Assessing the cross-cultural validity of the Wechsler Intelligence Scale for Children – 4th edition for use in Trinidad and Tobago

Korinne Gillian Louison

Supervised by:

Yulia Kovas

Alice Jones

This thesis is submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Goldsmiths, University of London

2016

DECLARATION

I hereby declare that thesis is my own work. None of the work detailed in the text of this document has been previously submitted as a requirement for a degree or an award. Work that have been conducted by sources have been appropriately acknowledged.

Signed ___________________________________

Date ___________________________________

ABSTRACT

More than ten years after its publication, there has been no prior attempt to investigate the validity of the WISC-IV (Wechsler, 2003a) for use in Trinidad and Tobago (T&T). This thesis is the first to assess the fit of a cross-cultural interpretive model of WISC-IV(US) measured intelligence in T&T children. The primary objectives were to: ascertain the psychometric properties of the WISC-IV(US); determine how the WISC-IV(US) subtests are associated with specified antecedent environmental variables; examine the relationship between WISC-IV(US) global ability and academic achievement; assess the fit of alternative interpretive models; and determine the predictive validity of adjusted IQ scores. Examination of the correlation matrix corroborated five alternative measurement models, with evidence of best fit for a direct hierarchical framework (Watkins et al., 2006). Multiple regression analyses demonstrated significant positive relationships between parental education and verbal comprehension, and between school performance and verbal comprehension, perceptual reasoning and global ability. Additionally, environmental deprivation was found to be negatively correlated with performance on all WISC-IV (US) composites. Children with higher global intelligence scores performed better than their low-scoring counterparts on two national tests of academic achievement. School performance was also shown to predict academic achievement in the sample. The results of the structural equation modelling analyses provided support for three distinct measurement models featuring the Wechsler indirect hierarchical model, the Watkins et al. direct hierarchical model, and an author-defined cross cultural direct hierarchical model. The antecedent variables of parental education, school performance and environmental deprivation and the outcome variable or academic achievement added significantly to the model. Adjusted factors scores that were derived from each path model accounted for a significant portion of variance in national test performance. If replicated, the current findings offer potentially useful alternative frameworks for interpreting test performance in T&T children.

STATEMENT OF AUTHORSHIP

The first phase of this thesis, the pilot study, involved collecting archival data from the records of the Student Support Services Division (SSSD) of the Ministry of Education (MOE). I executed all aspects of data collection, as well as data input, screening, and data analysis. For the additional studies that were described in Chapters 4 to 7, I prepared the consent form and the demographic questionnaire. I sought permission for the studies through personal visits and written correspondence to the MOE and through meetings with school principals. I administered the Wechsler Intelligence Scale for Children 4th Edition (WISC-IV) to over 1/3 of the sample. The rest were administered by graduates of University of the West Indies, Masters in Clinical Psychology programme. I also trained the testers prior to the start of testing. Secondary Entrance Assessment test scores and WISC-IV (US) test protocols were collected by me and I did the data input, screening and analyses for the 2nd dataset. I therefore consider the work presented in this thesis as my own.

ACKNOWLEDGEMENTS

I would like extend my sincerest gratitude and appreciation to my supervisor Dr. Yulia Kovas, who has been an invaluable mentor, advisor and teacher. Dr. Kovas, I am honoured to have benefitted from your expert advice, as well as your support and encouragement. I owe my development in the area of research to you and for that I will always be grateful. I would like to thank the members of InLab, who have provided kind support and helpful feedback for my thesis.

Thank you to the Ministry of Education of Trinidad and Tobago for giving me permission to do this important study. Also I would like to extend my appreciation to DERE, SSSD, school principals and the Catholic board for assisting me in data collection and in my recruitment of participants. Thanks to the participants and their families for taking part in my project. Also, I would like to thank my research assistants for volunteering their time to assist with data collection. A special word of gratitude is offered to Ms. Leticia Rodriguez-Cupid, a dear friend and colleague, without whom this thesis would not have been possible. Thank you so much for your support and well wishes, and for making this very challenging road much easier to negotiate.

A special thanks to members of my extended family and to my friends for your supportive and motivating words. To Katharina Hoelter, I am so grateful to you for taking care of me during my write up. Offering me a place of calm and making sure that I was comfortable, well fed, and in the right frame of mind to take on what seemed like an insurmountable task at the time.

To Joerg Hoelter, you are the definition of true friendship and selflessness. Thank you for providing a listening ear, a shoulder to cry on, and a buttress against the stumbling blocks of self-doubt and uncertainty that I have faced over the years. Your

support cannot really be repaid. This would not have been possible without you and for that I will be forever grateful.

To my loving family: my mother, Glenda Louison, my father, the late Kelton Louison, my sister Karen, and my brothers Gary and Greg, thank you for your love, understanding and support. This could not have been started or finished without all of you. Words will not do justice to how thankful and appreciative I am of the support and sacrifices you have made to ensure that I achieve my goals.

In loving memory of my father, Kelton Kenny Louison (1938–2010).

TABLE OF CONTENTS

Title..1
DECLARATION 2
ABSTRACT..3
STATEMENT OF AUTHORSHIP 5
ACKNOWLEDGEMENTS..6
TABLE OF CONTENTS 8
LIST OF TABLES...14
LIST OF FIGURES 18
LIST OF ABBREVIATIONS...20

Chapter 1: Intelligence: Definition, Internal Structure, Measurement and Aetiology of Variability..23
  The Internal Structure of Intelligence 27
    Specific abilities and the concept of *g* 27
    The hierarchical structure of intelligence 30
Intelligence Definition still "Under Construction": Measurement Issues .33
  Origins of intelligence testing: predicting occupational and academic success 34
  Modern intelligence testing: estimating abilities 38
Controversies in Intelligence Testing 44
  Inadequacy of intelligence models...44
  Ability of Achievement?...45
  Intelligence or a covariate of intelligence? Implications of the Flynn Effect. 46
  Culture, environment and the contextual nature of intelligence. 47
  Cultural bias in intelligence measurement. 48
Aetiology of Variability in Intelligence 50
  Genetics and environment..50
  DNA 54
  Home environment...55
  Parental education and income 56
  Schooling 56

The Outcomes of Intelligence..57
Conclusion 60

Chapter 2: The Wechsler Intelligence Scale for Children – 4th Edition WISC-IV(US) 62
    The WISC-IV (US) core subtests 63
    Scoring guidelines for the WISC –IV (US) subtests and composites...68
Psychometric Properties of the WISC-IV (US)...69
    Reliability..69
    Internal Validity..71
        Subtest inter-correlations..71
        Factor structure 71
    External validity..75
The Cross Cultural Validity of the WISC-IV (US) 77
Is structural invariance sufficient evidence of validity? 81
The WISC-IV (US) in Trinidad and Tobago (T&T) –Building a Rationale for a Cross-Interpretive Model...83
Possible Sources of Content Bias – Similarities and Picture Concepts.....84
Implications of Test Bias in Clinical Practice..85
Investigating the Relationship between Environmental Variables and WISC-IV (US) Structural Validity in a Referred Sample of T&T children -A Pilot Study...87
The Validity of the Current 4-factor Structure in a sample of Referred T&T children 93
    Results..95
Investigating Structural Invariance of the WISC-IV (US) based on SES .97
    Multi-sample analysis based on referral category. 98
    Results..99
    Testing for invariant factor loadings related to VCI, PRI, WMI, PSI. 101
Implications of the Pilot Study 102

Chapter 3: Methods...109

Participants......109
    Recruitment......109
    Description of the Sample......110
Materials......112
    Demographic Questionnaire......112
    Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) (US)......113
    The Secondary Entrance Assessment (SEA)......114
    Academic Performance Index (API)......115
Procedure......116
Training......116
    Interrater agreement......117
    Data Protection and Ethical Considerations......117
Data Analysis and Chapter Summary......118
    Chapter 4: Descriptive statistics and comparisons of means......119
    Chapter 5: Assessing the validity of the current four-factor model......119
        Inter-correlation analysis......119
        Exploratory factor analysis (EFA)......119
        Confirmatory factor analysis (CFA).......120
    Chapter 6: The relationship between environmental factors and performance on the WISC-IV (US)......121
    The relationship between IQ and environmental variables and performance on the SEA.......121
    Chapter 7: Investigating the fit of the cross-cultural interpretive model......122
Summary......123

Chapter 4: Examining the Performance of the T&T sample......125
Data Screening......125
Study 1......129
Results......129
    Between Group Comparisons of WISC-IV (US) performance (T&T vs US children)......129

