria, the result of the field-test supports the use of Method 2 for use in Study 2.

### **Conclusion**

Results of the field-test provide some support for the use of CMB Analogies subtest as a measure of children's performance in dynamic assessment tests in Study 2. The children in the sample were able to cope with the language and task demands of the test. In addition, no floor or ceiling effects were noted, although the Post-Teaching score of 1 child was close to the maximum score. The use of Method 2 scoring criteria appears to better reflect differences in children's scores. The Post-Teaching scores could be used as an outcome measure of children's performance in dynamic assessment tests after mediation. However, adjustments would need to be made to account for children's different initial performances at the pre-teaching phase, as suggested by Tzuriel (2001).

## APPENDIX L

### PRELIMINARY STUDY FOR THE SELECTION OF ITEMS FOR THE THEORY OF MIND (TOM) BATTERY

#### Aims

The aim of the study was to select suitable items that can be used in the ToM Battery. A series of tasks of mentalising abilities (see Table L1) which had been used in autism research were used and evaluated based on the following criteria:

1. The items show difficulty levels that is appropriate for the target sample in the present thesis, i.e. children with autism, aged 8 to 12 years old;
2. The items contribute positively to the internal reliability of the ToM Battery; and
3. Children's responses in the test items are relatively free from bias, i.e. are not affected by characteristics extraneous to theory of mind abilities, such as age and non-verbal abilities.

Table L1 : Items in ToM Battery (Preliminary Study)

Sub-tasks	Items
1 <sup>st</sup> order false belief	Sally-Ann
	Smarties
2 <sup>nd</sup> order false belief	Ice-cream man
	Birthday puppy
Interpretive diversity	Cow
	Dalmation
Deception	Penny Hiding
Advance false belief	Strange stories <sup>44</sup> – ToM
Aggregate ToM	Sum of scores from all items

#### Methods

##### Participants

Participants were sampled from White et al's (in prep.) study and included data from 27 children aged 8 to 10 years (mean age 9:9 yrs), all of whom had pre-existing diagnosis of autism spectrum disorder (see Table L2). The children were first language speakers of English. There is a wide range in participants' verbal (65 to 123) and nonverbal abilities (63 to 123).

<sup>44</sup> Only the theory of mind strange stories were included in this study. In White et al.'s study, two other types of strange stories were used as controls, namely the 'Physical Stories' and 'Jumbled Stories'.

**Table L2 : Participants' Profile (ToM Preliminary Study)**

	Minimum	Maximum	Mean	Std Dev.
Age	8:3yrs	10:10 yrs	9:9 yrs	8.5
Nonverbal IQ (SPM*)	63	125	98	18.8
Verbal IQ (BPVS**)	65	123	84	11.08

\*SPM : Standard Progressive Matrices (Raven, 1958)

\*\*BPVS : British Picture Vocabulary Scales (Dunn et al, 1982)

### *Measures*

All the test items from White et al's study were used. For the purpose of the present study, the items were grouped according to 5 sub-tasks, and a composite score 'Aggregate ToM' score was computed (see Table L3) which reflects children's overall theory of mind abilities. The Standard Progressive Matrices (SPM, Raven, 1958) was used to obtain an index of children non-verbal abilities; and the British Picture Vocabulary Scale (BPVS, Dunn et al., 1982) was used as a measure of verbal abilities used.

### *Procedures*

Each child was tested individually in a quiet room in his/her school by a trained researcher. The same order of the testing was kept for all participants, starting with the BPVS and SPM, and then followed by the items in the ToM Battery.

### **Results**

Results for the field-test are reported according to the stated evaluation criteria: (see 'Aims').

#### *First criterion: item difficulty*

For the first criterion, the items were evaluated in terms of the percentage of children that passed the tests adequately. As shown in Table L3, there are no items in which all participants obtained maximum score, i.e. no ceiling effects. The items in the first order false belief sub-task, namely the Smarties and Sally-Ann tests, appear to be the 'easiest' items, i.e. with the highest number of children achieving maximum possible score (51.9%). On the other hand, items in the advanced false belief sub-task appear to be the most difficult, with no child achieving the maximum possible score. For the other items, the percentage of children achieving maximum score ranges from 14.8% (Penny Hiding) to 55.6% (Birthday Puppy).

Closer analysis of the percentage passes in the advance false belief subtask revealed that the majority of children (85.1%) obtained the lower band score (see Table L4). This suggests that the difficulty level for the Strange Stories may not be appropriate for the children in the study.

Table L3: Participants' Performance in Individual Items in ToM Battery (Preliminary Study)

	Min	Max	Mean	Std Error	Max. possible score	Percentage with max possible score
<b>Aggregate ToM Score</b>	<b>7.67</b>	<b>33.0</b>	<b>19.47</b>	<b>1.27</b>	<b>52</b>	<b>0.0%</b>
<b>Sub-Tasks</b>						
<b>1<sup>st</sup> order false belief</b>	<b>0.00</b>	<b>2.00</b>	<b>1.22</b>	<b>0.72</b>	<b>2.00</b>	<b>51.9%</b>
Sally-Ann	0.00	1.00	0.60	0.09	1.00	59.3%
Smarties	0.00	1.00	0.63	0.09	1.00	63.0%
<b>2<sup>nd</sup> order false belief</b>	<b>0.00</b>	<b>2.00</b>	<b>0.96</b>	<b>0.16</b>	<b>2.00</b>	<b>33.3%</b>
Ice-cream man	0.00	1.00	0.41	0.09	1.00	40.7%
Birthday puppy	0.00	1.00	0.55	0.09	1.00	55.6%
<b>Interpretive diversity</b>	<b>0.00</b>	<b>2.00</b>	<b>0.66</b>	<b>0.21</b>	<b>2.00</b>	<b>20.0%</b>
Cow	0.00	1.00	0.27	0.12	1.00	26.7%
Dalmation	0.00	1.00	0.40	0.13	1.00	40.0%
<b>Deception</b>	<b>0.00</b>	<b>6.00</b>	<b>3.41</b>	<b>0.32</b>	<b>6.00</b>	<b>14.8%</b>
Penny Hiding	0.00	6.00	3.41	0.32	6.00	14.8%
<b>Advance false belief</b>	<b>6.67</b>	<b>21.00</b>	<b>13.51</b>	<b>0.85</b>	<b>40.00</b>	<b>0.0%</b>

Table L4 : Percentage of Participants in Each Score Range for Strange Stories (Preliminary Study)

Items	Range of Scores		
	Fail	Lower band	Upper band
Strange Stories	0	1 to 10	11 to 21

*Second criterion: internal consistency*

Table L5 shows the internal reliability (Cronbach Alpha) for the Aggregate ToM ( $\alpha=0.71$ ); and alpha values of the ToM Battery if each item is removed. This gives an indication of the consistency of scores across all items, as well as the extent to which a particular item affects the reliability of the Aggregate ToM Score. The overall alpha value of the ToM Battery exceeds the 0.6 level for 'adequate reliability' (Field, 2000). In addition, when the items were individually deleted from the aggregate index, alphas never dropped below 0.6, suggesting that most items were contributing positively to the reliability of the aggregate index.

However, for the interpretive diversity item (Cow/Dalmatian), the alpha values for Aggregate ToM actually increased (i.e. improved) when these items were removed. Closer analysis indicated that the interpretive diversity items did not show adequate inter-subtest correlation with other items in the Battery (see Table L6). All other subtasks correlated significantly with each other, indicating consistency in children's responses. The absence of correlation suggests that there may be some problems in the assumption about the consistency between children's performance in the interpretive diversity and the other false belief tasks in the ToM Battery.

**Table L5 : Internal Reliability of ToM Battery  
(Preliminary Study)**

Items	Cronbach Alpha
Aggregate ToM	0.71
<b>Alpha value if item is removed from Aggregate ToM score</b>	
Sally-Ann	0.68
Smarties	0.67
Ice-cream man	0.67
Birthday puppy	0.69
Cow/Dalmation	0.72
Penny Hiding	0.60
Strange stories - ToM	0.67

**Table L6 : Inter-Subtask Correlation in ToM Test Battery (Preliminary Study)**

	1 <sup>st</sup> order false belief	2 <sup>nd</sup> order false belief	Interpretive diversity	Deception	Advance false belief
1 <sup>st</sup> order false belief <i>Sally-Ann &amp; Smarties</i>	1	0.72**	0.06	0.62**	0.62**
2 <sup>nd</sup> order false belief <i>Ice-cream &amp; Birthday Puppy</i>	-	1	0.1	0.44*	0.57**
Interpretive diversity <i>Cow &amp; Dalmation</i>	-	-	1	0.19	0.19
Deception <i>Penny Hiding</i>	-	-	-	1	0.48*
Advance false belief <i>Strange Stories</i>	-	-	-	-	1

\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ .

### *Third criterion: freedom from bias*

As discussed (See section 4.4), three factors have been found to influence theory of mind scores, i.e. age, nonverbal and verbal abilities (Happe, 1995). One possible gauge of the validity of the ToM Battery is the extent to which the children's responses in the ToM Battery are distinct from these characteristics which (albeit related) are extraneous to ToM abilities.

Table L7 shows the correlation of ToM battery with these three factors, i.e. age, nonverbal and verbal abilities. It can be seen that only the Verbal IQ scores (BPVS) were significantly correlated with scores on the second-order and advance false belief subtasks. This suggests that between the three factors, verbal IQ (as measured by BPVS) seems to have a significant effect on scores for second order and advance false belief sub-tasks. On the other hand, differences in age and non-verbal abilities (as measured by SPM) did not have any significant affect on the ToM scores of the children in the field-test.

**Table L7 : Correlation between ToM and : Age; Nonverbal; and Verbal IQ (Preliminary Study)**

ToM Battery	Age	Nonverbal IQ (SPM)	Verbal IQ (BPVS)
Aggregate ToM Score	0.15	0.12	0.61**
<b>Sub-Tasks</b>			
1 <sup>st</sup> order false belief	0.13	0.05	0.38
2 <sup>nd</sup> order false belief	0.15	0.13	0.41*
Interpretive diversity	0.42	0.27	0.47
Deception	0.19	0.31	0.35
Advance false belief	0.24	0.7	0.57**

\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ .

To summarise, the results indicate the following:

- **Item difficulty:** With the exception of items in the advance false belief task, all other items in battery had appropriate difficulty level, i.e. no apparent ceiling or floor effects in the children's scores. As expected, the easiest tasks appear to be items in the 1<sup>st</sup> order false belief subtask, where more than 50% of the children obtained the maximum score, and the most difficult are items in the advance false belief tasks, where most participants achieved the lower band score. This suggests that the items in the advance false belief subtask may not be suitable for children in the 8 to 10 year old range (i.e. the items were too difficult) and may need to be replaced .
- **Internal reliability:** Analysis of the changes in alpha values if each item is removed from the Aggregate ToM score suggests that most of the items contributed positively to the overall reliability index. The alpha values for the Aggregate ToM exceeded the 0.6 level for adequate internal reliability. With the exception of the items for interpretive reality, all subtasks scores are positive correlated with one another. The lack of correlation between scores on interpretive diversity and other false belief tasks suggests that the exclusion of these items may actually improve the reliability of the Aggregate ToM score.
- **Freedom from bias:** Children's scores in the second-order and advance false belief subtasks correlated significantly with their verbal abilities, as measured by the BPVS. On the other hand, age and nonverbal abilities do not appear to affect children's scores in the ToM battery.

## Discussion

Based on the criterion of **item difficulty**, most of the items in battery were appropriate for the target sample. The exception is Strange Stories, which needed to be replaced. A more age-appropriate version of the advance false-belief tasks is the stories used by Happe (1995), which was described in Section 4.4. One issue that had been raised with the Happe stories is the equivalence of the 12 different story types, where each story corresponds to a different type of advance false belief attribution. However, it can be argued that this issue is of less relevance for the present thesis, as the items will be used to gauge children's overall theory of mind abilities (based on aggregate ToM), rather than an analysis of discrete types of advance false beliefs attribution.

Based on the criterion of internal reliability, with the exception of items for interpretive diversity, all other items contributed positively to the overall reliability for the Aggregate ToM score. The interpretive diversity items (i.e. Cow/Dalmatian) were not correlated with other false belief tasks in the battery, and its removal actually improved the reliability index of the Aggregate ToM. This suggests that to strengthen the reliability of the test battery, it may be necessary to exclude the interpretive diversity items from the test battery.

Analyses of response bias suggested that for the group of children tested, differences in Aggregate ToM scores were not due to age related differences, or differences in non-verbal learning abilities. Significant correlations were found between children's verbal abilities and scores on second-order and advance false belief subtasks. There are several possible explanations for the correlation. Firstly, it maybe the second-order and advance false belief subtasks include items with significantly higher language demands.

Another possible explanation for the correlation between verbal IQ and second-order and advanced false belief attribution is that in addition to mentalising abilities, children in the field-test needed a higher-level language competency to cope with the higher-level theory of mind tasks. This is consistent with the findings from other studies where it has been shown that compared with normal children, children with autism require a higher level of language skill to pass theory of mind tests (Happe, 1995; Fisher, 2005). The correlations between verbal abilities and second order and advanced false belief subtasks may reflect a characteristic that is unique to the sample of children in the study, i.e. children with autism. To investigate this further, the ToM Battery would need to be tested with normal children with similar age and ability levels. However, this is beyond the scope of the present field-test.

### Conclusion

Based on the findings of the field test, the items that would be suitable for use in the ToM Battery in the present thesis are shown in Table L8. Additional modifications are needed to make the content more culturally appropriate for children from Singapore, and these will be discussed as part of the methodology for study 2 (See Chapter 5).

**Table L8 : Items in ToM battery Used in Present Thesis**

Test Items	ToM Construct	Reference (original use)
1. Sally-Ann	1 <sup>st</sup> order false belief	Baron-Cohen et al (1985)
2. Smarties	1 <sup>st</sup> order false belief	Perner et al (1987)
3. Ice-cream man	2 <sup>nd</sup> order false belief	Baron-Cohen (1989)
4. Birthday puppy	2 <sup>nd</sup> order false belief	Sullivan et al (1994)
5. Penny Hiding	Deception	Baron-Cohen (1992)
6. Strange stories	Advance false-belief	Happe (1995)
7. Composite from 1 to 6	Aggregate theory of mind abilities	NA (present thesis)

## APPENDIX M

### DESCRIPTION OF THE THEORY OF MIND BATTERY

This description of the Theory of Mind Battery comprised of the following:

- Part 1 : Overview of ToM Battery
- Part 2 : Instructions for individual items
- Part 3 : Pictures of materials used
- Part 4 : Score sheets

#### PART 1 : OVERVIEW

The Theory of Mind (ToM) Battery used in the present thesis measures children's theory of mind abilities. The items were selected based on mentalising tasks that have been used in autism research. A small scale study was conducted to identify the items that show good reliability and adequate discriminant validity (see Appendix L). The identified items were then adapted to suit the culture and language of children in Singapore. The final items in the ToM Battery that is used in the present thesis are as follows:

Items	Construct	Reference (origin)
1. Jim-Rosie	1 <sup>st</sup> order false belief	Baron-Cohen et al. (1985)
2. Smarties	1 <sup>st</sup> order false belief	Perner et al. (1987)
3. Coin Hiding	Deception	Baron-Cohen (1992)
4. Ice-cream man	2 <sup>nd</sup> order false belief	Baron-Cohen (1989)
5. Birthday present	2 <sup>nd</sup> order false belief	Sullivan et al. (1994)
6. Strange Stories	Advance false belief	Happe et al. (1985)

In the ToM Battery, children's performance in all the items were combined to reflect their aggregated score in theory of mind abilities.