  Within-Group Comparisons of Means in the T&T Sample..........131
  Subtest mean score comparisons 131
  Index mean score comparisons 133
Discussion...134
  Explaining Differences in Mean IQ Scores 134
Conclusion 142

Chapter 5: Internal Validity of the WISC-IV(US) in the T&T sample

Chapter 5: Internal Validity of the WISC-IV(US) in the T&T sample...144
Study 2: An Examination of the Inter-correlations among Core Subtests of the WISC-IV (US)..144
Analyses...147
Results...148
Discussion...151
  Subtest Inter-Correlations 151
  Composite Inter-Correlations...153
Study 3: Exploratory Factor Analysis of the WISC-IV (US) Core Subtests in a T&T Sample...154
Analyses...157
Results...158
Discussion...164
  Picture Concepts, Working Memory and the Verbal Comprehension Subtests –An Education/Exposure Related Cluster? 165
Study 4: Confirmatory Factor Analysis of the WISC-IV (US) Core Subtests in a T&T Sample 168
Analyses...172
Results...174
Discussion...176
Conclusion 179

Chapter 6: The external validity of the WISC-IV (US)

Chapter 6: The external validity of the WISC-IV (US) 180
Study 5: The Relationship Between Environmental Variables and WISC-IV (US) Performance..180
Analyses...182

Results.....182
    Inter-correlations.....182
    Multiple Regressions.....184
    Discussion.....188
Study 6: School Performance and Global Ability as Predictors of Academic Performance.....194
Academic success in a developing nation: The Trinidad and Tobago education system.....199
    Differentiation among primary schools.....203
    Differentiation among secondary schools.....204
    The Secondary Entrance Assessment.....205
Analyses.....209
Results.....209
    SEA Language Arts achievement.....211
    Mathematics achievement.....213
Discussion.....214
Conclusion.....216

Chapter 7: Investigating the cross cultural interpretive model.....219
Study 7.....220
Adjusting IQ scores based on Socio Cultural variables – The SOMPA.....224
Analyses.....229
    Measurement model.....229
    Structural model.....229
    Structural Equation Modelling.....230
Results.....235
    Applying Model Constraints.....235
    Model 1.....236
    Results.....237
    Model 2.....240
    Results.....241
    Model 3.....245
    Results.....246
Discussion.....249

        Implication of the Findings..251
Study 8: Using Factor Score Estimates to Assess the Predictive Validity
of the Cross Cultural Interpretive Model...253
Analyses...254
Results..254
Discussion:
Revisiting Factor Score Estimations in Test Score Interpretations 255

Chapter 8: Conclusion..258
Investigating the Performance of the T&T Sample 260
Investigating Internal Validity 262
Environmental Variables and WISC-IV (US) Performance....................264
Environmental Variables, Global Ability and Academic Performance...265
Assessing the Cross-Cultural Interpretive Model 266
        Factor score estimation 267
Implications of the Findings 268
        Enhancing WISC-IV (US) score interpretations..........................268
        Watkins (2006) direct hierarchical measurement model 268
        Education/exposure measurement model 269
        Cross-cultural interpretive model 270
                Factor score estimation and cross-cultural
                means comparisons..270
                Factor score estimation and adjusted IQ scores 272
Strengths and Limitations of the Thesis..272
Future Directions 274
Final Remarks 276
References..279
APPENDIX A 305
APPENDIX B 315
APPENDIX C 328

LIST OF TABLES

Table 2.1. Split-Half and Test-Retest Reliability Coefficients (corrected r) for WISC-IV (US) Subtests and Composites........................70
Table 2.2. Correlational Matrix for the WISC-IV (US) Standardization Study 72
Table 2.3. Standardized Loadings of Core Subtests on WISC-IV Composites (All Ages)..73
Table 2.4. Correlations between WISC-IV (US) Composite Scores and Criterion Measures..76
Table 2.5. WISC-IV factor structure by country 80
Table 2.6. Demographic characteristics of a referred sample of T&T children 88
Table 2.7. Comparison of WISC-IV (US) IQ scores: Trinidad & Tobago vs. United States children with LD.......................................90
Table 2.8. Comparison of WISC-IV (US) IQ scores: Trinidad & Tobago vs. United States children with Mild Mental Retardation.....90
Table 2.9. Comparison of WISC-IV (US) IQ scores: Trinidad & Tobago vs.UnitedStates children with Moderate Mental Retardation90
Table 2.10. Comparison of WISC-IV (US) IQ scores: Trinidad & Tobago vs. United States children with Attention Deficit Hyperactivity Disorder...92
Table 2.11. Goodness of Fit Indexes for Evaluating Model Adequacy (N=354) 96
Table 2.12. Key demographic data by referral category for the T&T sample of referred children 98
Table 2.13 IQ score (WISC-IV (US)) by referral category 99
Table 2.14. Goodness of Fit Indexes for Evaluating Model Adequacy..100
Table 2.15. Goodness of fit indices for Unconstrained and Constrained Models 100
Table 2.16. Chi Square difference tests for comparing nested models...101
Table 2.17. Chi Square difference values for comparing Factor and subtest loadings..102
Table 3.1. Key demographic data for the T&T sample of 11 to

12-year-old children N = 203...111
Table 3.2. Comparisons of sample vs. population data by Gender Ethnicity and Parental Education.......................................112
Table 4.1. WISC-IV (US) IQ scores and skewness and kurtosis values for Trinidad & Tobago (N = 203) sample (11-12) .126
Table 4.2. WISC-IV (US) scores: Trinidad & Tobago sample (N = 203) and United States standardization sample (N = 200) 11 years 130
Table 4.3. WISC-IV (US) scores: Trinidad & Tobago sample (N = 203) and United States standardization sample (N = 200) 12 years 130
Table 4.4. Comparisons of Subtest Means using Repeated Measures ANOVA – Trinidad & Tobago sample (N = 203)............131
Table 4.5. Statistically Significant Pairwise Comparisons of 10 core subtests – Trinidad & Tobago sample (N = 203)..............132
Table 4.6. Comparisons of Index Means using Repeated Measures ANOVA – Trinidad & Tobago sample (N = 203)............133
Table 4.7. Statistically Significant Pairwise Comparisons of 4 composites – Trinidad & Tobago sample (N = 203) 134
Table 4.8. Sample Means comparisons of T&T sample with learning disabilities and US Learning Disability sample, and T&T sample without learning disabilities with US standardization sample..137
Table 4.9 Comparisons of Subtest Means using Repeated Measures ANOVA – Trinidad & Tobago sample without learning disability cases (N = 152).......................138
Table 4.10 Statistically Significant Pairwise Comparisons of Coding with 9 subtests – Trinidad & Tobago sample (N = 203) 138
Table 4.11. Comparisons of Index Means using Repeated Measures ANOVA – Trinidad & Tobago US Exposure subsamples...140
Table 4.12 Statistically Significant Pairwise Comparisons with Picture Concepts: Trinidad & Tobago US Exposure subsamples...141
Table 5.1. Correlational Matrices for the T&T and US Standardization

samples..149
Table 5.2. Correlational Matrices of the WISC-IV Composites for the T&T and US Standardization Combined (N = 203)..........150
Table 5.3. Goodness of Fit Indexes for Evaluating Model Adequacy for first order EFAs (N = 203)..159
Table 5.4. Factor Loading and Communalities....................................159
Table 5.5. Factor Correlations: Three Factor Solution 160
Table 5.6. Factor Loadings and Communalities – 4 factor solution....161
Table 5.7. Factor Correlations: Four Factor Solution..........................161
Table 5.8. Factor Loading and Communalities – 4 factor for the T&T sample 163
Table 5.9. Goodness of Fit Statistics for Confirmatory Factor Analysis of 10 core subtests – Wechsler (2003b) study 171
Table 5.10 Goodness of Fit Indexes for Evaluating Model Adequacy.174
Table 5.11. Standardized Loadings, Variances, Total and Common Variance Explained, Communalities, and Uniqueness values for the CFA of the direct hierarchical model......................175
Table 6.1. Intercorrelations for Environmental Variables and WISC-IV global and specific factors 183
Table 6.2. Standardized β coefficients for predictors of Verbal ComprehensionIndex(VCI) score after multiple regression185
Table 6.3. Standardized β coefficients for predictors of Perceptual Reasoning Index(PRI) score after multiple regression.......186
Table 6.4. Standardized β coefficients for predictors of Working Memory Index (WMI) score after multiple regression.......186
Table 6.5. Standardized β coefficients for predictors of Processing Speed Index (PSI) score after multiple regression 187
Table 6.6. Standardized β coefficients for predictors of Full Scale IQ (FSIQ) score after multiple regression 187
Table 6.7. Comparisons of No Exposure and Exposure group data by Type of School, Age and Parental Education 189
Table 6.8. Intercorrelations for Environmental variable IQ and Achievement 210
Table 6.9. Logistic Regression Analysis of the relationship between

School Quality and Global Ability and Language Arts Achievement controlling for Parental Education and Exposure (Reference category is 0-50 marks) 212
Table 6.10. Logistic Regression Analysis of the relationship between School Quality and Global Ability and Language Arts Achievement controlling for Parental Education and Exposure (Reference category is 51-85 marks) 212
Table 6.11. Logistic Regression Analysis of the relationship between School Quality and Global Ability and Mathematics Achievement controlling for Parental Education and Exposure (Reference category is 0-50 marks) 213
Table 6.12. Logistic Regression Analysis of the relationship between School Quality and Global Ability and Mathematics Achievement controlling for Parental Education and Exposure (Reference category is 51 - 85 marks) 214
Table 7.1. Goodness of Fit Indexes for Evaluating Model Adequacy Of Path Models (N = 203) 235
Table 7.2. Goodness of Fit Indices for Full and Constrained Models 237
Table 7.3. Unstandardized Coefficients, Standardized Coefficients and Significance Levels for Unconstrained Model 1..........239
Table 7.4. Goodness of Fit Indices for Full Model 2 and Constrained Models...241
Table 7.5. Unstandardized Coefficients, Standardized Coefficients and Significance Levels for Unconstrained Model 2..........243
Table 7.6. Goodness of Fit Indices for Full Model 3 and Constrained Models...246
Table 7.7. Unstandardized Coefficients, Standardized Coefficients and Significance Levels for Unconstrained Model 3..........248
Table 7.8. Goodness of Fit Indices for Evaluating Model Adequacy For Using Factor Score Estimates (Language Arts) 255
Table 7.9 Goodness of Fit Indices for Evaluating Model Adequacy For Using Factor Score Estimates (Mathematics) 255

LIST OF FIGURES

Figure 1.1. Cattell-Horn Model of Intelligence 29
Figure 1.2. Vernon’s Hierarchical Model of Intelligence........................30
Figure 1.3. Cattell-Horn-Carroll (CHC Model of Intelligence................32
Figure 1.4. Example of subtest similar to the Coding subtest – Adapted from the Wechsler Intelligence Scale for Children -4th edition (US) (WISC-IV (US)..36
Figure 1.5. The Feature Profile Test..38
Figure 2.1. Summary of the WISC-IV core and supplemental subtests .63
Figure 2.2. 2D Geometric design similar to those found in the Block Design subtest...65
Figure 2.3. Item similar to those found in the Picture Concepts subtest .65
Figure 2.4. Items similar to those found in the Matrix Reasoning test…66
Figure 2.5. Items similar to those found in the Coding subtest 67
Figure 2.6. Items similar to those found in the Symbol Search subtest...68
Figure 2.7. WISC-IV (US) current 4 factor structure..............................73
Figure 2.8. WISC-IV (US) core subtests within the CHC framework 75
Figure 2.9. Models 1 to 4: Confirmatory Factor Analysis.......................94
Figure 2.10. Standardized Loadings for the current WISC-IV four factor model 96
Figure 2.11. Simple path diagram for the cross cultural interpretive model for WISC-IV (US) measured intelligence in T&T children 108
Figure 4.1. Frequency distributions of WISC-IV (US) core subtests Block Design, Similarities, Digit Span, Picture Concepts, Coding, Vocabulary, Letter Number Sequencing, Matrix Reasoning, Comprehension, Symbol Search............127
Figure 4.2. Frequency distributions of WISC-IV (US) composites Verbal Comprehension, Perceptual Reasoning, Working Memory, Processing Speed and Full Scale IQ 128
Figure 6.1. Hypothetical WISC-IV (US) nomological network.............181

Figure 6.2. US Exposure Item of the Demographic Questionnaire 190

Figure 7.1. The Culture – Language Interpretive Matrix (C – LIM)
For the WISC-IV..223
Figure 7.2. Standardized coefficients: Model 1 238
Figure 7.3. Standardized coefficients: Model 2 131
Figure 7.4. Standardized coefficients: Model 3 132

LIST OF ABBREVIATIONS

ABASII	Adaptive Behaviour Assessment System – 2nd Edition
ACH	Achievement General Factor
ADHD	Attention Deficit Hyperactivity Disorder
AERA	American Educational Research Association
AGFI	Adjusted Goodness of Fit Index
AIC	Aikake Information Criterion
A’Level	Advanced Level
API	Academic Performance Index
BarOnEQ	BarOn Emotional Quotient
BIC	Bayesian Information Criterion
CA	Chronological Age
CAPE	Caribbean Advanced Proficiency Examinations
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CHC	Cattell – Horn – Carroll
C – LIM	Culture – Language Interpretive Matrix
CMS	Children’s Memory Scale
χ^2	Chi-square
χ^2/df	Normed chi-square
χ^2 diff	Chi-square difference
CSU	Cognition of Semantic Units
CTT	Classical Test Theory
df	Degrees of Freedom
DSM	Diagnostic and Statistical Manual
EFA	Exploratory Factor Analysis
EI	Emotional Intelligence
ELP	Estimated Learning Potential
EQ	Emotional Quotient

FSIQ	Full Scale IQ
g	General Intelligence
Ga	Auditory Processing
GAI	General Ability Index
Gc	Crystallized Intelligence
Gf	Fluid Intelligence
GFI	Goodness of Fit Index
Glr	Long Term Memory and Retrieval
Gq	Quantitative Reasoning
Grw	Reading and Writing Ability
Gs	Processing Speed
Gsm	Short Term Memory
Gt	Decision/Reaction Time
Gv	Visual Processing
IQ	Intelligence Quotient
IRT	Item Response Theory
KABC II	Kaufman Assessment Battery for Children 2nd Edition
km	Spatial – Mechanical Ability
LD	Learning Disability
MA	Mental Age
MOE	Ministry of Education
MR	Mental Retardation
O’Level	Ordinary Level
PMA	Primary Mental Ability
PRI	Perceptual Reasoning Index
p	Probability
PSI	Processing Speed Index
r	Correlation coefficient
rc	Tucker Coefficient of Congruence
RMSEA	Root Mean Square of Approximation

RPM	Raven’s Progressive Matrices
SEA	Secondary Entrance Assessment
SES	Socioeconomic Status
SEm	Standard Error of Measurement
SEM	Structural Equation Modelling
S – L	Schmid Leiman Orthogonalization
SLD	Speech/Language Disorder
SOI	Structure of Intellect
SOMPA	System of Multicultural Pluralistic Assessment
SRMR	Standardized Mean Square Residual
SSSD	Student Support Services Division
TLI	Tucker Lewis Index
T&T	Trinidad and Tobago
UNIT	Universal Non – Verbal Intelligence Test
US	United States
UK	United Kingdom
VCI	Verbal Comprehension Index
ved	Verbal – Educational Ability
WASI	Wechsler Abbreviated Scale of Intelligence
WIAT	Wechsler Individual Achievement Test
WISC	Wechsler Intelligence Scale for Children
WJIII	Woodcock Johnson Tests of Cognitive Abilities 3rd Edition
WMI	Working Memory Index
WPPSI	Wechsler Preschool and Primary Scale of Intelligence
WPT	Wonderlic Personnel Test

CHAPTER 1

INTELLIGENCE: DEFINITION, INTERNAL STRUCTURE, MEASUREMENT AND AETIOLOGY OF VARIABILITY

More than 100 years after Francis Galton's (1869) radical theory of inherited genius generated widespread scientific interest in the field, our understanding of human intelligence has remained fundamentally unchanged. Then, as now, intelligence was conceptualized as a latent capacity for success in one or more of life's pursuits. The field of individual differences, however, has not yet achieved consensus about the nature of this cognitive faculty. Is intelligence a capacity for learning or is it acquired knowledge? Is it a general ability or does it consist of multiple (correlated and uncorrelated) categories? Is it a universal construct, or is it culturally specific?

In 1997, a panel of renowned theorists agreed upon 25 empirically based conclusions about intelligence and an actual definition was proposed in Conclusion # 1:

> Intelligence is a very general capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking- smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – 'catching on', 'making sense' of things, or 'figuring out' what to do (Gottfredson, 1997a, p. 13).

The experts argued that this multifaceted human capacity has other important characteristics – it predicts important life outcomes; and what's more, "intelligence tests measure it well" (Gottfredson, 1997a, p.13). What remains unclear, is whether the definition of intelligence has been limited by the extent to which tests have been able to measure this faculty.