#### PART 2 : INSTRUCTIONS FOR INDIVIDUAL ITEMS

##### Introduction to the ToM Battery

Say: *I am going to tell you some stories and I want you to listen very carefully because I will ask you some important questions about the stories. Ready? Let's begin.*

<b>Item 1</b>	<b>'Jim and Rozie'</b>	
Description	The tester narrates the story and uses the materials to mime the actions.	
Materials	Jim & Rosie dolls; box with cover; open basket covered with a small handkerchief; a marble. See set-up position in picture 1.	
Instructions	<p><b>This is Jim and this is Rozie</b> (point to each doll respectively).  <b>Jim has a basket and Rozie has a box</b> (point to basket and box respectively).  <b>Jim has a marble</b> (show marble).  <b>He is going to put his marble in his basket</b> (mime actions) <b>to keep it safe while he goes out</b> (remove Jim from sight).  <b>But while Jim is out, naughty Rozie takes Jim's marble out of the basket and puts it in her box</b> (mime actions).</p>	
Questions:	Acceptable answers:	
<b>Control Q1: Where is the marble really?</b>	Box	
<b>Control Q2: Where did Jim put the marble in the beginning?</b>	Basket	
<b>Test Q: When Jim comes back, where will he think his marble is?</b>	Basket	

Scoring criteria	Score 1 for Test Q only if the answers for both control questions are correct.
------------------	--

Item 2	<b>Smarties</b>
Description	The tester asks the child to guess the contents of a chocolate container, then shows its actual content, i.e. a pencil. The child is then asked to predict the responses of his/her friend to the same stimulus.
Materials	Smarties or 'M&M' chocolate tube (emptied); and contents replaced with a pencil (see picture 2)
Instructions	<p>(Show the child the sealed Smarties tube) <b>What do you think is inside?</b>          (Open the tube) <b>What is it?</b>          (Shows the child the pencil) <b>It's a pencil!</b>          (Put back the pencil and closed the lid)</p> <p><b>In a minute, your friend is going to come in. He has not seen this tube yet. When he comes in I am going to show him this tube, closed up just like this. I am going to ask him 'What's in here?'</b></p>
Questions:	Acceptable answers:
<b>Test Q: What will he say?</b>	Smarties; chocolate; sweets; M&M
<b>Control Q: What is really inside?</b>	Pencil
Scoring criteria	Score 1 for the Test Q only if the answer for the control question is correct.

Item 3	<b>Coin Hiding Game</b>
Description	This test is based on a commonly played children's game. There are two sets of trials: in the first set the tester hides the coin while the child guesses; and the role is reversed for second set. The scores in this test are based on the child's attempts to hide the coin in the second set.
Materials	One coin.
Instructions	<p><b>Now we are going to play a game. I am going to hide this coin in one of my hands, and you will have to guess which hand I have hidden the coin in. We'll do this a few times and then you will have a go at hiding the coin and I will have to guess. First is my turn.</b></p> <p>(Set 1: The tester hides the coin with hands on his/her back; and then extend both hands with fists closed. Note: do not cross the hands (see picture 3). 5 trials are given, and the coin should be hidden in the following sequence: left; right; right; left; right).</p> <p><b>Now you have a go. Hide the coin in your hands; hide it well so that I can't see it. I will try and guess.</b></p> <p>(Set 2: The child is given 6 trials to hide the coin, and the tester makes a guess after each trial.)</p>
Scoring criteria	<p>For each of the 6 trials, score 1 if child performs the task of hiding the coin <u>without</u> any of the following errors:</p> <ul style="list-style-type: none"> <li>• Does not hide both hands (both or one hands can be seen throughout the game)</li> <li>• Does not bring both hands forward, i.e. extends only 1 hand out.</li> <li>• Hands not closed</li> <li>• Coin can be seen, i.e. not hidden.</li> </ul> <p>(Note : Total maximum score for Item 3 is 6)</p>

<b>Item 4</b>	<b>Ice Cream Story</b>	
Description	The tester narrates the story and uses the materials to mime the actions.	
Materials	John and Mary dolls; ice-ream van; 3-di set-up of housing estate showing Mary's house, John's house, a park and a field (see picture 4).	
Instructions	<p><b>This is John and this is Mary</b> (show dolls).  <b>They live in this housing estate. This is John's house, Mary's house, a park and a field</b> (show respective places in the 3-di set-up).  (Check child's understand) <b>Which is John and which is Mary?</b></p> <p><b>Here they are in the park. Along comes an ice-cream man</b> (mime the ice-cream van driving along the road and parking in the park near the children).</p> <p><b>John wants to buy an ice cream but he has left his money at home. He is very sad.</b>  <b>"Don't worry", says the ice cream man. "I'll be here in the park all afternoon".</b>  <b>"Oh good!" says John. "I'll come back here in the afternoon to buy an ice-cream".</b></p> <p>(Check understanding) <b>Where did the ice-cream man say to John he would be all afternoon?</b> (Answer: Park)</p> <p><b>So John goes home. He lives in this flat. Now the ice-cream man says, "I am going to drive my van to the field to see if I can sell my ice-cream there".</b></p> <p>(Check understanding) <b>Where did the ice cream man say he was going? Did John hear that?</b> (Answer: Field; No)</p> <p><b>The ice cream man drives over to the field. On his way, he passes John' flat. John sees him and asks, "Where are you going?" The ice cream man says, "I am going to sell my ice cream in the field". So off he drives to the field.</b></p> <p>(Check understanding) <b>Where did the ice cream man tell John he was going? Does Mary know that the ice cream man has talked to John?</b> (Answer : Field ; No)</p> <p><b>John goes to buy the ice cream. Now Mary goes home. She lives in this flat. Then she goes to John's flat. She knocks on the door and says, "Is John home?" "No", says John's mother, "He has gone to buy ice-cream".</b></p>	
Questions:	Acceptable answers:	
<u>Test Q:</u> <b>Where does Mary think John has gone to buy the ice cream?</b>	Park	
<u>Control Q1:</u> <b>Where did John really go to buy the ice cream?</b>	Field	
<u>Control Q2:</u> <b>Where was the ice cream man in the beginning?</b>	Park	
Scoring criteria	Score 1 for Test Q only if the answers for both control questions are correct.	

<b>Item 5</b>	<b>Birthday Present Story</b>	
Description	The tester narrates the story and uses the materials to mime the actions.	
Materials	Peter and Peter's Mum dolls; toy skateboard; 3-di mock-up of the interior of an apartment showing a bedroom; a kitchen; a storeroom; and a hall, where a 'phone' is placed (see picture 5).	
Instructions	<p><b>This is Peter, and this is his mum</b> (show dolls).  <b>This is their flat: this is the hall, bedroom, kitchen and storeroom</b> (point to respective places in the mock-up apartment).</p> <p><b>Tonight is Peter's birthday and his mum is surprising him with a special present: a new skateboard</b> (show the toy skateboard).  <b>She has hidden the skateboard in the storeroom</b> (put the toy skateboard in the storeroom).</p> <p><b>Peter says, "Mum, I really hope you get me a skateboard for my birthday".</b>  <b>Remember, mum wants to surprise Peter with the skateboard. So instead of telling Peter she has got him a skateboard, mum says, "Sorry Peter, I did not get you a skateboard for your birthday. I got you a great book instead".</b>  <b>Then mum went into the bedroom for a nap</b> (mime the actions with the doll).</p> <p><b>Peter says, "I'm going out to play". Before he goes out, Peter looks for his roller-skates. He looks for it in the storeroom, and finds his birthday skateboard!</b>  <b>Peter says to himself, "Wow, mum did not get me a book for my birthday, she really got me a birthday skateboard!"</b></p> <p><b>Control Q1: Does Peter know that his mum got him a skateboard for his birthday?</b>  (Answer: Yes)</p> <p><b>Control Q2: Does mum know that Peter saw the birthday skateboard?</b>  (Answer: No)</p> <p><b>Now the telephone rings</b> (point to the phone) <b>Ring!! Peter's mum picks up the phone. Peter's grandmother calls to find out what time the birthday party is.</b>  <b>Grandma asks mum on the phone, "Does Peter know what you really got him for his birthday?"</b></p>	
Questions:	Acceptable answers:	
	<b>Test Q1: What does mum say to grandma?</b>	She says "No"/ No.
	<b>Test Q2: Now remember, mum does not know that Peter saw what she got him for his birthday. Then grandma says to mum, "What does Peter think you got him for his birthday?"</b> <b>What does mum say to grandma?</b>	A book/ storybook.
Scoring criteria	Score 1 for each Test Qs only if the answers for both control questions are correct. (Note: Total maximum score for this item is 2)	

<b>Item 6</b>	<b>Strange Stories</b>
Description	The tester shows the child a series of short stories and reads them, one at a time, to the child. At the end of each story, two questions are asked.
Materials	Booklet of strange stories (see picture 6)
Instructions	<b><i>I am going to read to you some short stories. Listen carefully and at the end of each story, I will ask you some questions. Ready?</i></b>
Scoring criteria	<p>Each Strange Story has two questions:</p> <p><b>Q1. Comprehension Question</b>, e.g. "Is it true what X says?"  The score for the comprehension is 1 point for each correct answer.</p> <p><b>Q2. Justification Question</b>, e.g. 'Why did X say that?"  The score for the justification question can be 0, 1 or 2 points, depending on the child's response (see scoring guidelines for each story).</p> <p><b>Note:</b> If child gives a 1-point answer, prompt by saying '<b>Can you tell me some more?</b>'</p>
<b>Story 1</b>	<b><i>Kate and Emma are playing in the house. Emma picks up a banana from the fruit bowl and holds it up to her ear. She says to Kate, "Look! This banana is a telephone!"</i></b>
Questions	<p>Q1: <b>Is it true what Emma said?</b> (Answer: No)</p> <p>Q2: <b>Why did Emma say that?</b></p>
2 point	Reference to make belief/pretend play, e.g. they are only playing; she's pretending that the banana is a phone.
1 point	Reference to physical feature, e.g. because the banana looks like a telephone; because she can hold the banana like a telephone.
0 point	Factually incorrect or irrelevant responses, e.g. because it is a telephone.
<b>Story 2</b>	<b><i>Mary has a cough. All through lunch she coughs and coughs and coughs. Father says, "Poor Mary, you must have a frog in your throat!"</i></b>
Questions	<p>Q1: <b>Is it true what Mary's father said?</b> (Answer: No)</p> <p>Q2: <b>Why did Mary's father say that?</b></p>
2 point	Reference to simile, e.g. because she's croaking like a frog; her voice is sore and sounds like a frog croaking; that's a simile.
1 point	Factually correct information but missed the point of the simile, or incomplete explanation of the simile, e.g. because she has a sore throat; because she's coughing badly; because she is coughing repeatedly.
0 point	Factually incorrect or irrelevant responses.

<b>Story 3</b>	<b>Daniel and Kim see Mrs. Tan coming out of the hairdresser's one day. She looks a bit funny because the hairdresser has cut her hair much too short. Daniel says to Kim, "She must have been in a fight with a lawn-mower!"</b>
Questions	Q1: <b>Is it true what Daniel said?</b> (Answer: No) Q2: <b>Why did Daniel say that?</b>
2 point	Recognition that Daniel was making a joke; or that the statement was said in humour, e.g. Daniel was trying to be funny; he was joking; he made a joke about her hair.
1 point	Reference to physical/factual aspect, e.g. because her hair looks like that; because her hair is too short.
0 point	Factually incorrect or irrelevant responses.
<b>Story 4</b>	<b>One day Aunty Jane came to visit Peter. Now Peter loves his aunt very much, but today she is wearing a new hat, which Peter thinks is very ugly indeed. Peter thinks his aunt looks silly in it, and much nicer in her old hat. But when Aunty Jane asks Peter, "How do you like my new hat?" Peter says, "Oh, its very nice."</b>
Questions	Q1: <b>Is it true what Peter said?</b> (Answer: No) Q2: <b>Why did Peter say that?</b>
2 point	Reference to a white lie, or wanting to spare Aunty Jane's feelings, or some implications that this is for her benefit and not just to avoid rudeness, e.g. he's not telling the truth because he does not want to hurt her; because she will be upset if she tells her that the hat is ugly.
1 point	Reference to trait (e.g. because he is a nice boy; because he does not want to be rude) or relationship (e.g. because he loves his aunt). No reference to aunt's feelings or thoughts, or incomplete explanations.
0 point	Reference to incorrect or irrelevant feelings/facts, e.g. because he likes the hat; because he wants to trick his aunt.
<b>Story 5</b>	<b>William is a very untidy boy. One day his mother comes into his bedroom, and it is even more messy than usual! There are clothes, toys, and comics, everywhere. William's mother says to William, "This room is a pigsty!"</b>
Questions	Q1: <b>Are there pigs in William's room?</b> (Answer: No) Q2: <b>Why did William's mother say that his room is a pigsty?</b>
2 point	Reference to a simile or metaphor, e.g. she using the saying 'dirty like a pigsty'; a pig sty is dirty, and William's room is also dirty, so she's saying that his room is like a pigsty because its is also dirty like a pigsty.
1 point	Reference to the physical fact, e.g. because the room has comics, toys and clothes everywhere; or incomplete explanation of the metaphor, e.g. because the room is dirty.
0 point	Factually incorrect, e.g. because the room is full of toy pigs.

<b>Story 6</b>	<b><i>A burglar who has just robbed a shop is making his getaway. As he is running home, a policeman sees him drop his wallet. He doesn't know the man is a burglar; he just wants to tell him he dropped his wallet. But when the policeman shouts out to the burglar, "Hey, you! Stop!" The burglar turns round, sees the policeman and gives himself up. He puts his hands up and admits that he did the break-in at the local shop.</i></b>
Questions	Q1: <b><i>Did the policeman know that the burglar had just robbed a bank?</i></b> (Answer: No) Q2: <b><i>Why did the burglar put his hands up and admit that he did the break-in at the local shop?</i></b>
2 point	Reference to the burglar's belief that the policeman knew that he burgled the shop, e.g. because he thinks that he has been caught; because he thinks that the policeman was after him for robbing the bank.
1 point	Reference to something factually correct in the story, e.g. he has just robbed a shop; because the policeman said 'Hey you, Stop!'
0 point	Factually incorrect or irrelevant responses.
<b>Story 7</b>	<b><i>Sarah and Tom are going on a picnic. It is Tom's idea; he says it is going to be a lovely sunny day for a picnic. But just as they are unpacking the food, it starts to rain, and soon they are both soaked to the skin. Sarah is angry. She says, "Oh yes, a lovely day for a picnic alright!"</i></b>
Questions	Q1: <b><i>Is it true what Sarah said?</i></b> (Answer: No) Q2: <b><i>Why did Sarah say that?</i></b>
2 point	Reference to understanding of sarcasm, e.g. she was angry but she does not want to scold him directly/obviously; she is showing her anger by saying the exact opposite of what she really means.
1 point	Reference to physical fact, e.g. it's raining; because Tom said it was going to be a lovely sunny day; or incomplete explanation, e.g. because she's angry; because she cannot have the picnic.
0 point	Factually incorrect or irrelevant responses.
<b>Story 8</b>	<b><i>During the war, the Red army captures a member of the Blue army. They want him to tell them where his army's tanks are; they know they are either by the sea or in the mountains. They know that the prisoner will not want to tell them, he will want to save his army, and so he will certainly lie to them. The prisoner is very brave and very clever; he will not let them find his tanks. The tanks are really in the mountains. Now when the other side asks him where his tanks are, he says, "They are in the mountains."</i></b>
Questions	Q1: <b><i>Where are the Blue army's tanks really, in the mountains or by the sea?</i></b> (Answer: mountains) Q2: <b><i>Why did the prisoner say the tanks are in the mountains?</i></b>

2 point	Reference to double bluff, e.g. he knows that the other army will not believe him and will look in the other place; he was pretending to lie when in fact he was telling the truth so that the Red army will do the opposite and look at the wrong place.
1 point:	Reference to outcome, e.g. he wants to save his army tanks; incomplete explanation of the double bluff, e.g. he wants to confuse the other army.
0 point:	Reference to motivations that missed the double bluff, e.g. he was very scared, he did not want to lie to them; because he had to tell the truth or they will kill him.
<b>Story 9</b>	<b><i>Chris is going to a fancy dress party. He is going as a ghost. He wears a big white sheet with eyes cut out to see through. As he walks to the party in his ghost costume, he bumps into Mr. Wang. It is dark, and Mr. Wang says, "Oh! Who is it?" Chris answers, "I'm the ghost, Mr. Wang!"</i></b>
Questions	Q1: <i>Is it true what Chris said; was he really a ghost?</i> (Answer: No) Q2: <i>Why did Chris say that?</i>
2 point	Reference to make believe or role-play, e.g. because he's pretending to be a ghost in the party; because he's acting as a ghost in the party.
1 point	Reference to factual information, e.g. because he's wearing a ghost costume; he looks like a ghost.
0 point	Factually incorrect or irrelevant responses.
<b>Story 10</b>	<b><i>Brian is always hungry. This morning he is going for a picnic with his friends in school. His teacher has prepared special snacks for everyone; it is his favorite food – chicken burgers. He is a very greedy boy, and he would like to have more chicken burgers than anyone else, even though his mother will have made him a nice lunch when he gets home! But everyone is allowed one burger and no more. When his teacher gives out the burger to him, he says, "Oh, please can I have three burgers, because I won't be having any lunch when I get home!"</i></b>
Questions	Q1: <i>Is it true what Brian said?</i> (Answer: No) Q2: <i>Why did Brian say that?</i>
2 point	Reference to the fact that Brian is eliciting sympathy (e.g. he wants the teacher to feel sorry for him); or being deceptive (e.g. he is lying/cheating so that the teacher will give him more burgers).
1 point	Reference to his state (e.g. he is greedy/hungry); or the intended outcome (e.g. so that he can get more sausages); or factual (e.g. because each person can get only one).
0 point	Factually incorrect or irrelevant responses.
<b>Story 11</b>	<b><i>Today, Jessie wants to go on the swings in the playground. But to get to the playground she knows she has to pass old Mr. Wong's house. Mr. Wong has a nasty fierce dog, and every time Jessie walks past the house the dog jumps up at the gate and barks. It scares Jessie very much, and she hates walking past the house because of the nasty dog. But Jessie does so want to play on the swings. Jessie's mother asks her, "Do you want to go out to the playground?" Jessie says, "No".</i></b>
Questions	Q1: <i>Is it true what Jessie said?</i> (Answer: No) Q2: <i>Why did Jessie say that?</i>

2 point	Reference to <u>both</u> conflicting emotions, e.g. Jessie wants to go to the playground but she's afraid of the dog.
1 point	Reference to <u>one</u> state only, e.g. she is scared of the dog; or the dog is nasty.
0 point	Factually incorrect or irrelevant responses.

<b>Story 12</b>	<b><i>Siti is playing in the garden with her doll. She leaves her doll in the garden when her mother calls her to come in for lunch. While they are having lunch, it starts to rain. Siti's mother asks Siti, "Did you leave your doll in the garden?" Siti answers, "No, I brought her in with me, mummy".</i></b>
Questions	Q1: <i>Is it true what Siti said?</i> (Answer: No) Q2: <i>Why did Siti say that?</i>
2 point	Reference to forgetting (e.g. she thinks she has brought the doll in; she forgot that she has left the doll outside); or reference to deception (e.g. she does not want her mum to scold her so she lied).
1 point	Reference to factual information but no reference to forgetting/deception, e.g. she left her doll in the garden.
0 point	Factually incorrect or irrelevant responses.

<b>Story 13</b>	<b><i>Late one night, old Mrs. Lim is walking home. She doesn't like walking home alone in the dark because she is always afraid that someone will attack her and rob her. She is really a very nervous person! Suddenly, out of the shadows comes a man. He wants to ask Mrs. Lim what time is it, so he walks towards her. When Mrs. Lim sees the man coming towards her, she starts to tremble and says, "Take my purse, and just don't hurt me please!"</i></b>
Questions	Q1: <i>Did the man really wanted to rob Mrs Lim?</i> (Answer: No) Q2: <i>Why Mrs Lim say that?</i>
2 point	Reference to Mrs Lim's belief that he was going to harm/rob her or ignorance of his real intentions, e.g. she thinks that he was going to rob her; she does not know that he just want to ask her what time.
1 point	Reference to her trait (e.g. she's a very nervous person); or her state (e.g. she was scared); or her intentions (e.g. she said that so that he would not hurt her). No suggestion that the fear was unnecessary.
0 point	Factually incorrect or irrelevant responses.

<b>Story 14</b>	<p><b><i>Helen waited all year for Christmas, because she knew at Christmas she could ask her parents for a rabbit. Helen wanted a rabbit more than anything in the world. At last Christmas Day arrived, and Helen ran to unwrap the big box her parents had given her. She felt sure it would contain a little rabbit in a cage. But when she opened it, with all the family standing round, she found her present was just a boring old set of encyclopaedia, which Helen did not want at all! Still, when Helen's parents asked her how she liked her Christmas present, she said, "It's lovely, thank you. It's just what I wanted.</i></b></p>
<b>Questions</b>	<p>Q1: <b><i>Did Helen really like the presents?</i></b> (Answer: No)</p> <p>Q2: <b><i>Why did Helen say that?</i></b></p>
2 point	Reference to white lie or wanting to spare her parents' feelings, e.g. she told a lie because the truth would hurt them; implication that this is for the parents' benefit and not just her desire to avoid rudeness or insult, e.g. she does not want to make her parents feel sad/disappointed that they had given her the wrong presents.
1 point	Reference to trait (e.g. because she's a nice/polite girl); or relationship (e.g. because she likes/loves her parents); or purely motivational justification without reference to parents' thoughts or feelings (e.g. so that they won't shout at her or scold her); or incomplete explanations (e.g. she's lying/pretending to like the books).
0 point	Factually incorrect or irrelevant responses.
<b>Story 15</b>	<p><b><i>John hates going to the dentist, because every time he goes to the dentist he needs a filling, and that hurts a lot. But John knows that when he has a toothache, his mother always take him to the dentist. Now John a has bad toothache at the moment, but when his mother notices he is looking ill and asks him, "Do you have toothache, John?" John says, "No, mummy".</i></b></p>
<b>Questions</b>	<p>Q1: <b><i>Did John have a bad toothache?</i></b> (Answer: Yes)</p> <p>Q2: <b><i>Why did John say that?</i></b></p>
2 point	Reference to lying, e.g. he was bluffing because he does not want to go to the dentist.
1 point	Incomplete explanations without reference to lying, e.g. because he does not want to go to the dentist.
0 point	Factually incorrect or irrelevant responses.
<b>Story 16</b>	<p><b><i>One day, while she was playing in the house, Anna accidentally knocks over and breaks her mother's favourite crystal vase. Oh dear, when mother finds out she will be very angry! So when Anna's mother comes home and sees the broken vase and asks Anna what happened, Anna says, "The dog knocked it over, it wasn't my fault!"</i></b></p>
<b>Questions</b>	<p>Q1: <b><i>Did the dog knocked over the vase?</i></b> (Answer: No)</p> <p>Q2: <b><i>Why did Anna say that?</i></b></p>

2 point	Reference to lying, e.g. she's not telling the truth; she blamed the dog because she does not want to get scolded.
1 point	Incomplete explanation, without reference to lying, e.g. because she does not want to get scolded.
0 point	Factually incorrect or irrelevant responses.
<b>Story 17</b>	<i>Today James is going to Claire's house for the first time. He is going over for tea, and he is looking forward to seeing Claire's dog, which she talks about all the time. James likes dogs very much. When James arrives at Claire's house, Claire runs to open the door, and her dog jumps up to greet James. Claire's dog is huge, it's almost as big as James! When James sees Claire's huge dog he says, "Claire, you haven't got a dog at all! You've got an elephant!"</i>
Questions	Q1: <i>Did Claire have an elephant or a dog?</i> (Answer: Dog) Q2: <i>Why did James say that?</i>
2 point	Recognition that James was making a joke or was saying things in humour, e.g. he is a joker; he is trying to be funny.
1 point	Reference to state without referring to joke/humour, e.g. because the dogs looks like an elephant; because it is very big.
0 point	Factually incorrect or irrelevant responses.
<b>Story 18</b>	<i>Mark and Adam are having great fun! They have turned the kitchen table upside down and they are sitting in it, paddling along with rolled-up newspapers. When their mother comes in she laughs. "Whatever are you two doing?", she asks. "This table is a pirate ship", says Adam, "And you had better get in too before you sink – because you are standing in the sea!"</i>
Questions	Q1: <i>Is it true what Adam said?</i> (Answer: No) Q2: <i>Why did Adam say that?</i>
2 point	Reference to make belief, e.g. they are only playing; it's a pretend game.
1 point	Reference to physical feature, e.g. because the table looks like a ship.
0 point	Factually incorrect or irrelevant responses.
<b>Story 19</b>	<i>Ann's mother has spent a long time cooking Ann's favourite meal: fish and chips. But when she brings it in to Ann, she is watching T.V. and she doesn't even look up, or say thank you. Ann's mother is angry and says, "Well that's very nice, isn't it! That's what I call politeness!"</i>
Questions	Q1: <i>Is it true what Ann's mother said?</i> (Answer: No) Q2: <i>Why did Ann's mother say that?</i>

2 point	Reference to understanding of sarcasm, e.g. Ann's mother wanted to show her anger by saying the exact opposite of the truth; she was saying one thing but meant the opposite.
1 point	Reference to factual information or incomplete explanations of sarcasm, e.g. because she is angry; because Ann is rude.
0 point	Factually incorrect or irrelevant responses.
<b>Story 20</b>	<b><i>Sam is a big liar. Sam's brother Jim knows this; he knows that Sam never tells the truth! Now yesterday Sam stole Jim's ping-pong bat, and Jim knows Sam has hidden it somewhere, though he can't find it. He's very angry. So he finds Sam and he says, "Where is my ping-pong bat? You must have hidden it either in the cupboard or under your bed, because I've looked everywhere else. Where is it, in the cupboard or under your bed?" Simon tells him the ping-pong bat is under his bed.</i></b>
Questions	Q1: <b><i>Where did Simon said he has hidden the ping-pong bat?</i></b> (Answer: under the bed) Q2: <b><i>Why would Jim look for the ping-pong bat in the cupboard?</i></b>
2 point	Reference to Jim knowing that Simon lies, e.g. Jim knows that Simon is a liar and will make him look in the wrong places.
1 point	Reference to factual information without implication/reference to lying, e.g. that's really where it is; because he thinks it is in the cupboard.
0 point	Reference to general non-specific information, e.g. because he has looked everywhere; because it's either in the cupboard or under the bed; or inaccurate interpretation, e.g. he looked in the bed but it was not there so he looked for it in the cupboard.
<b>Story 21</b>	<b><i>Jill wanted to buy a kitten, so she went to see Mrs. Wee, who had lots of kittens she didn't want. Now Mrs. Wee loved the kittens, and she wouldn't do anything to harm them, though she couldn't keep them all herself. When Jill visited she wasn't sure she wanted one of Mrs. Wee's kittens since they were all males and she had wanted a female. But Mrs. Wee said, "If no one buys the kittens, I'll just have to drown them!"</i></b>
Questions	Q1: <b><i>Will Mrs Wee really drown the kittens if no one buys them?</i></b> (Answer: No) Q2: <b><i>Why did Mrs Wee say that?</i></b>
2 point	Reference to the fact that Mrs Wee is eliciting Jill's sympathy for the kittens, e.g. she wanted Jill to feel sad/sorry for the kittens; or being deceptive, e.g. she was lying so that Jill thinks she will have to save the kittens from drowning.
1 point	Reference to outcome (e.g. to make Jill keep the kittens); or factual information (e.g. she couldn't keep all the kittens; she loved the kittens).
0 point	Factually incorrect or irrelevant responses.

<b>Story 22</b>	<p><b>Jane and Sally are best friends. They both entered the same painting competition. Now Jane wanted to win this competition very much indeed, but when the results were announced it was her best friend Sally who won, not her. Jane was very sad she had not won, but she was happy for her friend, who got the prize. Jane said to Sally, "Well done, I'm so happy you won!" Jane said to her mother, "I am sad I did not win that competition!"</b></p>
<b>Questions</b>	<p>Q1: <b>What did Jane say to Sally?</b> (Answer: I'm happy); <b>What did Jane say to her mum?</b> (Answer: I'm sad).</p> <p>Q2: <b>Why did Jane say that?</b></p>
2 point	Reference to both conflicting emotions, e.g. Jane is happy for Sally but sad that she herself did not win the competition.
1 point	Reference to one state only, e.g. Jane is sad that she did not win; or Jane is happy that Sally won.
0 point	Factually incorrect or irrelevant responses, e.g. they are best friends.
<b>Story 23</b>	<p><b>At school today John was not present. He was away ill. All the rest of Ben's classes were at school though. When Ben got home after school his mother asked him, "Was everyone in your class at school today?" Ben answers, "Yes mummy".</b></p>
<b>Questions</b>	<p>Q1: <b>Is it true what Ben said?</b> (Answer: No)</p> <p>Q2: <b>Why did Ben say that?</b></p>
2 point	Reference to forgetting, e.g. Ben thought that everyone was in school; or ignorance, e.g. he did not know that John was absent.
1 point	Reference to factual information but no reference to forgetting, e.g. because John was absent in school.
0 point	Factually incorrect or irrelevant responses.
<b>Story 24</b>	<p><b>On Christmas Eve Alice's mother takes her to the big shopping center in town. They go to look in the toys department. In the toys department Mr. Lim, Alice's next-door neighbour is dressed up as Santa Clause, giving out sweets to all the children. Alice thinks she recognizes Mr. Lim, so she runs up to him and asks, "Who are you?" Mr. Lim answers, "I'm Santa Clause!"</b></p>
<b>Questions</b>	<p>Q1: <b>Is it true what Mr Lim said; is he really Santa Clause?</b> (Answer: No)</p> <p>Q2: <b>Why did Mr Lim say that?</b></p>
2 point	Reference to make believe/pretence, e.g. because he is pretending to be Santa; so that the children will think he's the real Santa.
1 point	Reference to factual information, e.g. because he's wearing a Santa's costume; because he looks like Santa.
0 point	Factually incorrect or irrelevant responses.