Indeed, while this definition has been successfully operationalized, like other conceptual definitions, it may lack universal applicability. This is because existing measures of intelligence chiefly estimate cognitive abilities associated with educational and occupational outcomes (Gottfredson & Saklofske, 2009). This is an approach that seems to exclude a wide range of non-Western perspectives. For instance, in some Asian societies, intelligent behaviours do not only promote academic and occupational success, but help to sustain communities (e.g. empathy, adaptation to change, humour, interpersonal skill, self-knowledge and even intellectual humility) (Zhang & Wu, 1994 cited in Cocodia, 2014; Yang & Sternberg, 1997). Also, African notions of intelligence regard academic ability as just one of a number of intellectual abilities that also include social problem solving, initiative, respect and responsibility (Grigorenko, Geissler, Prince, Okatacha, et al., 2001).

Howard Gardner (2000) and Robert Sternberg (1985) offer expansions on these cross-cultural themes that are of broader scope than traditional Western constructs. The Gardner model for example identifies eight different intelligences: linguistic, logical–mathematical, spatial, musical, bodily–kinesthetic, interpersonal, intrapersonal, and naturalistic (Gardner, 2000). Gardner asserts that each of the intelligences is specific to the requirements of a particular context (Reeve & Bonaccio, 2011). With the exception of the interpersonal, intrapersonal and naturalistic intelligences that reflect capacity for self-knowledge as well as awareness of others and the environment, Gardner's intelligences clearly predict success in a variety of occupational pursuits.

Measuring the social, emotional and naturalistic abilities has been challenging. For instance, Emotional Quotient (EQ) inventories, inspired by the Goleman (1996) theory of Emotional Intelligence (EI), have been designed to gauge an individual's awareness of their own and others' emotions. Such inventories generally have faced questions regarding their validity as measures of ability (Reeve & Bonaccio, 2011). In fact, EI measures have been said to tap into personality traits rather than cognitive ability (Robert, Schulze, & MacCann, 2008).

Sternberg's tri-faceted theory features three distinct principal factors: creative intelligence, academic/analytical intelligence; and also a kind of acquired knowledge that facilitates survival within indigenous environments (practical intelligence) (Sternberg, 1985; Sternberg & Hedlund, 2002; Reeve & Bonaccio, 2011). Sternberg's practical intelligence provides a useful framework for understanding the contextual nature of intelligence. While most multiple intelligence models are unsupported by empirical evidence (Reeve & Bonaccio, 2011), cross cultural studies have provided some compelling support for Sternberg's theory.

Practical intelligence is characterized by the acquisition of knowledge and skills that are learnt in an unstructured manner and without the person's awareness (Reber, 1989; Sternberg & Hedlund, 2002). Studies have compared this tacitly gained intelligence with crystallized intelligence, which is knowledge that is mainly acquired in formal academic settings (Sternberg & Hedlund, 2002) and that is measured by Western tests of cognitive ability. As an example, one study found that indigenous Kenyan children's performance on practical intelligence items was negatively correlated with their performance on established tests of crystallized intelligence (Sternberg, Nokes, Geissler, Prince et al., 2001). The authors argued that the time taken to gain one kind of knowledge negatively impacted acquisition of the other. Children who dropped out of school early were better able to gather the necessary skills for successful adaptation to their natural environment. Conversely, they were less likely to demonstrate formalized or academic knowledge. In another study, two tests of crystallized intelligence, each specific to an ethnic region, were administered to Filipino children from the two respective communities (Church & Katigbak, 1987). Results showed that children performed better on tests that were relevant to their indigenous surroundings.

If there are cultural differences in the way intelligence is defined and measured, then no universal definition of intelligence can really exist (Sternberg, 2004). If a universal definition could be formulated, it would be infinitely wordy and complex. Sternberg (2004) argues that while the cognitive mechanisms associated with intelligence are universal, cultural and environmental differences demand that they are operationalized differently across contexts.

Current definitions about what makes a person intelligent range from implicit concepts like Gardner's (2000) social and naturalistic intelligences, to empirically supported hierarchical models (e.g. Spearman, 1904). Although early frameworks of intelligence emerged from psychometric concepts, recent models have been articulated as a more complex network of what Cronbach and Meehl (1955) defined as nomologicals or laws that surround a central construct (Reeve & Bonaccio, 2011). These nomological models provide a framework by which we can understand how environmental variables foster or limit the development of abilities. Additionally, the network tells us how cognitive abilities are related to other individual characteristics or outcomes. In other words, whatever is the nature of intelligence, there is a network of factors around these intellectual faculties that give them additional meaning.

From this perspective, the chapter will explore the psychometric structure as well as provide an overview of the construction, reliability, and validity of intelligence tests. The influence of the model on test design and controversies in assessment will also be discussed. It will be shown that in spite of the broad definition of intelligence, test score interpretation has continued to focus narrowly on the measurement of abilities. Additionally, the exclusion of environmental variables from interpretive and scoring models is argued to perpetuate a narrow framework for the interpretation of cross cultural differences in intelligence. It is proposed that to assess the cross cultural validity of any intelligence test both the internal and external aspects of the network must be considered. Later sections will discuss the antecedent and outcome variables of the intelligence network. The chapter will conclude with a brief outline of the research aims of this thesis.

The Internal Structure of Intelligence

Specific abilities vs the concept of g. In 1904, through the efforts of Charles Spearman, the theoretical definition of intelligence achieved some clarity. By examining the correlational matrices of a dataset of academic scores, Spearman (1904) articulated the first psychometric model of intelligence. He extracted a single general factor of intelligence (*g*) which he deemed to be common to all of the tests. What remained unaccounted for by the general factor, he named the specific factors (*s*). These *s* factors defined those aspects of the test that remained unique to the test alone and were not correlated with each other.

Ensuing exploratory studies by psychometrician, Louis Thurstone (1938), challenged Spearman's model. Using another method of factor analysis, Thurstone proposed a multifactor model of specific abilities which he called primary mental abilities or PMAs. These abilities were identified as follows: verbal, space (visual spatial), number (quantitative), perceptual, memory, word fluency, inductive and deductive reasoning, and closure. He argued that these mental abilities were uncorrelated (Thurstone, 1938). Thurstone’s statistical methods prohibited the extraction of a general factor, an approach that was severely criticized by Spearman (1939). Eventually Thurstone conceded that a higher general factor was statistically reasonable, but still not as important as the specific abilities in defining human intellect (Reeve & Bonaccio, 2011).

Guilford (1966) hypothesized a model that was influenced by Thurstone’s PMAs. Like Thurstone, he argued that because of too much demographic variability within the sample, Spearman’s *g* theory was weak. Since most ability tests tend to correlate with age and education, Guilford suggested that on the basis of these factors alone, these tests are likely to correlate with each other. Therefore, by controlling for age and education, there is likely to be a cluster of zero correlations in the matrix, making it impossible to extract a general factor (Guilford, 1966).

Using a different statistical method, Guilford proposed his Structure of Intellect (SOI) model which he presented as a 120 celled three-dimensional matrix. Each cell represented a specific ability, but furthermore, each ability could be conceptualized as a composite of specific operations with specific contents, and with specific products (Guilford, 1966). As an example, consider one of Guilford’s abilities, CSU, or Cognition of Semantic Units where Cognition represents an operation, Semantic represents the content, and Unit stands for the product (Guilford, 1966). The CSU factor would reflect an individual’s ability to understand an abstract concept. Take the concept of ‘freedom’ as an example. On the basis of the SOI model, the process by which a person understood ‘freedom’ may be explained by this sequence of steps:

1. Operation (creating mental pictures);
2. Content (images of freedom); and
3. Product (verbal definition of freedom).

The Guilford model, otherwise known as a facet model, has not found empirical support (Carroll, 1993) and has been essentially abandoned in the literature (Reeve & Bonaccio, 2011).

Traditional factor analytic models have been more widely accepted. As an example, the Cattell model of intelligence, which later became the Cattell-Horn model, was famous for its proposal of a dichotomous framework comprising crystallized intelligence (*Gc*; knowledge and skills acquired through education and experience); and fluid intelligence (*Gf*; ability to reason and solve novel problems without the benefit of prior knowledge) (Cattell, 1971/1987; See Figure 1.1). Crystallized ability was said to be the result of the investment of fluid reasoning ability over time (Cattell, 1971/1987).