### **PART 3 : Pictures of Materials used in ToM Battery**

#### **Picture 1**

##### **Item 1 : Materials for 'Jim and Rosie'**



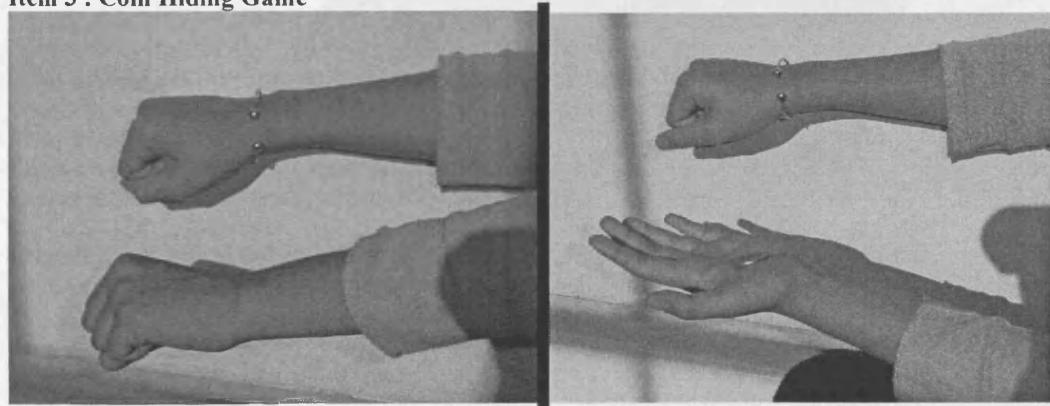
#### **Picture 2**

##### **Item 2 : Materials for 'Smarties'**



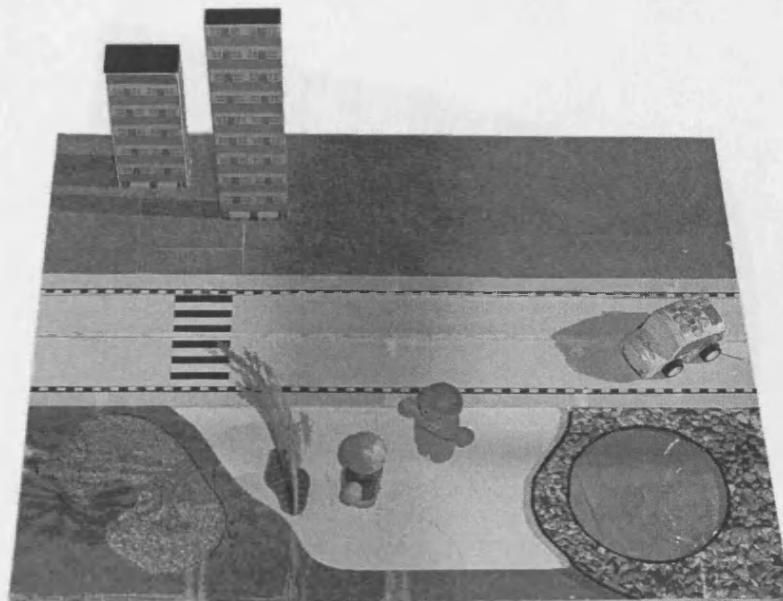
#### **Picture 3**

##### **Item 3 : Coin Hiding Game**



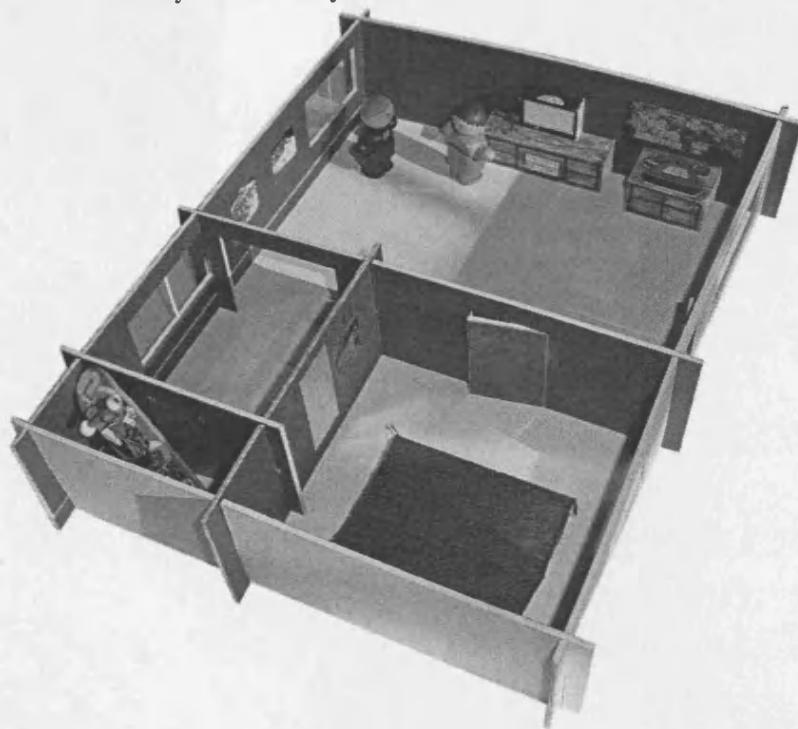
**Picture 4**

**Item 4: Materials for Ice Cream Story**

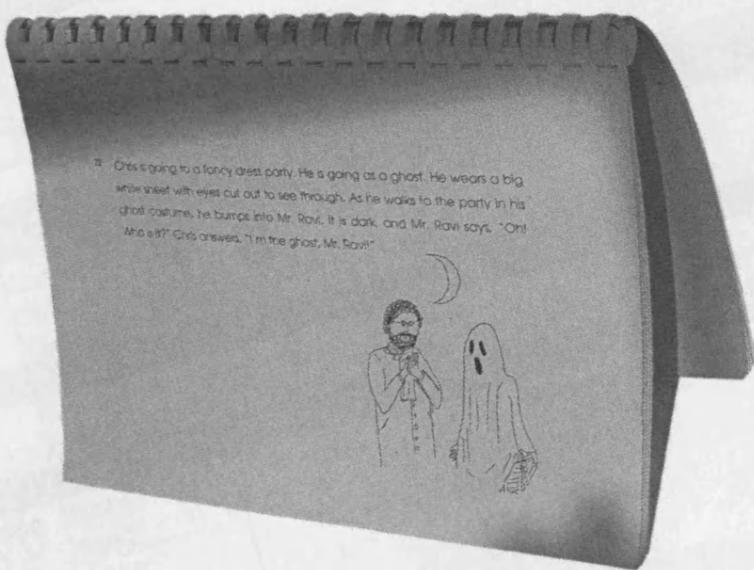


**Picture 5**

**Item 5 : Materials for Birthday Present Story**



Picture 6  
Item 6: Strange Stories Booklet



**PART 4 : Score sheet for the ToM Battery**

**REPORT OF MIND BATTERY - SCORE SHEET**

Demographic Information		Birth & Test Dates		
Name		Year	Month	Day
Sex		Birth date		
Class		Test Date		
School		Test Age		

Item No.	Subsidiary Test	Score	Child's responses
1. Jim & Rozie	Control Q1	Pass	Fail
	Control Q2	Pass	Fail
	Test Q1	1	0
2. Smarties	Test Q	1	0
	Control Q	Pass	Fail
3. Coin Hiding	Trial 1	1	0
	Trial 2	1	0
	Trial 3	1	0
	Trial 4	1	0
	Trial 5	1	0
	Trial 6	1	0
4. Ice Cream Story	Tes Q	1	0
	Control Q1	Pass	Fail
	Control Q2	Pass	Fail
5. Birthday Present Story	Control Q1	Pass	Fail
	Control Q2	Pass	Fail
	Test Q1	1	0
	Test Q2	1	0

Story	Comprehension	Score	Responses
Story 1	Comprehension	1 Pass	0 Fail
	Justification	2	1 0
Story 2	Comprehension	1 Pass	0 Fail
	Justification	2	1 0
Story 3	Comprehension	1 Pass	0 Fail
	Justification	2	1 0
Story 4	Comprehension	1 Pass	0 Fail
	Justification	2	1 0

Story 5	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 6	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 7	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 8	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 9	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 10	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 11	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 12	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 13	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 14	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 15	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 16	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 17	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 18	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 19	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0
Story 20	Comprehension	1 Pass	0 Fail	
	Justification	2	1	0

Story 21	Comprehension	1	Pass	0	Fail
	Justification	2		1	0
Story 22	Comprehension	1	Pass	0	Fail
	Justification	2		1	0
Story 23	Comprehension	1	Pass	0	Fail
	Justification	2		1	0
Story 24	Comprehension	1	Pass	0	Fail
	Justification	2		1	0

SUMMARY		
Construct	Item	Total Score
1 <sup>st</sup> order false belief	1. Jim-Rosie	
1 <sup>st</sup> order false belief	2. Smarties	
Deception	3. Coin Hiding	
2 <sup>nd</sup> order false belief	4. Ice-cream man	
2 <sup>nd</sup> order false belief	5. Birthday present	
Advance false belief	6. Strange Stories	
Aggregate ToM Score (Grand Total)		

## APPENDIX N

### STUDY 2 : SAMPLE OF THE INVITATION LETTER AND INFORMED CONSENT FORMS

To Parents,  
Thru'  
(Name of Principal and school)

Dear Parent,

#### **Invitation to Participate in Research on Autistic Spectrum Disorder**

I am a Chartered Educational Psychologist from the Ministry of Education and currently conducting a research study on the assessment of special educational needs of children with autistic spectrum disorder (ASD) in Singapore. The aim of this study is to identify reliable and valid measures for assessing the learning abilities and needs of children with ASD in Singapore. This research is undertaken as part of a PhD programme in Psychology at University College London<sup>45</sup>.

I would like to invite your child to participate in this study. Your child's participation would involve 3 sessions of individual assessment with me. Each session would be between 1 to 3 hours and involve activities such as reading stories, puzzles and games. All information about individual children will be kept strictly confidential. Parents could have access to feedback the findings for their child, if requested.

The outcomes of this study would go a long way towards improving our understanding of the learning needs of children with ASD in Singapore, for example, it could highlight the most effective and appropriate ways of assessing the learning abilities in children with autism. I would greatly appreciate it if you would accept this invitation by completing the enclosed reply slip. If you need further clarifications, please feel free to contact me either at this e-mail address [s.aljunied@ucl.ac.uk](mailto:s.aljunied@ucl.ac.uk) or telephone [02078365420](tel:02078365420).

Thank you very much.

Yours sincerely,

Mariam Aljunied (C.Psychol)  
Psychology Department  
University College London

---

<sup>45</sup> This study is approved by the University College London Ethics Committee on the Ethics of Non-NHS Human Research.

## ***Reply Slip***

Name of child :

School : Class :

Age :

**Name of parent(s):**

E-mail :

Address :

Contact : Tel (Home) Tel (Office)  
Tel (HP)

✓ I accept the invitation to participate in the research. I am happy for the researcher to contact me to provide more details on the study.

\_\_\_\_\_  
Name of parent

Signature: \_\_\_\_\_

Date :

***PLEASE RETURN THE COMPLETED FORM TO (NAME OF CONTACT PERSON IN SCHOOL) BY  
(DATE)***

### **Informed Consent Form**

(This form to be completed independently by the parent)

Title of Project :

#### **ASSESSMENT OF SPECIAL EDUCATIONAL NEEDS OF CHILDREN WITH ASD IN SINGAPORE.**

	<b>YES</b>	<b>NO</b>
I have read the invitation letter, which gives background information to the study.		
I have had a discussion with Ms Aljunied about the nature and purpose of the project and had the opportunity to ask questions.		
I understand that my child's involvement in the study is voluntary and that I can withdraw my child's participation at any time.		
I understand that following the study I can request feedback on the results concerning my child. I understand that the same information can be given to my child's school, with my consent.		

Signed : ..... Date: .....

Name in capitals : .....

## APPENDIX O

### STUDY 3: POWER-POINT SLIDES (STIMULUS PRESENTATION) USED IN FOCUS GROUP

**Focus Group Discussion**

**Assessment of Special Educational Needs of Children with Autism**

02 Feb 2005  
PARB Training Room  
Teacher's Network (Singapore)

**Objective**

- To obtain practitioners' views on the treatment utility of selected tests that could be used in the assessment of children with autism.

**Background**

- Part of an on-going research identifying ways indicators which are able to distinguish children with autism who are able to cope with mainstream school, from those who require special schooling.
- The research is focusing on two aspects : predictive validity and treatment validity.
- This FGD focuses on treatment validity:
  - The extent to which the indicators provide information that can be used to plan interventions for children with autism.

**Clarify terms...**

- ❖ Practitioners
  - Professionals who are directly involved in the assessment and providing advice for special educational needs children with autism
- ❖ Indicators
  - Indicators which has been shown to have significant correlations with special educational needs of children with autism:
    - IQ scores
    - Behavioural Assessment of Dysexecutive Syndrome
    - Theory of Mind battery
    - Cognitive Modifiability Battery

**Clarify terms...**

- Treatment utility
  - the extent to information from tests shows relevance and usefulness, particularly as it relates to social consequences (i.e. what the tests is used for in society).
  - contains the idea of cost-benefit analysis
  - contains an aspect *incremental validity* in that it requires an assessment procedure to improve prediction over and above existing procedures.

**Selection of Indicators**

- Based on a review of theories of autism, e.g.:
  - Theory of Mind hypothesis;
  - Executive Function theory;
  - Central Coherence theory;
- Based on a review of theories of assessment of learning abilities, e.g.:
  - Theory of general intelligence;
  - Theory of cognitive modifiability.
- Measures for each indicator were identified and tested with a sample of 52 children with autism in Singapore.

## Selection of Indicators

- From the study, 4 indicators showed significant correlations with children's special educational needs:
  - IQ scores (WISC-111, S'pore)
  - Scores on Theory of Mind battery
  - Behavioural Assessment of Dysexecutive Syndrome
  - Analogy Subtest of the Cognitive Modifiability battery (Tzuriel, 2001)

## What is the treatment utility?

- What types of test information are useful, particularly for children with autism?
- What are the potential benefits of using the test as part of assessment of children with autism?
- What are the potential difficulties of using the test as part of the assessment of children with autism?
- In what ways can the use of the test improve current practices in the assessment of children with autism?

## WISC-111 S'pore

- Based on WISC-111, an established and well used test
- Standardised test for intelligence
- Based on theory of 'general intelligence'
- Singapore norms for 6 to 11:12 yrs
- 12 subtests, yielding:
  - Performance IQ index
  - Verbal IQ index
  - Full scale IQ index
  - Factor based index, e.g.
    - Verbal comprehension

## BADS-C

- Test of executive function for children
- Original version (for adults) is a well established clinical test;
  - children's version is new (October 2003)
- Standardised test
  - Norms for children age 8 to 16 years
  - Norms according to 3 IQ bands

## BADS-C : Description

- Executive function is:
  - Executive function is the ability to consider alternatives in planning.
  - An umbrella term for all of the complex set of processes that underlie **flexible, goal-directed responses to novel or new situations**, in particular, novel situations that involve:
    - planning and decision making;
    - error correction or troubleshooting;
    - initiation of novel sequences of actions; or
    - the need to overcome a strong habitual response.
  - (Shallice & Burgess, 1991)
- Theory of Executive Function deficit in autism

## BADS-C : Description

- Six subtests, each designed to tap selected aspects of executive function skills, such as :
  - Flexibility
  - Novel problem solving
  - Sequencing actions
  - Using feedback
  - Planning
  - Impulsivity
  - Following instructions

### BADS-C : Examples of subtests

- Playing Card test
  - Testing flexibility & inhibition
- Zoo Map test 1
  - Novel problem solving; Sequencing; Planning, Impulsivity.
- Six Part test
  - Planning; flexibility and perseveration; using feedback

[Brief demonstration](#)

### BADS-C information

- Scaled scores for each subtest
  - range 1 to 19
- Overall BADS-C Scaled score
  - Mean 100, SD 15
- Observational record of relevant behaviors, e.g.:
  - time spent on planning (if any); perseveration; strategies used (e.g. whether response was systematic, efficient).
- DEX-C : Accompanying parent/teacher checklist on behaviours at home/school that reflects deficits in executive function (with norms for interpretation )
- Means scores for control and clinical groups provided for BADS-C subtest scores and DEX-C
  - Clinical groups : PDD, ADHD, TBI, DCD

### Ground rules...

- Confidentiality
- Anonymity
- No right answers
- Aim is to obtain views and opinions of test users/practitioners, not to judge
- Diversity of opinions is welcomed.

### Theory of Mind battery

- Test battery developed for previous study;
- Based on a selection of theory of mind tasks that have been used in autism research;
- Theory of Mind hypothesis (Baron-Cohen et al,19??)
  - Impairments in mentalising abilities (or the ability to think about thoughts) is the core deficit in autism
- Tasks selected to tap the whole range of mentalising abilities:
  - Field-tested to sub-select most-appropriate and discriminating tasks;
  - Adapted for local context

### Theory of Mind battery

- First order false beliefs:
  - 'Sally-Ann'
  - Smarties
- 2<sup>nd</sup> order false beliefs:
  - Ice-cream story
  - Birthday story
- Advance false beliefs:
  - Strange Stories

[Brief demonstration](#)

### Theory of Mind battery

- Total scores reflect impairments in theory of mind;
- Sub-component scores reflect deficits in specific mentalising abilities, e.g. 1<sup>st</sup> VS 2<sup>nd</sup> order;
- Qualitative observation of children's verbal comprehension and inferential skills, e.g.:
  - Can children understand factual information that does NOT involve mentalising skills?
  - Can children make inferences based on physical attributes?

### Cognitive Modifiability Battery (Tzuriel, 2001)

- A dynamic assessment test
  - The tests involve the interaction between the tester and the child;
  - There is a focus on learner metacognitive processes and responsiveness to intervention; and
  - Assessment procedures which follows the pretest-intervention-posttest format.

### CMB

- Measures "Cognitive modifiability"
  - The individual's receptivity or responsiveness to instructional intervention
- CMB is administered in three phases:
  - Pre-teaching, Teaching & Post-Teaching.
- Pre and Post-Teaching phases:
  - structured and standardized,
  - no assistance is provided,
  - yield quantitative information (norms for Kindergarten to Grade 3)
- Teaching phase:
  - full mediation (or assistance) is given
  - yields clinical information, i.e. qualitative description of children's responses.

### CMB – Analogies subtest

- Target skill : visual-spatial and analytical skills
- Completion of three dimensional pattern based on analogical transformation, e.g:
  - if A → B, C → ?.
- Transformations are based on changes in number, colour, height and position of wooden blocks in the window plates.
- Items are in increasing levels of difficulty.
- Each difficulty level has 3 parallel items (for use in pre, teaching and post phases).

*Brief demonstration*

### CMB Analogies subtest

- Pre-test scores
  - Children's current abilities in visual-spatial tasks;
- Post test scores
  - Children's abilities in visual-spatial tasks, after teaching;
- Teaching phase:
  - Qualitative observations of the types of mediation/teaching strategies that the child responded to;
  - Observations on the child's behaviour in learning situations that may interfere / facilitate teaching and learning.

## APPENDIX P

### STUDY 3: MODERATORS' GUIDE FOR FOCUS GROUP DISCUSSIONS

<b>Moderator's Guide for the Focus Groups Discussion</b>	
<u>Opening remarks by moderator</u>	
<ul style="list-style-type: none"> <li>• Re-iterate ground-rules:           <ul style="list-style-type: none"> <li>• Confidentiality &amp; anonymity of all participants assured</li> <li>• No 'right answers' for all the FGD questions</li> <li>• The aim is to obtain views and opinions of test users/practitioners; and not to judge</li> <li>• Diversity of opinions is welcomed.</li> </ul> </li> <li>• Introduce participants by first names</li> </ul>	
Questions	Probes
<p>With reference to the each of the four tests:</p> <ol style="list-style-type: none"> <li>1. What information derived from this test do you find useful for children with autism?</li> <li>2. In what ways would the use of this test improve the current practice of assessment of children with autism?</li> <li>3. What are the possible benefits of introducing the use of this test in the assessment of children with autism?</li> <li>4. What are the possible difficulties in using this test in the assessment of children with autism?</li> </ol>	<p>In what ways does information from the test tells you about children's level of functioning and learning problems?    Why is this information useful for you?    In what ways can information from the test help you to decide/advice what type/level of support/intervention needed by children with autism?    What other ways do you find information from the test useful for children with autism?</p> <p>Probe for examples/cases for clarity.    Is this not currently in practice? Why?</p> <p>Probe for examples/cases for clarity.    What other benefits do you see in using this test?    Why is this important for children with autism?</p> <p>Probe for examples/cases for clarity.    What other difficulties do you see in using this test?    Why or how do these difficulties arise?    Can these difficulties be overcome?</p>
<u>Closing remarks</u>	
Moderator to re-iterate anonymity and confidentiality; and invite participants to their express final thoughts.	

## APPENDIX Q

### STUDY 3: SAMPLE OF INFORMED CONSENT FORM FOR PARTICIPANTS OF FOCUS GROUPS

(This form to be completed independently by the participant)

Title of Focus Group Discussion :  
Assessment of Special Educational Needs of  
Children with Autism

	YES Please tick ✓
I have read the invitation letter, which gives background information to the focus group discussion.	
I understand that my involvement in the focus group discussion is voluntary and that I can withdraw my participation at any time.	
I understand that all information and opinions expressed during the discussion will be kept strictly confidential, with no participants' or specific institutions' names being used in the reporting of results.	
I understand that solely for the purpose of data analysis, the discussion will be audio-taped and the recording will be destroyed upon completion of this study. In the transcripts of the focus group discussion, all personal identifying information will be omitted.	

Signed : ..... Date: .....

Name in capitals : .....

## APPENDIX R

### STUDY 3: OPERATIONAL DEFINITIONS FOR ALL CATEGORY HEADINGS

**Code: Utility for intervention**

Description: Use of test info. to plan and design intervention for children.

Inclusion: Identifying general approaches to intervention; identify level of support needed; planning academic, cognitive, social and other interventions; planning IEP; evaluate intervention outcomes.

Exclusion: Use of test for diagnosis, referral to special schools, referral to other professionals (see code : Other Uses).

**Code: Utility for intervention. Identify general strategy/approach**

Description: Use of test info. to identify general teaching / intervention approach.

Inclusion: General statements about utility of the test for intervention, e.g. modality matching, use of test to highlight the need for a 'more structured' approach, to use more 'visual strategies'; to identify general guides/direction for teachers/parents.

Exclusion: Use of test to plan interventions for a specific aspect, i.e. academic, cognitive, social, IEP.

**Code: Utility for intervention. Identify level of support**

Description: Use of test information to determine the level, degree or extent of support or intervention needed.

Inclusion: Use of test information to decide if a particular intervention is needed in greater or lesser amount; to determine the urgency in which the support is needed; to decide how much time is needed for a particular intervention.

Exclusion: Use of test to evaluate intervention outcomes (see code: Plan/design Intervention - Evaluate Intervention/Outcomes).

**Code: Utility for intervention. Evaluate intervention/outcome**

Description: Use of test info. to evaluate intervention outcomes.

Inclusion: Use of test as a pre and post intervention measure; use of test to measure children's progress over time, after intervention.

Exclusion: Use of test to measure children's current level of functioning (see code: Assess Function); use of test to assess the quality of environment or current teaching input (see code: Assess Environment/Input).

**Code: Utility for intervention. Plan Academic intervention**

Description: Use of test info. for intervention that relates directly to academic tasks or subjects.

Inclusion: Intervention related to school curriculum, project work, school programmes and class-based activities (e.g. visual organizers for class time-table); use of test info. to decide class or group placements; use of test info. for IEPs; use of test infl. for exam accommodations.

Exclusion: Use of test info. to identify general strategy or guide for teachers (see code: Plan/Design Intervention - General Strategy/Approach).

**Code: Utility for intervention. Plan Social Intervention**

Description: Use of test information to plan interventions focusing on child's social and adaptive skills.

Inclusion: Interventions focusing on behaviours related to social interaction skills, e.g. adaptive skills in social contexts, role switching, perspective taking; mentalising (ToM) skills; avoiding bullies.

Exclusion: Use of test info to identify general strategies or guide for teachers (see code: Plan/Design Intervention - General Strategy/Approach).

**Code: Utility for Assessment. Assess environment/input**

**Description:** Use of test to assess the quality of learning environment, and the current teaching input provided for the child (i.e. before intervention).

**Inclusion:** Use of test to identify the strategies that had worked or not worked well for the child (prior to planned intervention); and assess the level of structure in the teaching/learning environment.

**Exclusion:** Use of test in evaluating the outcomes of a planned intervention, use of test for planning IEPs (see code: Evaluate intervention outcomes; and Plan academic intervention).

**Code: Utility for assessment. Assess General Cognitive Ability or Profile**

**Description:** Use of test to gauge children's general abilities.

**Inclusion:** Assessment of overall cognitive profile, discrepancies between verbal and non-verbal abilities/reasoning; measurement of learning potential.

**Exclusion:** Use of test to assess specific cognitive skills, e.g. planning, inhibition or perspective taking skills (see Code: Assess Specific Cognitive Skill).

**Code: Utility for Assessment. Assess specific cognitive skill**

**Description:** Use of test to assess children's ability in specific cognitive skills.

**Inclusion:** Assessment of memory skills, planning skills, inhibition, ability to transfer skills, perspective taking, mentalising abilities.

**Exclusion:** Assessment of general ability, e.g. potential; language functions; social-behavioural functions (see other sub-codes for Assess Function).

**Code: Utility for Assessment. Assess Language.**

**Description:** Use of test to assess child's language function.

**Inclusion:** Assessment of children's ability to follow/understand instructions; verbal reasoning, comprehension, vocabulary skills.

**Exclusion:** Use of test to obtain general cognitive profile, e.g. VIQ, PIQ discrepancy (see code

Assess Function: General Cognitive Function); and use of test to obtain qualitative information about children's responses to tasks (see code Assess Function: Qualitative Responses).

**Code: Utility for Assessment. Assess qualitative responses**

**Description:** Use of test to obtain qualitative information about a child's responses to learning tasks.

**Inclusion:** Use of test to observe the processes that the child demonstrates during testing, e.g. how the child responds to different task presentations; the quality (and not mere accuracy) of the children's responses.

**Exclusion:** Benefits arising from actual test properties, e.g. suitability of task and language demands for children with autism; flexibility in test procedures (see code Benefit).

**Code: Utility for Assessment. Assess social skills**

**Description:** Use of test to obtain information about children's social skills and behavioural difficulties associated with social dysfunction.

**Inclusion:** Use of test to assess rigidity, discipline problems in school, behaviour in social contexts, e.g. classrooms.

**Exclusion:** Use of test to obtain qualitative information about child's responses to tasks (see code : Assess qualitative responses); use of test to assess specific cognitive function, e.g. cognitive flexibility, perspective taking (see code: Assess specific cognitive function).

**Code: Other Uses**

**Description:** Benefits of the test for purposes other than planning intervention or assessment.

**Inclusion:** Usefulness of the test for diagnosis, referral to special school, referral to other professionals for follow-up.

**Exclusion:** Use of test related to intervention and assessment of children's function (see code: Utility for intervention; and Utility for assessment).

**Code: Other uses. Diagnosis**

**Description:** Usefulness of the test for use in diagnosing ASD.

**Inclusion:** Use of test in making differential diagnosis, e.g. ASD vs ADHD.

**Exclusion:** Use of test for referral to special schools and for follow-up assessments; use of test for planning intervention (see code: Utility for intervention).

**Code: Other uses. Referral**

**Description:** Use of test to inform decisions about special school placements.

**Inclusion:** Use of test in referral procedures; and in deciding schools placements, e.g. mainstream or special schools.

**Exclusion:** Use of test to plan group or class placements within a school (see code: Utility for intervention. Plan Academic Intervention)

**Code: Other uses. Legal purposes**

**Description:** Use of test for purposes that has legal ramifications in Singapore.

**Inclusion:** Use of test info. to see exemption from national service (conscript army).

**Exclusion:** Use of test for diagnosis (See code: Other uses. Diagnosis)

**Code: Other uses. Referral to other professionals**

**Description:** Use of test info. to decide if the child needs referral or follow-up assessments by other professionals/therapists.

**Inclusion:** Use of test info. to refer the child for follow-up tests, e.g. speech and language, occupational therapy assessments.

**Exclusion:** Use of test for planning intervention or for special school.

**Code: Improvements to current practice**

**Description:** Comments expressing the view that the use of the test could improve current assessment practices for children with autism.i.e. which are NOT already/currently in practice.

**Inclusion:** The ways in which the test could introduce new insights/dimensions of functioning that are ignored/overlooked in current assessment practices., e.g. use of test in improving (expanding) focus of assessment; highlighting aspects of children's functioning that is currently overlooked in assessments.

**Exclusion:** Usefulness of test information for planning intervention, assessment and features of the test that are beneficial for children with autism (Consider alternative codes: Utility for Intervention; Utility for Assessment; Benefits).

**Code: Benefit**

**Description:** Reference to benefits/advantages of using the test that arise from specific test properties.

**Inclusion:** Benefits relating to test properties, i.e. task demands (language, materials used), contents & activities, test procedures; resources (time and training); and ease in communicating results to others (parents/teachers).

**Exclusion:** Usefulness of test in planning intervention, diagnosis, referral etc. (see code: Usefulness); and improvements in current assessment procedures (see code: Practical Improvements).

**Code: Benefit. Communicating to others**

Description: Information from test that can be usefully / easily communicated to others.

Inclusion: Information from test that can be used for feedback to teachers, parents, school administrators, ease of translating test information into recommendations for teachers/parents etc.

Note: Comments must include direct reference to parents/teachers/school/third parties.

Exclusion: Use of test for planning intervention, assess functions, (see code: Utility for Intervention; Utility for Assessment); use of test for diagnosis and referral (see code: Other Uses).

**Code: Benefit. Language/culture**

Description: Properties of the test that reduces the demands on children's language skills or cultural-specific knowledge.

Inclusion: Test features such as visual cues or prompts in addition to verbal instructions; items that can be easily translated/substituted by culturally appropriate items.

Exclusion: Suitability of the activities & materials for children's across a wide age or ability levels (see code : Benefit: Activity/Material).

**Code: Benefit. Procedures**

Description: Aspects of test procedures that make it easier for the tester to accommodate children's needs/concerns during testing.

Inclusion: Flexibility in administration protocol, well designed record forms, availability of appropriate norms etc.

Exclusion: See code: Benefit. Activity/Materials; Language & Culture.

**Code: Benefit. Activities/materials**

Description: Activities/materials used in the test that facilitates its use by children of all age ranges or ability levels.

Inclusion: Use of concrete materials; appropriate item difficulty level.

Exclusion: Aspects of test administration that facilitates its use by the tester (see code: Benefit. Procedures).

**Code: Cost/Difficulties**

Description: Reference to costs/difficulties of using the test, arising from specific test properties.

Inclusion: Difficulties relating to test properties, e.g. task demands (language, materials used), contents & activities, test procedures; resources (time, cost and training required); and in communicating test results to others.

Exclusion: Direct reference to fact that test is not useful for intervention purposes (see code: Not Beneficial/Useful).

**Code: Cost/Difficulties. Communicating info to others**

Description: Difficulties in communicating test information to others.

Inclusion: Difficulties/issues abt feeding back information to teachers, parents, school administrators, difficulties in translating test info into recommendations that are relevant for parents/school.

Exclusion: Direct reference to the test's lack of usefulness for planning intervention (see code: Not Beneficial/Useful).

**Code: Cost/Difficulties. Resources: time and training**

Description: Difficulties arising from additional demands on time and training costs.