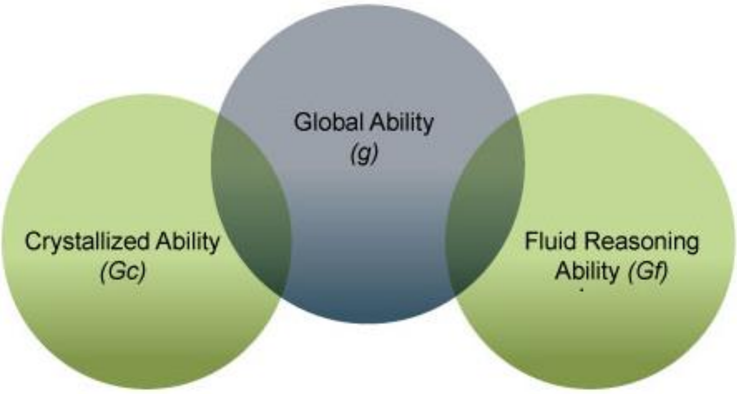


Figure 1.1. Cattell-Horn Model of Intelligence (Cattell, 1972/1987).

In 1997, Horn and Noll produced a refined version of the Cattell (1943) two specific factor model, in which featured higher order crystalized ability and fluid ability factors with 10 specific abilities on the lower level (Horn & Noll, 1997). Like Thurstone’s original model, this model did not include a higher order global factor in spite of the existence of a positive manifold (Reeve & Bonaccio, 2011). Later, using factor analytic methods, it was possible to extract a *g* factor from Horn and Noll’s (1997) data. Additionally, the studies demonstrated that the *Gf* factor and the *g* factor were almost indistinct (Reeve & Bonaccio, 2011).

The hierarchical structure of intelligence. Vernon (1961) is credited with postulating the first hierarchical model of intelligence (Reeve & Bonaccio, 2011). Although Spearman and Thurstone theorized a model of primary mental abilities with a summary general factor, they did not present it as part of a hierarchical model. The difference was that the Vernon theory summarized intelligence as a multileveled arrangement of skills and abilities. Observable behaviours were on the lowest level, and higher up the hierarchy, were the latent capacities that are measured by clusters of these observable behaviours. The idea was that correlations among tests of ability may in fact be explained by latent ability factors (Carroll, 1993). The higher up the hierarchy the more important are the factors, each successive level explaining the correlation among an increasing number of skills or abilities. Vernon's model hypothesized an overarching general intelligence factor, and on the level below are two 2nd order factors of verbal-educational ability (*ved*) and spatial-mechanical ability (*km*). On the 1st level are skills that are considered to be associated with educational success (verbal fluency, numerical skills and reasoning skills) and also mechanical achievement (psychomotor coordination and spatial ability) (Vernon, 1961; see Figure1.2).

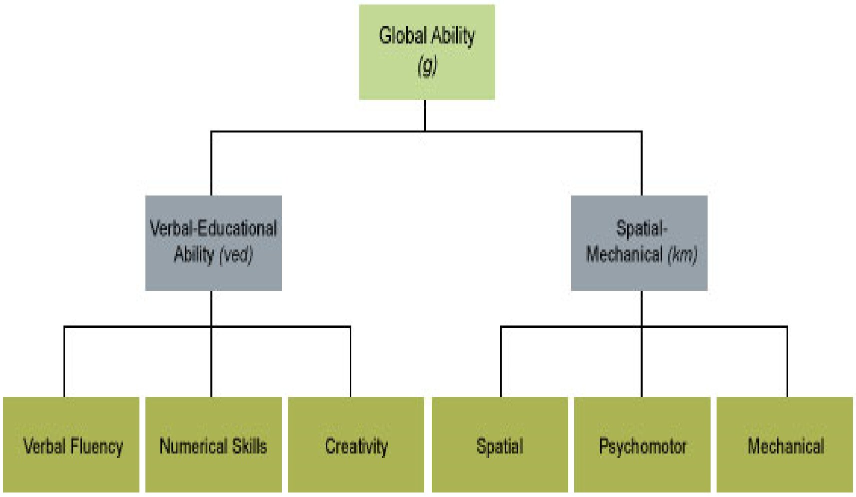


Figure 1.2. Vernon's Hierarchical Model of Intelligence (Vernon, 1961).

Carroll (1993) analyzed over 400 data sets and presented a three levelled model of intelligence consisting of 60 specific abilities on the first stratum, 8 broad abilities on the 2nd level and a higher order *g* factor. The 1st level abilities – fluid reasoning, crystallized ability, general memory and learning, processing speed, broad cognitive speediness, broad retrieval ability, broad auditory perception and broad visual perception are inter-correlated. The Cattell (1943), Vernon (1961), Horn (1968) and Carroll (1993) models of psychometric intelligence were essentially enhancements of the Spearman theory - differing in complexity but comparable in basic configuration: a higher order general factor, with specific factors on the lower levels of the hierarchy (Sternberg & Kaufman, 1996). McGrew (2009) developed the 3 stratum Cattell-Horn-Carroll (CHC) model which can be described as a synthesis between the Cattell-Horn (Horn & Noll, 1997) and Carroll (1993) models. The CHC model includes *g* at the top of the hierarchy, followed by 10, 2nd order broad ability domains, and over 70 task specific skills on the 1st stratum (Taub, Floyd, Keith, & Mc Grew, 2008; Reeve & Bonaccio, 2011) (See Figure 1.3.). The broad ability domains include the following: crystallized intelligence (*Gc*), fluid intelligence (*Gf*), quantitative reasoning (*Gq*), short term memory (*Gsm*), long term memory and retrieval (*Glr*), processing speed (*Gs*), visual processing (*Gv*), auditory processing (*Ga*), reading and writing ability (*Grw*) and decision/reaction time (*Gt*) (McGrew, 2009; Reeve & Bonaccio, 2011).

The CHC taxonomy has been the preferred model of reference for test developers (Newton & McGrew, 2010). Many intelligence tests measure subsets of CHC abilities based on their associations with key outcomes. School psychologists, for example, are interested in the relationship between cognitive abilities and academic performance as such knowledge can have positive implications for diagnosis and instructional planning in schools.

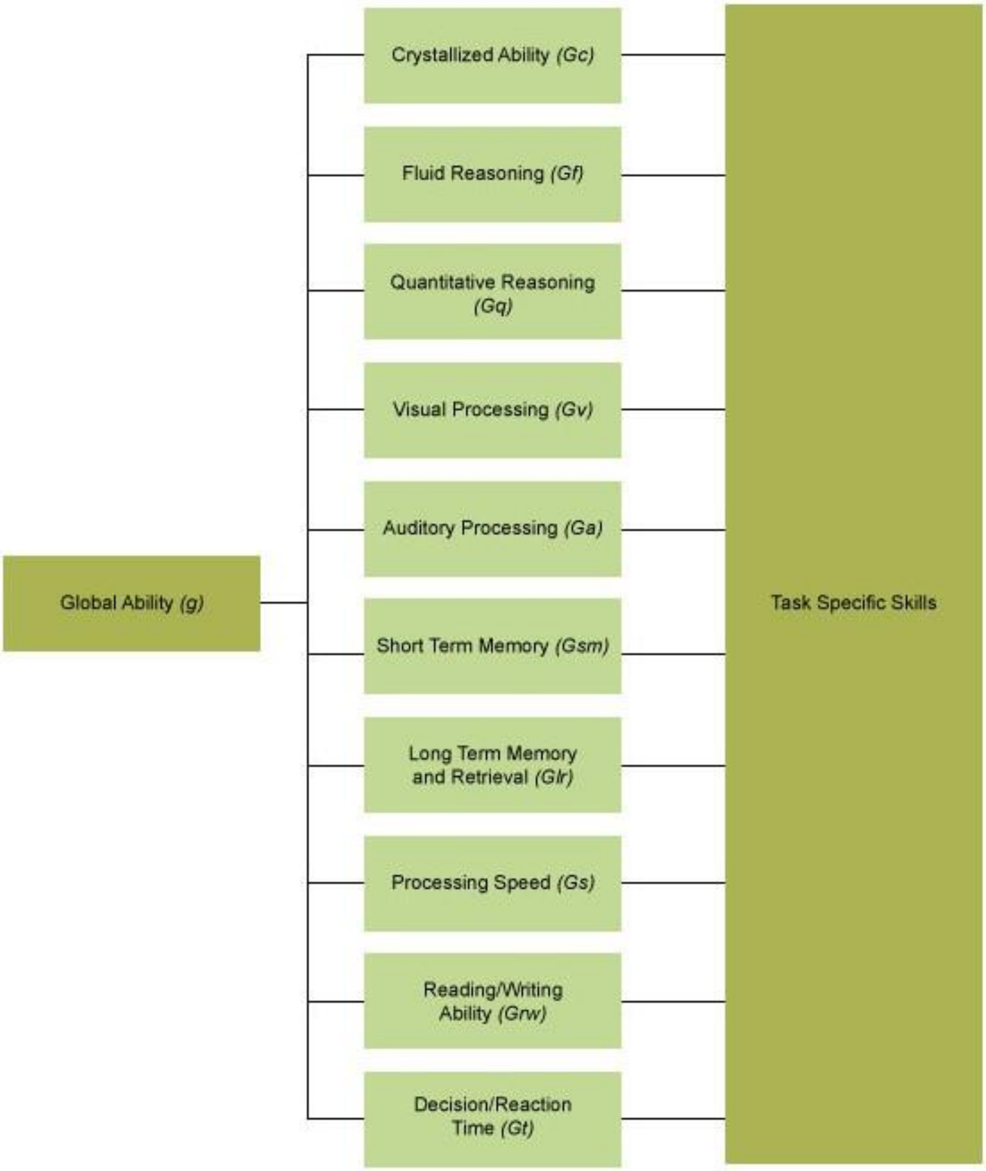


Figure 1.3. Cattell-Horn-Carroll (CHC) Model of Intelligence (McGrew, 2009) – Adapted from Reeve and Bonaccio (2011).