Inclusion: For example, the test takes a longer time to complete; costs needed for additional or specialised training.

Exclusion: Difficulties related to lack of flexibility/clarity in test procedures (see code : Cost/Difficulties. Procedures)

**Code: Cost/Difficulties. Activities/materials**

Description: Activities/materials used in the test that may limit its use by children of all age ranges or ability levels.

Inclusion: Lack of age appropriate norms, or ability-based norms; task demands not appropriate for some groups of children, e.g. low functioning children; possible sex bias.

Exclusion: Difficulties in performing the tasks due to language or cultural bias (see code: Cost/Difficulties. Language/Culture).

**Code: Cost/Difficulties. Language/culture**

Description: Properties of the test that places high demands on language skills or cultural-specific knowledge.

Inclusion: Lack of features such as visual cues or prompts in addition to verbal instructions; items that are not easily translated/substituted by culturally appropriate items.

Exclusion: Task demands that are not appropriate for age or ability levels (See: Cost/Difficulties. Activities/materials).

**Code: Cost/Difficulties. Procedures**

Description: Difficulties arising from aspects of test procedures.

Inclusion: Lack of flexibility in administration protocol; poorly designed record forms; difficulties arising from the need to adhere to standard administration protocol; lack of clarity/ambiguity in administration procedures.

Exclusion: Difficulties arising from high level of tester's skills and knowledge needed to interpret the test (See code: Cost/Difficulties. Time & Training).

**Code: Not beneficial/useful for intervention**

Description: Direct reference to the fact that the test is NOT useful for intervention purposes.

Inclusion: Comments expressing reservation and apprehension about its utility, relevance for intervention; and specific examples of test's lack of utility.

Exclusion: Difficulties or limitations related to properties of the test, e.g. task and language demands (see code: Cost/Difficulties. Activities/materials; Language/culture; Procedures).

## REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement for Education (NCME). (1966) *Standards for Educational and Psychological Testing*. Washington DC: American Psychological Association

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorder, Fourth Edition (DSM-IV)* Washington : American Psychiatric Publishing.

American Psychological Association (1985) *The Standards for Educational and Psychological Testing*. Washington DC : American Psychological Association

American Psychological Association (APA), American Educational Research Association (AERA) & National Council on Measurement for Education (NCME). (1999) *Standards for Educational and Psychological Testing*. Washington DC: American Psychological Association

Anastasi, A. & Urbina, S. (1996) *Psychological Testing (8th edition)*. New York: Prentice Hall.

Anastasi, A. (1988). *Psychological Testing (4th edition)*. New York: Macmillan.

Asarnow, R.F., Tanguay, P.E., Bott, L. and Freeman, B.J. (1987) Patterns of intellectual functioning in non-retarded autistic and schizophrenic children. *Journal of Child Psychology and Psychiatry*, 28, 273-280.

Barnett, D. W., & Macmann, G. M. (1992). Aptitude-achievement discrepancy scores: Accuracy in analysis misdirected. *School Psychology Review*, 21, 494-508.

Baron-Cohen S (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, 30, 285-297.

Baron-Cohen S (1992). Out of Sight or Out of Mind? Another look at deception in Autism. *Journal of Child Psychology and Psychiatry*, 7, 1141-1155.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen D.J. (Eds.) (2000) *Understanding Other Minds: Perspectives From Developmental Cognitive Neuroscience* (2nd Edition). Oxford: Oxford University Press.

Baron-Cohen, S., Leslie, A.M. and Frith, U. (1985) Does the autistic child have a 'theory of mind'? *Cognition*, 4, 37-46.

Baron-Cohen, S.; Leslie, A.M. and Frith, U. (1986) Mechanical, behaviour and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, 4, 113-125.

Baron-Cohen, S.; Tager-Flusberg, H. & Cohen, D.J (Eds) (1995) *Understanding Other Minds : Perspectives from Autism*. Oxford: Oxford Medical Publications.

Bauman, M.L., & Kemper, T.L. (1985) Histoanatomic observations of the brain in autism. *Neurology*, 35, 866-874.

Bauman, M.L., & Kemper, T.L. (Eds.) (1994) Neuroanatomic observations of the brain in autism. *The Neurobiology of Autism*. Baltimore: John Hopkins University Press.

Beadle-Brown, J., Murphy, G., Wing, L., Gould, J. Shah, A., and Holmes, N. (2002) Changes in social impairment for people with intellectual disabilities: a follow-up study of the Camberwell cohort. *Journal of Autism and Developmental Disorders*, 32(2) 195-206.

Bers, T.H (1978) The popularity and problems of focused group research. *College and University*, 64, 260-268

Bettleheim, B. (1967) *The Empty Fortress: Infantile Autism And The Birth Of The Self*. New York: Free Press.

Blaxill, M. F. (2002) Any changes in prevalence of autism must be determined. *British Medical Journal*, 324(7332) 296.

Bolton, P., MacDonald, H., Pickles, A., Rios, P. Goode, S. Crowson, M., Bailey, A. and Rutter, M. (1994) A case-control family history study of autism. *Journal of Child Psychology and Psychiatry*, 35, 877-900.

Boring, E.G. (1923) Intelligence as the tests test it. *New Republic, June*, 35-37.

Bradley, R., Danielson, L., & Hallahan, D. P. (Eds.) (2002). *Identification of Learning Disabilities: Research to Practice*. Mahwah, NJ: Lawrence Erlbaum.

Brundtland, G. : *WHO Conference on Health and Disability*. Internet address: [http://www.who.int/directorygeneral/speeches/2002/english/20020418\\_disabilitytrieste.html](http://www.who.int/directorygeneral/speeches/2002/english/20020418_disabilitytrieste.html). (Accessed 2002).

Campione, J. (1989) Assisted assessments: taxonomy of approaches and outline of strength and weaknesses. *Journal of Learning Disabilities*, 22 (3), 151- 65

Campione, J.C. & Brown, A.L. (1987) Linking dynamic assessment with school achievement. C. Lidz (Ed.) *Dynamic Assessment : an Interactional Approach to Evaluating Learning Potential*. New York : Gilford Press.

Carlson, J. & Heiz Wiedl, K. (2000) The validity of Dynamic Assessment. In C.S. Lidz & J. Elliott (Eds.) *Dynamic Assessment : Prevailing Models and Applications*. Greenwich: Jai Press.

Carpendale, J. & Chandler, M. (1996) On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development*, 67, 1686-1706.

Cattell, R.B. and Kline, P.(1977) *The Scientific Analysis of Personality and Motivation*. London: Academic Press.

Charman, T. & Cambell, A. (1997) Reliability of theory of mind task performance by individuals with a learning disability : a research note. *Journal of Child Psychology & Psychiatry*, 38, 725-730.

Cichetti, D.V. & Sparrow, S.A. (1981) Developing criteria for establishing inter-rater reliability of specific items : applications to assessment of adaptive behaviour. *American Journal of Mental Deficiency*, 86, 127-137.

Classification, Assessment, Surveys and Terminology Team, World Health Organisation (CAS, WHO) (2002) *The ICF checklist : Development and Application*. Meeting of Heads of WHO Collaboration Centers for the Family of International Clasifications. Doc ref. WHO/HFS/CAS/C/02.95 (Accessed 2002).

Cohen, D.J. Caparulo, B.K. Gold, J.R., Waldo., M.C., Shaywitz, B., & Rimland, B. (1977) Agreement in diagnosis : clinical assessment and behaviour rating scales for pervasively disturbed children. *Journal of the American Academy of Child Psychiatry*, 17, 589-603.

Coleman, M. & Gillberg, C. (1985) *The Biology of Autistic Syndromes*. New York: Prager.

Cone, J.D. (1989). Is there utility for treatment utility? *American Psychologist*, 44, 1241-1242.

Cronbach, L. J., & Furby, L. (1970) How should we measure change - or should we? *Psychological Bulletin*, 74, 68-80.

Cummins, J.C. (1984) *Bilingual and Special Education: Issues in Assessment and Pedagogy*. TX: PRO-ED.

Cunningham, G.K. (1986) *Educational and Psychological Measurement*. New York : MacMillan

De Villiers, J. (2000) Language and theory of mind: what are the developmental relationships? In S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen (Eds.) *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience* (2nd Edition). Oxford: Oxford University Press.

Deutsch, R & Reynolds, Y. (2000) The use of Dynamic Assessment by EP's in the UK. *Educational Psychology in Practice*. Vol 16, 3. 311-331

Dockrell, J. and McShane, J. (1993) *Children's Learning Difficulties: A Cognitive Approach*. Oxford : Blackwell Publishers.

Ehlers, S. & Gillberg, C. (1993) The epidemiology of Asperger Syndrome. A total population study. *Journal of Child Psychology and Psychiatry*, 31, 1327-1350.

Eisenmajer, R., Prior, M., Leekam, S., Wing, L., Ong, B., Gould, J. & Welham, M. (1998) Delayed language onset as a predictor of clinical symptoms in pervasive developmental disorders. *Journal of Autism & Developmental Disorders*, 28(6), 527-533.

Elliott, C.D. (1996) *The British Ability Scales II*. Windsor Berkshire: NFER-NELSON Pub. Company.

Elliott, C.D., Murray D.J. & Pearson, L.S. (1983) *British Ability Scales*. Slough: NFER.

Elliott, J. & Lauchlan, F. (1997) Assessing potential - the search for the philosopher's stone? *Educational and Child Psychology*, 14 (4), 6-16.

Elliott, J. (2000) Dynamic assessment in educational contexts: purpose and promise. In C.S. Lidz & J.G. Elliott (Eds) *Dynamic Assessment: Prevailing Models and Applications*, 6, 713-740.

Emslie, H., Wilson, C., Burden, V., Nimmo-Smith, I. & Wilson, B. (2003) *Behavioural Assessment of Dysexecutive Syndrome for Children (BADS-C)*, UK: Thames Valley Test Company

Farrell, P. (1997) The integration of children with severe learning difficulties: a review of recent literature. *Journal of Applied Research in Intellectual Disabilities*, 10, 1-14.

Farrell, P. (2004) Factors associated with effective inclusion of primary-aged pupils with Down Syndrome. *British Journal of Special Education*, 31(4) 184-190.

Fasotti, L. (2003). Executive function retraining. In J. Grafman & I.H. Robertson (Eds.) *Handbook of Neuropsychology*. Von Cramon, DY: Elsevier.

Ferari, M. (1982) Can differences in diagnostic criteria be stopped? *Journal of Autism and Developmental Disorders*, 12, 85-88.

Feuerstein, R. Rand, Y. & Hoffman, M.B. (1979) *The Dynamic Assessment of Retarded Performers : The Learning Potential Assessment Device, Theory, Instruments and Techniques*. Baltimore : University Park Press.

Field, A. (2000) *Discovering Statistics Using SPSS for Windows*. London : SAGE Publishers.

Fisher, N., Happé, F., & Dunn, J. (2005) The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry*, 46 (4), 409-419,

Flanagan, D.P. & Harrison, L.H. (Eds) (2005) *Contemporary Intellectual Assessment: Theories, Tests and Issues* (2nd edition). London: Guildford Press.

Flanagan, D.P., Andrews, T.J., & Genshaft, J.L. (1997) The functional utility of intelligence tests with special education populations. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds) *Contemporary intellectual assessment: theories, tests and issues*. London: Guildford Press.

Fletcher P., Happé F, Frith U, Baker S, Dolan R, Frackowiak R, Frith C (1995). Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition* 57, 109-128.

Foley, J. (1998) *English in New Cultural Contexts: Reflections from Singapore*, Singapore: Oxford University Press.

Frederickson, N. & Woolfson, H. (1987) Integration: the social dimension. *Educational Psychology in Practice*, 3, (2), 42-48.

Frederickson, N., & Cline, T. (2002). *Special Educational Needs, Inclusion and Diversity*. Buckingham: Open University Press.

Frederickson, N., Dunsmuir, S., Lang, J., & Monsen, J.J. (2004). Mainstream-special school inclusion partnerships: pupil, parent and teacher perspectives. *International Journal of Inclusive Education*, 8, (1) 37-57.

Frederickson, N., Webster, A. and Wright, A. (1991) Psychological assessment: a change of emphasis. *Educational Psychology in Practice*, 6(5); 23-32.

Freeman, B.J.; Lucas, J.C.; Forness, S.R.; & Ritvo, E.;(1985) Cognitive processing of high-functioning autistic children: Comparing the K-ABC and the WISC-R. *Journal of Psychoeducational assessment*, 4, 357-362.

Frith U (2003). *Autism: Explaining the Enigma (2nd Edition)*. Oxford : Blackwell Publishers

Frith, U. (1989). *Autism : Explaining the Enigma*. Oxford: Blackwell Publishers.

Gallery, M. E. & Hofmeister (1978) A method for assessing the treatment validity of tests in special education. *The Exceptional Child*, 25(2), 105-113.

Gardner, H. (1993) *Intelligence Reframed: Multiple Intelligences for the 21st century*. New York: Basic Books.

Gillberg, C (1996). Autism and its spectrum disorders. In G. Forrest & P. Vostanis (eds) *Advances in the Assessment and Management of Autism, Occasional Papers No.13*, 3- 6. Oxford: Association for Child Psychology and Psychiatry.

Gillberg, C., Rosenthal, U., & Johansson, E. (1983) Auditory brainstem responses in childhood psychosis. *Journal of Autism and Developmental Disorders*, 13, 181-195.

Gilliam, H.E. (1995) *Gilliam Autism Rating Scale*. Austin, TX: Pro-Ed.

Glutting, J. J., McDermott, P. A., Watkins, M. W., Kush, J. C., & Konold, T. R. (1997). The base rate problem and its consequences for interpreting children's ability profiles. *School Psychology Review*, 26, 176-188.

Goleman, D. (1995). *Emotional Intelligence*. New York: Bantam Books.

Goodwin, L.D. (2002) Changing conceptions of measurement validity : an update on the new Standards. *Journal of Nursing Education*, 41 (3), 100-106.

Gray, C. (1994) *The New Social Story Book*. TX: Future Horizons, Inc.

Gresham, F. M. (2002). Responsiveness to intervention: an alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.) *Identification of Learning Disabilities: Research to Practice* (pp. 467-519). Mahwah, NJ: Lawrence Erlbaum.

Grigorenko, E., & Sternberg, R.J. (1998) Dynamic testing. *Psychological Bulletin, 124*, 75-111.

Guilford, J.P. (1956). *Psychometric Methods*. New York: McGraw-Hill.

Gutheke, J. Beckmann, J. & Dobat, H. (1997) Dynamic testing - problems, uses, trends and evidence for validity. *Educational and Child Psychology, 14*, 17-32.

Guthke, J. & Beckmann, J. F. (2000) The Learning concept : application and practice. In C.S. Liditz & J. Elliott (Eds.) *Dynamic Assessment : Prevailing Models and Applications*, 213-266. Greenwich : Jai Press,

Hadwin, J., Baron-Cohen, S., Howlin, P. & Hill, K. (1997). Does teaching theory of mind have an effect on the ability to develop conversation in children with autism? *Journal of Autism and Developmental Disorders, 27* (5) 519-538.

Happe F. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able, mentally handicapped and normal children and adults. *Journal of Autism and Developmental Disorders, 24*, 129-154.

Happe F. (1995). The role of age and verbal ability in theory of mind task performance of subjects with autism. *Child Development, 66*, 843-855.

Happe, F. (1997) Central coherence and theory of mind : reading homographs in context. *British Journal of Developmental Psychology, 15*, 1-12.

Hassles, M.G.P (1997) Low IQ but high learning potential: why Zeinab and Mousa do not belong in special education. *Educational and Child Psychology, 14* (4), 121-135.