It therefore is not surprising that the assessment tools used in special education draw so liberally from the CHC broad ability domains. Such breadth in design is validated by studies demonstrating the significant correlation between specific abilities and a wide range of academic tasks (Fiorello & Primerano, 2005; Swanson & Jerman, 2006; Floyd, McGrew & Evans, 2008; Taub, Floyd, Keith & McGrew, 2008; McGrew & Wendling, 2010). The Woodcock Johnson Tests of Cognitive Abilities – III (WJ III, Woodcock, McGrew & Mather, 2001), for example, measures nine out of the ten CHC broad ability domains (Taub & McGrew, 2004). Similarly, the Wechsler Intelligence Scale for Children – 4th Edition (WISC-IV; Wechsler, 2003a) taps into between four and six 2nd stratum CHC factors (Flanagan & Kaufman, 2004; Lecerf, Rossier, Favez, Reverte et al., 2010), and the Kaufman Assessment Battery for Children – 2nd Edition (KABC – II, Kaufman & Kaufman, 2004) assesses five CHC factors (Reynolds, Keith, Fine & Fisher, 2007).

Intelligence Definition still "Under Construction": Measurement Issues

The previous sections presented an overview of the definition and internal structure of intelligence. This modern intelligence paradigm, albeit multi-layered and complex, appears to suffer from a lack of generalizability. This may be because individual cultures define intelligence as behaviours that promote their own unique aspirations. Cross cultural studies have shown that people generally tend to perform better on tasks that demonstrate some relevance to their particular environments (e.g. Church & Katigbak, 1987; Sternberg et al., 2001). Such notions offer a strong rationale for shifting the focus from universal to context based models of intelligence. For example, multiple intelligence theories argue that abilities should be defined in relation to clearly defined outcomes, such as success in music or interpersonal relationships, excellence in sports or academics, or survival in challenging environments (Sternberg, 1985; Gardner, 2000). This perspective has practical value, because it encourages specificity about the abilities we are seeking to measure. For example, global intelligence has been shown to predict health behaviours (Gottfredson & Deary, 2004; Wraw, Deary, Gale & Der, 2015); however, there are conceivably 'highly intelligent' people who do not adhere to the best health practices. This could be explained by genetic and biological factors (e.g.

alcoholism), and also variables within the environment (such as environmental toxicity, availability of good or affordable health care, personality or motivation). But there is also reason to assume that health behaviours may be predicted by as yet unaccounted for specific intellectual abilities. As such, research deliberately focused on identifying these particular capacities can provide much needed insight into health related behaviour. The same rationale can be perhaps applied to studying intelligences related to academic, occupational and other outcomes.

The following sections will show that while prominent Western notions have described intelligence as a universal construct, early intelligence tests were restricted in focus to behaviours associated with a limited range of life outcomes, namely occupational and academic success. Additionally, by retaining many of the early test items, modern instruments have perpetuated this traditional narrow focus.

Origins of intelligence testing: predicting occupational and academic success. In just over 100 years, intelligence measurement has evolved from a non-scientific study of intellectual differences into a science of psychometrics. Early examples of this late 19th century innovation differed from modern day instruments because they did not rely on empirically based models for their design, but rather on speculative efforts to account for individual differences in achievement (White, 2000). Indeed, intelligence as a psychometric construct was defined after tests were developed to measure it.

Pioneers in intelligence testing placed less emphasis on the structure of intelligence than the extent to which it predicted an outcome. The idea provided a useful basis for the systematic observation of behaviours that predicted success and was a valuable first step towards the operationalisation of intelligence. Curiously, these naturalistic observations were not reflected in the test design of that time. For example, Galton deduced that intellectual ability was indistinct from and estimable by other inheritable physical, neurological and sensory traits (Beins, 2010). Galton’s assessments therefore placed considerable focus on the measurement of anthropometric features (Boake, 2002) such as head circumference, reaction time, sensory discrimination and grip strength (Fancher, 2009). Predictably, the first mental tests were criticized for their weak inter-item correlations and poor convergent validity (Fancher, 2009). Spearman (1904) remarked on the merit of Galton's hypotheses, but lamented that his theory failed to develop sufficiently as a result of a lack of scientific study. Additionally, influenced by his cousin, Charles Darwin’s, theories of natural and artificial selection, Galton’s ‘science’ became muddled within a eugenic ideology, the goal of which was to enhance the intellectual make-up of the population. The eugenic perspective held that society's condition could be vastly improved if persons of highest intellectual ability were encouraged to (and less 'fit individuals were discouraged from) procreating (Galton, 1904). This idea of selective breeding received widespread support throughout Europe and the United States of America. The tragic legacy of Galton’s controversial ideas is well known, and understandably, intelligence testing continues to be viewed with considerable distrust.

In spite of all of its limitations, Galton's theory can be credited with igniting a movement of scientific investigation into human intelligence that has influenced test design today. James McKeen Cattell (1890) studied intelligence through what he deemed to be measurable individual differences in 'bodily, psychophysical and mental' measurements. Cattell’s ‘mental tests’ measured grip strength, sensory acuity, reaction time, working memory and rate of movement (Cattell, 1890). It would soon be shown that Cattell's measures of 'intelligence', like those of Galton, demonstrated few reliable associations with real life outcomes (Fancher, 1985). Still, there were noteworthy exceptions. One key intellectual skill, memory of randomly presented numbers, has been enhanced from its early format into what is today used as a measure of working memory (Boake, 2002) – an ability that studies have linked with language processing, reasoning, vocabulary and grammar learning, reading and mathematics (Engle, Tuholski, Laughlin & Conway, 1994; Baddeley, 2002; Henry & MacLean, 2003; Colom, Rebollo, Palacios, Juan-Espinosa & Kyllonen, 2004; Andersson & Lyxell, 2007; Swanson & Kim, 2007; Passolunghi & Cornoldi, 2008; Alloway & Alloway, 2010; Verhagen & Leseman, 2016). Another survivor from early instruments is the Substitution test that was designed by Cattell's students, R. Woodworth and F.L. Wells (1911) (Boake, 2002). The visual association measure has evolved into the modern day Coding test, a measure of learning speed and attention (Wechsler, 2003b). Figure 1.4 provides an example of a test that is similar to the Coding subtest of the Wechsler Intelligence Scale for Children – 4th Edition (US) (WISC-IV (US); Wechsler, 2003a).

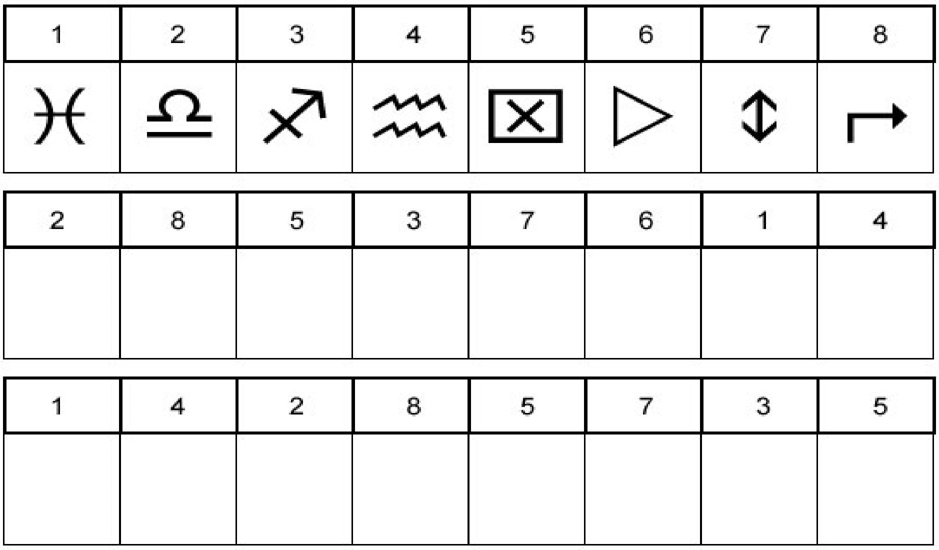


Figure 1.4. Example of subtest similar to the Coding subtest – Adapted from the Wechsler Intelligence Scale for Children – 4th Edition (US) (WISC-IV (US); Wechsler, 2003a).

In 1904, the French Government commissioned Alfred Binet to develop a tool for identifying students who were likely to fail at school (Thorndike, Cunningham, Thorndike & Hagen, 1991). Binet designed items to determine how well children met their age milestones. For each age level, Binet assigned 6 items that could be answered by a child of corresponding age and average ability. For example, an average six-year-old would be capable of identifying parts of the body, or defining simple words (e.g. Items 14 and 17; Binet & Simon, 1905/1916.). An average 13-year-old was expected to demonstrate more advanced verbal concept formation and visuospatial imagery (e.g. Items 55 and 57; Binet & Simon, 1905/1916). Performance on the test was summarized as the age level at which all tests were passed, increased by a year for each additional five tests that were successfully completed (Binet & Simon, 1905/1916). Therefore, children were diagnosed as functioning at a particular ‘Mental Age’, with guidelines for diagnosis of intellectual deficiency depending on how far the Mental Age fell below their chronological age (Binet & Simon, 1905/1916).

An American version of the Binet Scales was published by Lewis Terman and his colleagues in 1917. In this format, estimations of intelligence were summarized as an Intelligence Quotient or IQ (Terman, Lyman, Ordahl, Ordahl et al., 1917). The IQ score was calculated by dividing Binet’s Mental Age by the chronological age and multiplying by 100 (MA/CA x 100) (Terman, 1917). IQ was understood as the child performing at a certain percentage of their expected capacity.

The Binet-Simon and Stanford Binet scales proved effective as predictors of academic achievement, but they lacked an underlying theoretical construct. Additionally, the Stanford Binet scales were criticized for their heavy verbal content which made the battery inappropriate for use with illiterate persons or non-English speakers (Boake, 2002). In spite of its imitations, the Binet Scales could be credited with influencing the verbal content of modern instruments (Boake, 2002). Early tests of nonverbal ability also influenced the content of modern intelligence tests. For example, the Picture Arrangement test of Wechsler Intelligence Scale for Children-III (WISC-III; Wechsler, 1991) was adapted from an early measure of sequential deductive reasoning ability designed by Healy and Fernald (1911). Another example, the Feature Profile Test required test-takers to assemble a puzzle of a human face (Boake, 2002; See Figure 1.5). This test later appeared in the Wechsler scales as the Object Assembly test (e.g. Wechsler, 1974/1991).

Figure 1.5. The Feature Profile Test. Obtained from Boake (2002).

Modern intelligence testing: estimating abilities. In spite of their historical commonalities, modern tests of intelligence display considerable variety in content because they are based on differing paradigms. The Standards for Educational and Psychological Testing of the American Educational Research Association (AERA, 1999), suggest that a clear conceptual framework should underlie all test design. The Woodcock Johnson III (WJIII) for example has measured nine CHC abilities which have been shown to predict performance on a wide range of academic tasks (McGrew, Flanagan, Keith, & Vanderwood, 1997; Keith, 1999; Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001;

Floyd, Evans, & McGrew, 2003; Fiorello & Primerano, 2005; Swanson & Jerman, 2006). Other measures such as the Raven’s Progressive Matrices (RPM, Raven, Raven & Court 1998), the Universal Non Verbal Intelligence Test (UNIT; Bracken & McCallum, 1998), and the Wonderlic Personnel Test (WPT; Wonderlic, 1992) have been designed to measure higher order general ability. The relevance of general intelligence is argued to surpass that of specific abilities, because even instruments designed to measure specific abilities load heavily on g (Lubinski, 2004; Watkins, Glutting & Lei, 2007; Watkins, 2010). This view is supported in the general literature which associates g with a variety of life outcomes, including health behaviour, academic and job performance and psychological well-being (Hunter & Hunter, 1984; Jensen, 1998; Schmidt & Hunter, 2004; Glutting, Watkins, Konold, & McDermott, 2006; Gottfredson & Saklofske, 2009).

Item selection and analysis. To obtain reliable and valid estimates of intelligence, test designers must elicit behaviours associated with the underlying ability. Accuracy in item selection is therefore crucial to ensuring good test reliability and validity. In intelligence measurement, item selection is not a straightforward matter. Whereas criterion referenced measures such as achievement tests consist of items designed to assess mastery of clearly defined skills (Kline, 2000); this is not necessarily true of intelligence tests. The difference is that intelligence test items are tapping into a latent ability which cannot be measured directly but can only be inferred from performance on a test; therefore, intelligence test content need not look like what they are purporting to measure (Gottfredson & Saklofske, 2009).

After a period of review, items are administered to a trial sample of the target population (Kline, 2000). Pilot data will reveal the psychometric properties of the test items by summarizing their structure. Additionally, pilot analyses may determine their relation to other items of the test as in classical test theory (CTT) models (Lord, 1980;

Hambleton & Jones, 1993) or use more advanced methods as in item response theory (IRT) to evaluate items according to how well they discriminate among persons of differing ability (Hambleton & Jones, 1993).

Standardization. Score interpretation in intelligence testing is based on comparisons with similar individuals within the population (Kline, 2000). In other words, a raw score on any intelligence test is meaningless unless we know whether it is within, below, or above expectations based on group membership. Therefore, accuracy of interpretation is only possible if normative data is gathered from a representative sample (Kline, 2000). Comparisons are made by examining *z scores* which are a function of the mean and the standard deviation of the distribution of scores in the normative sample (Fischer & Milfont, 2010). A *z score* is equivalent to the standard deviation, so if the mean score is 25 and the standard deviation is 10, then a score of 35 will have a *z score* of +1 and a score of 15 will have a *z score* of -1. Many tests convert *z scores* to normalised standard scores that are easier to interpret (Kline, 2000). Scales like the Wechsler Intelligence Scale for Children – IV (Wechsler, 2003a) utilize a mean of 100 and standard deviation of 15. Using the normal distribution, standard scores can also be presented as a percentile score, which summarizes the percentage of individuals in the normal sample that fall beneath that score (Kline, 2000). Therefore, a percentile score of 80, means that the individual's performance exceeded 80 percent of the normative sample.

Assessing reliability. A reliable intelligence test makes stable estimates of an individual's ability. The concept is well articulated through classical test theory (CTT). According to CTT, the individual's true IQ score is a sum of the test score and random error (Hambleton & Jones, 1993). Therefore, the extent to which the estimate of intelligence is different from the true intelligence is as a result of error (T = O + E, where T is the true score, O is the observed score, and E is random error; (Rust & Golombok, 1992; Hambleton & Jones, 1993). A test's reliability can be thus demonstrated via the degree of error. It is important to stress that a reliable test is not immune to error; rather, error is minimal. The standard error of measurement (SEm) is an indication of the amount confidence we can place in the observed score as an accurate estimate of the true score (Kline, 2000; Wechsler, 2003b). Through the SEm, test developers can produce confidence intervals or the range of scores to be expected if the person is tested on multiple occasions. Tests of higher reliability tend to have smaller confidence intervals (Wechsler, 2003b). In IQ testing, confidence intervals are calculated at a percentage value (e.g. 90% or 95%), which suggests that if we were to test the individual on multiple occasions, then 90 or 95 percent of the time, the scores are expected to fall within the particular confidence range.

A reliable measure of intelligence will, barring random error, estimate intelligence similarly each time. *Test retest reliability,* determines the extent to which intelligence scores remain the same on subsequent administrations of a test, and is determined by calculating the correlation between scores obtained at different administration times (Rust & Golombok, 1992). Any score variation should be the result of naturally occurring and random error (e.g. room temperature, mood). The correlation coefficient (*r*) summarizes agreement between scores collected at two different administration times. Agreement values in excess of .90 are considered to be acceptable in intelligence testing (Rust & Golombok, 1992). Of course one must be aware of some of the factors associated with test-retest reliability that may either artificially elevate or attenuate *r* values, such as learning of items or maturation of the participant (Kline, 2000). The *split half reliability* approach which determines the internal consistency of the test is particularly useful at reducing or eliminating learning and maturation effects because only one administration is needed (Kline, 2000). In this case, the correlation between scores on different halves of the test is calculated. *Split half reliability* calculations will likely run into difficulty if the test is simply split into early items and later items (Rust & Golombok, 1992). This is particularly true if, as in IQ tests, the items increase in difficulty as testing proceeds (Kline, 2000). The prescribed approach is to calculate agreement of scores between odd and even items.