Hayes, N. (2000) *Doing Psychological Research : Gathering and Analysing Data*. Buckingham : Open University Press.

Haywood, H.C. & Tzuriel, D. (1992). The status and future of interactive assessment. H.C. Haywood and D. Tzuriel (Eds.). In *Interactive Assessment*, 503-505. New York: Springer-Verlag.

Haywood, H.C. & Tzuriel, D. (Eds.). (1992). *Interactive Assessment*, New York: Springer-Verlag.

Haywood, H.C., Brown, A.L., & Wingenfeld, S. (1990) Dynamic approach to psycho-educational assessment. *School Psychology Review, 19*, 411-422.

Heaton, R.K. (1993) *Wisconsin Card Sorting Test Manual : Revised*. Psychological Assessment Resources.

Hill, E. & Frith, U. (2003) Understanding autism: insights from mind and brain. London : *Philosophical Transactions. Royal Society of London, 358*, 281-289.

Hill, E. (2004) Executive dysfunction in autism. *TRENDS in Cognitive Science, 8* (1), 26-32.

Hill, E.L. (2004) Evaluating the theory of executive dysfunction in autism. *Developmental Review 24* (2004) 189–233

Hill, V. (2005) Through the past darkly: a review of the British Ability Scales Second Edition. *Child and Adolescent Mental Health, 10* (2), 87-98.

Hilliard, A. G. (1979) Standardisation and cultural bias as impediments to the scientific study and validation of "intelligence". *Journal of Research and Development in Education, 12*, 282-289.

Hodgdon, L. (2000) *Visual Strategies for Improving Communication*. USA: Quirkroberts Publishing.

Howe, M.J.A (1991) *IQ In Question. The Truth About Intelligence*. London: Sage Publications.

Hughes, C. & Graham, A. (2002) Measuring executive functions in childhood : problems & solutions. *Child and Adolescent Mental Health*, 7 (3), 131-142.

Hughes, C., & Dunn, J. (1998) Understanding mind & emotion : longitudinal associations with mental state talk between young friends. *Developmental Psychology*, 34, 1026-1037.

Hughes, C., Adlam, A., Happe, F., Jackson, J., Taylor, A. & Caspi, A. (2000) Good test re-test reliability for standard and advance false belief tasks across a wide range of abilities. *Journal of Child Psychology & Psychiatry*, 41: 483-490.

Humphreys, G.W. & Riddoch, M.J. (1987) The fractionation of visual agnosia. G.W. Humphreys & M.J. Riddoch (Eds) In *Visual Object Processing : A Cognitive Neuropsychological Approach*. Hillsdale, NJ : Lawrence Erlbaum Associates.

Hynd, G., Cohen, M.J., Ricio, C.A., & Arceneaux, J.M. (1998) A. Prifitera & D.H. Saklofske (Eds.) *WISC-III clinical use and interpretation: scientist-practitioner perspective*. San Diego: Academic Press.

Jeffrey, I. & Tzuriel (1999) *The Cognitive Modifiability Battery - Assessment & Intervention : Applications of a Dynamic Assessment Instrument in a Speech Language Therapy*. Unpublished manuscript, Vancouver, BC, Canada, The Variety Learning Center,

Jeffrey, I. (1997) The Cognitive Modifiability Battery - Assessment & Intervention : Clinical Perspective of a language therapist. Paper presented at the 6th Conference for Cognitive Education (IACE), Stellenbosch, South Africa.

Jenkins, J.R., & Pany, D. (1978) Standardized achievement tests: how useful for special education? *Exceptional Children*, 44, 448-453.

Jensen, A.R. (1980) *Bias in Mental Testing*. New York : Free Press.

Jensen, A.R. (1998) *The G Factor: the Science of Mental Ability*. Wesport: Praeger.

Kalyanpur, M. (1996) The influence of western special education on community-based services in India. *Disability & Society*, 11 (2), 249-270.

Kanner, L. (1943) Autistic disturbances of the affective contact. *Nervous Child*, 2, 217-250.

Kaufman, A.S. & Kaufman, N.L. (1983) *K-ABC Interpretive Manual*. Circle Pines : American Guidance Service.

Kaufman, A.S. (1979) *Intelligent Testing with the WISC-III*. New York: Wiley-Interscience.

Klin, A., & Volkmar, F.R. (2000). Treatment and intervention guidelines for individuals with Asperger syndrome. In A. Klin, F.R. Volkmar, & S.S. Sparrow (Eds.), *Asperger Syndrome* (pp. 340-366). New York: Guilford Press.

Klin, A., Carter, A., Volkmar, F., Cohen, D.J., Marans, W. and Sparrows, S. (1997) Developmentally based assessments – assessments issues in children with autism. In D.J. Cohen and F. Volkmar (Eds) *Handbook of Autism and Pervasive Developmental Disorder*. New York: John Wiley, 223-281.

Kline, P. (2000) *The Handbook of Psychological Testing* (2nd Edition) London : Routledge.

Klinger, L. G., & Renner, P. (2000). Performance-based measures in autism: Implications for diagnosis. *Journal of Clinical Child Psychology*, 29, 479-492.

Kruegar, R. A. (1988) *Focus Groups: A Practical Guide for Applied Research* London: Sage

Lamb, M.R., Robertson, L.C. & Knight, R.T. (1990) Component mechanisms underlying the processing of hierarchically organized patterns : inferences from patients with unilateral cortical lesions. *Journal of Experimental Psychology : Learning, memory, and Cognition*, 16, 471-483.

Landis, J. and Koch, G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Lauchlan, F. and Elliott, J. (1997) The use of the Cognitive Modifiability Battery as an intervention tool for children with complex learning difficulties. *Educational and Child Psychology*, 14(4), 137-148.

Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M. & McLennan, J. (1989) Autism Diagnostic Interview: A standardised investigator based instrument. *Journal of Autism and Developmental Disorders*, 19, 363-387.

Lee, C. (1997) *Alternatives to Cognition : A New Look at Explaining Human Social Behaviour*. New Jersey : Lawrence Erlbaum Associates.

Leekam, S.R., Libby, S.J., Wing, L., Gould, J. & Taylor, C. (2002) The Diagnostic Interview for Social and Communication Disorders: algorithms for ICD-10 childhood autism and Wing and Gould autistic spectrum disorder. *Journal of Child Psychology & Psychiatry*, 43 (3), 327-342.

Leslie, A. & Frith, U. (1988) Autistic children's understanding of seeing, knowing & believing. *British Journal of Developmental Psychology*, 6, 316-324.

Leslie, A.M. & D. Roth (1993) What autism teaches us about metarepresentation. In S. BaronCohen, H. Tager-Flusberg and D.J. Cohen (Eds.) *Understanding Other Minds*. Oxford: Oxford University Press, 83-111.

Lidz, C. (1987) *Dynamic Assessment : an Interactional Approach to Evaluating Learning Potential*. New York : Guilford Press.

Lidz, C. S. (1991) *Practitioner's Guide to Dynamic Assessment*. New York: Guilford Press.

Lidz, C.S. (2000). The Application of Cognitive Functions Scale (ACFS): An example of curriculum-based dynamic assessment. In C.S. Lidz & J.G. Elliott (Eds.). *Dynamic Assessment: Prevailing Models and Applications*. Amsterdam: JAI/Elsevier Science, 407-439.

Lidz, C. & Elliott, J. (2000) *Dynamic Assessment : Prevailing Models and Applications (Vol. 6)*. Amsterdam: JAI/Elsevier Science.

Lim H.C., Chan, T., & Yoong, T. (1994) Standardisation and Adaptation of the Denver Developmental Screening Test (DDST) and Denver II for Use in Singapore Children. *Singapore Medical Journal*, 35, 156-160.

Lincoln, A.J., Allen, M., & Kilman, A. (1995). The assessment and interpretation of intellectual abilities in people with autism. In E. Schopler, & G. Mesibov (Eds.), *Learning and Cognition in Autism* (pp. 89-117). New York, NY: Plenum.

Lincoln, A.J., Courchesne, E., Kilman, B.A., Elmasian, R. & Allen, M.H. (1988) A study of intellectual abilities in people with autism. *Journal of Autism and Developmental Disorders*, 18(4), 505-524.

Lockyer, L. & Rutter, M. (1970. A five to fifteen year follow-up study of infantile psychosis IV. *British Journal of Social and Cognitive Ability*, 9, 152-63.

Lord, C. & Schopler, E. (1988) Intellectual and development assessment of autistic children from preschool to school age: clinical implications of two follow-up studies. In E. Schopler & G. Mesibov (Eds) *Diagnosis and Assessment in Autism*. New York : Plenum Press.

Lord, C. (1982) Sex differences in autism. *Journal of Autism and Developmental Disorders*, 12, 317-330.

Lord, C. (1995) Diagnostic instruments in Autism Spectrum Disorders. In D.J.Cohen and F. Volkmar (Eds) *Handbook of Autism and Pervasive Developmental Disorder* (2nd Edition). New York: John Wiley & Sons.

Lord, C., & Pickles, A. (1996). Language level and non-verbal social-communicative behaviours in autistic and language-delayed children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 1542-1550.

Lord, C., Rutter, M. & DiLavore, P. (1989) Autism Diagnostic Observation Schedule : a standardised observation of communicative and social behaviour. *Journal of Autism and Developmental Disorders*, 19, 185-212.

Lovaas, O.I., Koegal, R.L., & Schreibman (1979) Stimulus overselectivity in autism: a review of research. *Psychological Bulletin*, 86, 1236 - 1254.

Luckett T, Powell S, Messer D, Thornton Me, & Schulz J (2002). Do children with autism who pass false belief tasks understand the mind as active interpreter? *Journal of Autism and Developmental Disorders*, 32, 127-140.

Lyen, K. (1997) A Brief History of Special Education in Singapore. Lyen, K., Lee, E., & Tham-Toh, J. (eds) *Rainbow Dreams – How to Help Your Child with Developmental Delay*. Singapore : The Rainbow Centre.

Mackintosh, N.J. (1998) *IQ and Human Intelligence*. Oxford : Oxford University Press.

Manly, T., Nimmo-Smith, I., Watson, P., Anderson, V. Turner, A. & Robertson, I. (2001) The differential assessment of children's attention : the test of everyday attention for children (TEA-Ch), normative sample and everyday performance. *Journal of Child Psychology & Psychiatry*, 42, 1065-1081.

Marshall, R., Karow, C.; Morelli, C.; Iden, K.; Dixon, J &, & Cranfill. T. (2004) Effects of interactive strategy modeling training on problem-solving by persons with traumatic brain injury. *Aphasiology*, 18 ,8, 659-673.

Matarazzo, J.D. (1990) Psychological assessment versus psychological testing. *American Psychologist*, 45, 999-1017.

Mayes, K., Klin, A., Tercyak, K.P, Cicchetti, D. & Cohen, D.H. (1996) Test retest reliability of false belief tasks. *Journal of Child Psychology & Psychiatry*, 37, 313-319.

Mayes, S.D. & Calhoun, S.L. (2003) Ability profiles in children with autism: Influence of age and IQ. *Autism*, 6, 65-80.

McCallum, R.S. & Bracken, B. (2005) The universal nonverbal intelligence test. In D.P. Flanagan & P.L. Harrison (Eds.) *Contemporary Intellectual Assessment* (2nd Edition). London: Guilford Press, 356-378.

McGrew, K.S. & Flanagan, D.P. (1993) The field of intellectual assessment : a current perspective. *The School Psychologist*, 49, 7-14.

McLeod, S. & Bleile, K. (2004) The ICF: a framework for setting goals for children with speech impairment. *Child Language Teaching and Therapy* 20,3 199-219

Messick, S. (1988) The once and future uses of validity: assessing the meaning and consequences of validity. In H. Wainer & H.I. Braun (Eds.) *Test Validity*. Hilldale: Lawrence Erlbaum, 112-156.

Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, K.G., Moreland, K.L., Dies, R.R., Eisman, E.J., Kubiszyn, T.W., & Reed, G.H. (2001) Psychological testing and psychological assessment: a review of evidence and issues. *American Psychologist*, 56 (2), 128-165.

Miles, M. B. & Hubermann A. M (1994), *Qualitative Data Analysis*. London: SAGE Publications

Ministry of Education (MOE) Press Release: *Enhancing Support for Children with Special Needs*. Internet address: <http://www.moe.gov.sg/press/2004/pr20040918.htm>. (Accessed 18 Sep 2004).

Ministry of Education, Singapore (2000a) *Developmental Continuum of Oracy, Literacy and Numeracy Skills*. Singapore : MOE.

Ministry of Education, Singapore (2000b) *Singapore Adaptive Behaviour Checklist*. Singapore : MOE.

Ministry of Education, Singapore (2000c) *Singapore Math Achievement Test*. Singapore : MOE.

Minschew, N.J., Sweeney, J.A., & Bauman, M.L. (1997) Neurological aspects of autism. In D.J.Cohen and F. Volkmar (Eds) *Handbook of Autism and Pervasive Developmental Disorder*. New York: John Wiley.

Morton, J. and Frith, U. (1995) Causal modelling: a structural approach to developmental psychopathology. Cichetti, D. and Cohen, D.J (Eds.) *Manual of Developmental Psychopathology*. New York : Wiley.

Mottron, L., Burack, J., Iarocci, G., Belleville, S. & Enn, J. (2003) Locally oriented perception with intact global processing among adolescents with high-functioning autism: evidence from multiple paradigms. *Journal of Child Psychology and Psychiatry*, 44, 904-913.

Naglieri, J.A. (1997) *Naglieri Nonverbal Ability Test*. San Antonio, TX: Psychological Corporation.

National Research Council (2001). *Educating Children With Autism. Report by the Committee on Educational Interventions for Children with Autism*. New York: Vintage.

Norris, C., & Dattilo, J. (1999). Evaluating effects of a social story intervention on a young girl with autism. *Focus on Autism and Other Developmental Disabilities*, 14, 180-186.

Nunnally, J.C. (1978). *Psychometric Theory*. New York: McGraw-Hill.

O'Brien, G., Pearson, J., Berney, T. and Bernard, L. (2000) *Measuring Behaviour in Developmental Disability: A Review of Existing Schedules*. UK: Developmental Psychiatry Research Unit, University of Newcastle

Oliver, M. & Barnes, C. (1998) *Disabled People and Social Policy : From Exclusion to Inclusion*. London : Longman

Organisation for Economic Cooperation and Development (2000) *Special Needs Education - Statistics and Indicators*. Paris : OECD.

Ortiz, S.O. & Dynda, A.M. (2005) Use of intelligence tests with culturally and linguistically diverse populations. D.P. Flanagan & P.L. Harrison (Eds.) *Contemporary Intellectual Assessment (2nd edition)*. London: Guilford Press.

Oswald D & Ollendick T (1989). Role taking and social competence in autism and mental retardation. *Journal of Autism and Developmental Disorders*, 19, 119-127.

Ozonoff, S., & Miller, JN (1995). Teaching theory of mind: A new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders*, 25, 415-433.

Ozonoff, S., (1995) Executive function in autism. In E.Schopler & B.B. Mesibov (Eds.) *Learning and Cognition in Autism*. New York: Plenum Press, 143-176.

Ozonoff, S., Pennington, B.F., & Rogers, S.J. (1991) Executive function deficits in high-functioning autistic individuals: relation to theory of mind. *Journal of Child Psychology and Psychiatry*, 32, 1081-1105.

Pennington, B. & Ozonoff, S. (1996) Executive functions and developmental psychopathology. *Journal of Child Psychology & Psychiatry*, 37, 51-87.

Perner J, Leekam S, & Wimmer H (1987). Three year olds' difficulty with false belief: The case for conceptual deficit. *British Journal of Developmental Psychology*, 5: 125-137.

Pfeiffer, D. (1998) The ICIDH and the need for its revision. *Disability & Society*, 13 (4), 503-523.

Premack, D. & Woodruff, G. (1978) Does the chimpanzee have a 'theory of mind'? *Behaviour and Brain Sciences*, 4, 515-526.

Prior, & Hoffmann, (1990) Neuropsychological testing of autistic children through exploration with frontal lobe tests. *Journal of Autism and Developmental Disorder*, 20, 581-590.

Quay, M. L. (1993) Special Education in Singapore. In Quah, M. L., Gopinathan, S. and Chang A.S. (Eds) *Towards Education for All in Singapore : a Country Report to the South-East Asian Research and Review Advisory Group (SEARRAG)*. Singapore : National Institute of Education.

Raven, J. (1956). *Coloured Progressive Matrices*. New York: Psychological Corporation.

Reschly, D. J., & Gresham, F. M. (1989). Current neuropsychological diagnosis of learning problems: A leap of faith. *Handbook of Clinical Neuropsychology* (pp. 503-519). New York: Plenum.

Reschly, D.J. (1997) Diagnostic and treatment validity of intelligence. D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds) *Contemporary Intellectual Assessment: Theories, Tests and Issues*. London: Guildford Press.

Reschly, D.J. (1997) Diagnostic and treatment validity of intelligence. D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds) *Contemporary Intellectual Assessment: Theories, Tests and Issues*. London: Guildford Press.

Resing, W. (2000) Assessing the learning potential for inductive reasoning in young children. In C.S. Liditz & J. Elliott (Eds.) *Dynamic Assessment : Prevailing Models and Applications*. Greenwich : Jai Press.

Richardson, K. (1996) *Understanding Intelligence*. Milton Keynes: Open University Press.

Ricks, D.M. & Wing, L. (1975) Language, communication and use of symbols in normal and autistic children. *Journal of Autism and Childhood Schizophrenia*, 5, 191-221.

Rothman, H. & Semmel, M. (1990) Dynamic assessment : a comprehensive review of literature. In Gaylord-Ross (Ed) *Issues and Research in Special Education*. New York : Teachers College Press.

Rubia, K., Taylor, E., Smith, A. Oksanen, H. Overmayer, S. & Newman, S (2001) Neurological analyses of impulsiveness in childhood hyperactivity. *British Journal of Psychiatry*, 179, 138-143.

Rushton, J.P., Brainerd, C.J. & Pressley, M. (1983) Behavioural development and construct validity : the principle of aggregation. *Psychological Bulletin*, 94, 18-38.

Russell, J. & Hill, E. L. 2001 Action-monitoring and intention reporting in children with autism. *Journal of Child Psychology and Psychiatry* 42, 317-328.

Rust, J., Golombok, S. and Trickey, G. (1993) *Manual of the Wechsler Objective Reading Dimensions (WORD)*. London: The Psychological Corporation

Sansosti, F.J., Powell-Smith, K.A. & Kincaid, D. (2004) A research synthesis of social stories intervention for children with ASD. *Focus on Autism and Other Developmental Disorders*, 19 (4), 199-204.

Sattler, J.M. (1992) *Assessment of Children* (3<sup>rd</sup> edition). San Diego, CA: Author.

Schopler, E. & Mesibov, G (Eds) (1985) *Communication Problems in Autism*. New York: Plenum Press.

Schopler, E. & Mesibov, G.B. (Eds) (1987) Neurobiological issues in autism. In *Current Issues in Autism* (series). New York: Plenum Press.

Schopler, E. & Mesibov, G.B. (Eds) (1988) Diagnosis and assessment of autism. In *Current Issues in Autism* (series). New York: Plenum Press.

Schopler, E. & Mesibov, G.B. (Eds) (1995) Learning and cognition in autism. In *Current Issues in Autism* (series). New York: Plenum Press.

Schopler, E. & Reichler, R. (1972) How well do parents understand their own psychotic child? *Journal of Autism and Child Schizophrenia*, 2, 387-400.

Schopler, E. (1979) Do autistic children come from upper-middle-class parents? *Journal of Autism and Developmental Disorders*, 9, 139-152.

Schopler, E. , Reichler, R., DeVellis, R., & Daly, K. (1988) *The Childhood Autism Rating Scale (CARS)*. Los Angeles : Western Psychological Services.

Schopler, E. (1971) Parents of psychotic children as scapegoats. *Journal of Contemporary Psychotherapy*, 4, 17-22.

Shah, A., & Frith, U. (1983) An islet of ability in autistic children: a research note. *Journal of Child Psychology & Psychiatry*, 24, 613-620.

Shallice, T. & Burgess, P. (1991) Deficits in strategy application following frontal lobe damage in man. *Brain*, 114, 727-741.

Shallice, T. (1982) Specific impairments in planning. *Philosophical Transactions of the Royal Society of London*, 298, 199-209.

Siegel, B. (2003) *Helping Children with Autism Learn*. Oxford : Oxford University Press.

Siegel, B., Anders T. F., Ciaranello, R., Bienenstock, B. & Kraemer, H.C. (1986) Empirically derived subsclassification of autistic syndrome. *Journal of Autism and Developmental Disorders*, 16(3), 275-293.

Siegel, L.S. (1992) IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities*, 22, 618-629.

Simmeonsson, R., Leonard, M., Lollars, D., Bjorck-Akesson, E., Hollenweger, J. and Martinuzzi, A. (2003) Applying the International Classification of Functioning, Disability and Health (ICF) to measure childhood disability. *Disability and Rehabilitation*, 25 (11-12), 602-610.

Singapore Statistics Department (2000) *Census of Population 2000*. Singapore: Singapore Statistic Department.

Skoff, B.F., Misrsky, A.F., & Turner, D. (1980) Prolonged brainstem transmission time in autism. *Psychiatric Research*, 2, 157-166.

Snowling, M. & Frith, U. (1986) Comprehension in 'Hyperlexic' readers. *Journal of Experimental Child Psychology*, 42, 392-415.

Sodian, B. & Frith, U. (1992) Deception and Sabotage in autistic, retarded, and normal children. *Journal of Child Psychology & Psychiatry*, 33, 591-605.

South, M., Williams, J., McMahon, W., Owley, T., Filipek, P., Shernoff, E., Corsello, C., Lainhart, J., Landa, R. & Ozonoff, S. (2002) Utility of Gilliam Autism Rating Scale in Research and Clinical Populations. *Journal of Autism and Developmental Disorders*, 32 (6); 593-599.

Spearman, C. (1904) 'General Intelligence', objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Sternberg, R.J. & Detterman, D.K. (1986) *What is Intelligence? Contemporary Viewpoints on Its Nature & Definitions*. New Jersey: Ablex.

Sternberg, R.J., & Grigorenko, E.L. (2002) *Dynamic Testing*. New York: Cambridge University Press.

Sternberg, R.J. (Ed) (2000) *Handbook of Intelligence*. Cambridge : Cambridge University Press

Stroop, J.R. (1935) Studies on interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.

Stucki, G., Ewert, T., and Cieza, A. (2002) Value and application of ICF in rehabilitative medicine. *Disability and Rehabilitation*, 24 (17), 932-938.

Stucki, G., Cieza, A., Ewert, T., Kostanjsek, N. Chatterji, S. and Ustun, T., (2003) Application of international Classification of Functioning, Disability and Health (ICF) in clinical practice. *Disability and Rehabilitation*, 24 (5), 281-282.

Sullivan K, Zaitchik D & Tager-Flusberg H (1994). Preschoolers can attribute second order beliefs. *Developmental Psychology*, 30, 395-402.

Swaggart, B.L., Gagnon, E., Bock, S.J., Earles, T.L., Quinn, C., Myles, B.S., & Simpson, R.L. (1995). Using social stories to teach social and behavioural skills to children with autism. *Focus on Autistic Behaviour*, 10, 1-16.

Swanson, H.L. (1996) *Swanson-Cognitive Processing Test*. Austin, TX : Pro-Ed.

Swanson, H.L. (2000) Swanson Cognitive processing test – review and application. C.S. Lidz & J. Elliott (Eds.) *Dynamic Assessment : Prevailing Models and Applications*. Greenwich : Jai Press.

Tabachnick, B.G. & Fidell, L.S. (2001) *Using Multivariate Statistics (4th edition)*. London: Allyn & Bacon.

Tager-Flusberg, H. (1981) Sentence comprehension in autistic children. *Applied Psycholinguistics*, 2, 5-24.

Tager-Flusberg, H. (1991) Semantic processing in the free recall of autistic children : further evidence of cognitive deficit. *British Journal of Developmental Psychology*, 9, 417-430.

Thorndike, R.M. (1997) The early history of intelligence testing. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.) *Contemporary Intellectual Assessments*. New York: Guildford Press.

Thurstone, L.L. (1938) The absolute zero in intelligence measurement. *Psychological Review*, 35, 175-179.

Thurstone, L.L. (1938) The factorial isolation of seven primary abilities. *Psychological Bulletin*, 33, 780-781.

Tomlinson, S. (1982) *A Sociology of Special Education*. London : Routledge & Kegan Paul.

Tzuriel, D. & Caspi, N. (1992). Dynamic assessment of cognitive modifiability in deaf and hearing preschool children. *Journal of Special Education*, 26, 235-251

Tzuriel, D. & Feuerstein, R. (1992). Dynamic group assessment for prescriptive teaching: differential effects of treatment. In H.C. Haywood and D. Tzuriel (Eds.). *Interactive Assessment*, 503-505. New York: Springer-Verlag.

Tzuriel, D. & Klein, P.S. (1985). Analogical thinking modifiability in disadvantaged, regular, special education and mentally retarded children. *Journal of Abnormal Child Psychology*, 13, 539-552.

Tzuriel, D. & Klein, P.S. (1985). Analogical thinking modifiability in disadvantaged, regular, special education and mentally retarded children. *Journal of Abnormal Child Psychology*, 13, 539-552.

Tzuriel, D. (1995) *The Cognitive Modifiability Battery (CMB) Assessment & Intervention. Instruction Manual*. Bar-Ilan University, School of Education : Israel.

Tzuriel, D. (1996) Mother-child mediated learning strategies in free-play and structured situations among low, medium and high SES level. *Child Development and Care*, 126, 57-82.

Tzuriel, D. (2000) The Cognitive Modifiability Battery (CMB) - assessment & intervention : development of a dynamic assessment instrument. C.S. Liditz & J. Elliott (Eds.) *Dynamic Assessment : Prevailing Models and Applications*. Greenwich : Jai Press.

Tzuriel, D. (2001) *Dynamic Assessment of Young Children*. New York : Kluwer Academic/Plenum Publishers.

Tzuriel, D., Samuels, M, & Feuerstein, R. (1988). Non-intellective factors in dynamic assessment. In R.M. Gupta & P. Coxhead (Eds.), *Cultural Diversity and Learning Efficiency*. London: Macmillan, 334-367.

Ueda, S., & Okawa, Y. (2003) The subjective dimension of functioning and disability: what is it and what it is for? *Disability and Rehabilitation*, 25(11-12), 596-601.

UNESCO (2002) *Education for all : Status and Trends 2002. Report for the International Consultative Forum on Education for All*. Paris : UNESCO.

Ungerer, J.A. & Sigman, M. (1987) Categorisation of skills and receptive language development in autistic children. *Journal of Autism and Developmental Disorders*, 17, 3-16.

Ustun, T., Chatterji, S., Bickenbach, J., Kostanjsek, N. and Schneider, M. (2002) The International Classification of Functioning, Disability and Health: a new tool for understanding disability and health. *Disability and Rehabilitation*, 25 (11-12), 565-571.

Ustun, T., Chatterji, S., Kostanjsek, N. and Bickenbach, J. (2003) WHO's ICF and functional status information in health records. *Health Care Financing Review*, 24 (3), 77-88.

Valdes, G. & Figueroa, R.A. (1994) *Bilingualism and Testing: a Special Case of Bias*. New Jersey: Ablex.

Vaughn, S., Schumm, J.S. & Sinagub, J. (1996) *Focus Group Interviews in Education and Psychology*. London: Sage Publications.

Wahlberg, T., Obiakor, F., Burkhardt, S. & Rotatori, A (Eds) (2001) *Autism Spectrum Disorders : Educational and Clinical Intervention*. Oxford : Elsevier Science Ltd.

Ward, J., Bochner, S. Center, Y., Outred, L. & Pieterse, M. (Eds.) (1987) *Educating Children with Special Needs in Regular Classrooms: An Australian Perspective*. North Ryde NSW: Special Education Centre, MacQuarie University.

Wasserman, J.D. & Tulsky, D.S. (2005) A history of intelligence assessment. D.P. Falanagan & P.L. Harrison (Eds.) *Contemporary Intellectual Assessment: Test, Theories & Issues* (2nd Edition). London: Guilford Press.

Wechsler, D. (1939) *The Measurement of Adult Intelligence*. Baltimore: Willimas & Wilkins.

Wechsler, D. (1991) *Manual for The Wechsler Intelligence Scales for Children - Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1993) *Manual for The Wechsler Intelligence Scales for Children - Third Edition (UK)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1996) *Manual for The Wechsler Intelligence Scales for Children - Third Edition (Singapore Version)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2004) *Manual for The Wechsler Intelligence Scales for Children - Fourth Edition - Intergrated*. San Antonio, TX: The Psychological Corporation.

White, S. & Frith, U. (in prep). Unpublished PhD thesis in preparation. *Subtypes in the Autism Spectrum: from Cognition to Behaviour*. University College London.

Wilson, B.A., Alderman, N., Burgess, P., Emslie, H., & Evans, J. (1996) *Behavioural Assessment of the Dysexecutive Syndrome*. Thames Valley Test Company : UK.

Wimmer H, & Perner J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13: 103-128.

Wing, L. & Gould, J. (1979) Severe Impairments of social interaction and associated abnormalities in children: epidemiology and classification. *Journal of Autism and Developmental Disorders*, 9, 11-29.

Wing, L. (1988) The continuum of autistic characteristics. E.Schopler & S. Mesibov (Eds.) *Diagnosis and Assessment in Autism*. New York: Plenum, 153-176.

Wing, L. (1997) Syndromes of autism and atypical development. In D.J.Cohen & F.Volkmar (Eds) *Handbook of Autism and Pervasive Developmental Disorder*. New York: John Wiley.

Wing, L. (1999) *Diagnostic Interview for Social and Communication Disorder (DISCO)*. UK: National Autistic Society.

Wing, L., Leekam, S., Libby, S., Gould, J., and Larcombe, M. (2002). The Diagnostic Interview for Social and Communication Disorders: background, inter-rater reliability and clinical use. *Journal of Child Psychology and Psychiatry*, 43, 307-325.

Witkins, H., Oltman, P. Raskin, E. & Karp, S. (1971) *A Manual for Embedded Figures Test, Children's Embedded Figures Test & Group Embedded Figures Test*. California : Consulting Psychologists Press.

World Health Organisation (1995) *International Classification of Disability - Version 10 (ICD-10)* Geneva : WHO

World Health Organisation (2000) *WHO Presentation on ICIDH-2 Beta-2 Field Trial Results* (power-point presentation). Geneva : WHO

World Health Organisation (2001) *International Classification of Functioning, Disability and Health (ICF)*. Geneva : WHO

Zetlin, A.G. (1987). The social status of mildly learning handicapped high school students. *Psychology in the Schools*, 24, 165-173.

Zoppi, K.A. & Epstein, R.M. (2001) *Interviewing in Medical Settings*. New York : Springer-Verlag.