Assessing validity. Reliable tests are not necessarily valid. This can be attributed to the difference between two types of error, systematic error, and random error (Kline, 2000). As opposed to random errors that are outside the control of the test, systematic errors are inherent in the test and efforts should be made, both in test development and in providing guidelines for test administration, to ensure that systematic errors are minimized (Kline, 2000).

Early in the history of intelligence testing, researchers were aware that investigations were necessary to determine the extent to which test items correlated with each other and by how well they correlated with outcome measures. It was through the work of Axel Oehrn and other researchers that associations among simple mental processes, such as perception, memory and association, were investigated (Spearman, 1904). A major development arose through the work of Franz Boas who extended Oehrn's internal validity studies to investigate the relationship between children's test scores and teacher estimates of intellectual ability (Spearman, 1904). Many other external validity studies had been conducted during the decade since Galton's initial proposals, but they were criticized by Spearman (1904) for lacking a strong methodological basis and a disregard for error and nuisance factors. Spearman eventually demonstrated associations between sight, sound and weight discrimination and different real world measures of intelligence such as class rank, teacher ratings, and non-academic intelligence. Furthermore, Spearman demonstrated that a discrimination skills factor and a general intelligence factor were perfectly correlated. Terman also investigated the predictive validity of a category of adult feeblemindedness on occupational outcomes. He demonstrated that it was very unlikely to find individuals fitting the diagnostic category for feeblemindedness in sustainable employment (Terman, 1917). Clearly, early validity studies were chiefly concerned with establishing relationships among test scores, and also between test scores and real life outcomes. Today, the exploration of validity has become more complex.

Whereas reliable intelligence tests increase our confidence that the test is measuring the same *construct* each time, valid tests assure us that the construct is actually intelligence. As such, strong evidence of validity provides empirical support for score-based interpretations. Evaluation of the internal validity of tests can be by factor analyses which examine sample scores to determine whether item inter-correlations fit the model that is proposed by the test's authors (Kline, 2006). *Content validity* studies investigate how well the test design (items, rules about administration and scoring) reflect the construct that is being measured (Rust & Golombok, 1992; Kline, 2000). Exploration of content can provide useful information, however, this may not be a strict requirement in intelligence measurement because the appearance of compatibility with the underlying construct is not essential.

Explorations of the broader construct assess the relationships between test scores and other *criterion* measures. External validity is generally assessed through correlations with a future criterion or other measures. Investigations of validity can ascertain how well a test correlates with other measures of the same construct (Rust & Golombok, 1992). Often in test development, researchers obtain sample scores on other established intelligence tests. High correlations among similar tests reflect good *convergent validity*. Conversely, weak correlations are anticipated with tests of differing or opposing constructs. Examining a test score's correlation with a future criterion (test performance, future behaviour, or future life outcome (Rust & Golombok, 1992; Kline 2000) also provides a useful estimate of a test's external validity. In intelligence testing, the tester is often interested in predicting future academic performance. Tests with good criterion validity therefore can be useful for ruling out learning disabilities in children. For example, a child with average intellectual ability (50% percentile) who is performing in the lower extreme percentile (<2%) in reading may very well meet the criteria for a Specific Learning Disability in reading, provided all other conditions for acquiring reading skills are within appropriate levels (American Psychological Association (APA), 2000).

Controversies in Intelligence Testing

Inadequacy of intelligence models. In spite of finding widespread empirical validation, intelligence testing has been weighed down by controversy, attributed in part to its inauspicious beginnings, but also to the conceptual limitations of current intelligence models. As an example, while psychometric intelligence has been positively correlated with academic outcomes, it has been shown to explain performance on some subjects better than others. A study by Deary, Strand, Smith and Fernandes (2007) demonstrated that intelligence explained 58.6%, 48% and 18.1% of variance in Mathematics, English, and Art performance respectively. Similarly, Rimfeld, Kovas, Dale and Plomin (2015) demonstrated that correlations between intelligence and academic performance ranged from .36 for Art to .56 for Mathematics. Since academic subjects also load differently on the general academic factor (Kline, 2000), intelligence may have stronger associations with subjects that have higher academic loadings. The findings however, may have highlighted deficiencies in existing psychometric models which appear to explain some academic outcomes better than others.

The insufficiencies of current psychological models expose the pragmatic side of intelligence theory and measurement. It is clear that the extent to which developers tap into one or another factor is partially based on the ease with which a factor can be operationalized. There are indeed abilities (e.g. emotional intelligence and ability to acquire and recall explicitly taught rules) that have proven to be very complicated to measure (Guthke, Beckmann & Dobat, 1997; Reeve & Bonnacio, 2011) and therefore show up less on psychometric tests. Emotional quotient tests, for instance, have demonstrated weak associations with other established intelligence measures (e.g. Wechsler, 2003a/2003b), and have been argued to tap into personality more than ability (Robert, Schulze, & MacCann, 2008). Additionally, dynamic tests of learning ability have suffered from problems of reliability and also validity; and as a consequence, with a very few exceptions, they virtually have been abandoned by test developers, (Guthke, Beckmann & Dobat, 1997). Other concerns regard the extent to which current instruments measure latent ability at all (Godfredson & Saklofske, 2009). Even if we accept the argument that modern intelligence tests fundamentally estimate a person’s capacity for academic success, more complicated issues regard the extent to which intelligence tests are measuring the latent capacity or achievement itself.

Ability or achievement? The fact that so many of the ability tests that are currently used have content that resemble the knowledge and skills gained through education has been the source of much discussion. For example, the actual nature of *Gc* is not clear (Kan, Kievit, Dolan & van der Mass, 2011). Does *Gc* represent latent ability or achievement?

Cattell’s (1971/1987) investment theory appears to suggest the latter. According to the theory, *Gc* (acquired knowledge and skills) is developed through the application of reasoning and problem solving ability over time. If we accept this, then conceivably *Gc* can be described as an outcome of learning rather than an actual capacity to learn – a formative variable and not a latent variable (Kan et al., 2011). From such a perspective, *Gc* is not conceived as a causal factor, but as a statistical entity that is constructed from latent variables (Kan et al., 2011). The extent to which *Gc* is developed therefore would not merely depend on learning aptitude, but also on access to and quality of learning opportunities. Indeed, studies have demonstrated that *Gc* is positively correlated with environmental variables such as socioeconomic status (SES) and years of education (Cahan & Cohen, 1989; Rinderman, Flores-Mendoza & Mansur-Alves, 2010).

Intelligence or a covariate of intelligence? Implications of the Flynn Effect. James R. Flynn (1984) noted an increase of 13.8 points in American intelligence test scores over a period of 46 years. Also, international studies revealed intergenerational gains in IQ of 5 to 25 points (Flynn, 1987). This phenomenon has been described in the literature as *The Flynn Effect*, and has sparked much debate over the implications of such a drastic increase in population mean IQ over such a relatively short time. Flynn also argued against the suggestions that SES, education and even increases in test sophistication could be proposed as major influential variables. At best, according to Flynn (1987) these variables together may explain a 5pt inter-generational increase in test scores. Flynn posited that the residual variance could be potentially accounted for by as yet unidentified environmental variables.

But the most salient of Flynn’s arguments challenged the notion that IQ tests measured intelligence at all. According to Flynn (1987), all IQ tests, crystallized ability and fluid reasoning ability measures alike, do not measure intelligence, rather a variable that is linked to but has a weak causal relationship with intelligence. This suggests that any cross generational differences in test scores are not reflective of increases in intelligence per se, but increases in this yet unidentified covariate. The same could be said for between group differences which Flynn (1987) attributed to cultural differences rather than to differences in the latent trait. What are these cultural differences?

Culture, environment and the contextual nature of intelligence. Barber (2005) analysed data from over 81 countries and found that agricultural societies demonstrated lower mean scores, and countries with higher levels of secondary school enrolment demonstrated higher mean scores on measures of intelligence. The study also found statistically significant correlations between literacy and IQ (Barber, 2005). Barber's findings are consistent with those of the practical intelligence studies described earlier in the chapter (see Sternberg, Nokes, Geissler, Prince et al., 2001; Sternberg & Hedlund, 2002), which identified a positive correlation between formal education and crystallized ability test performance. Barber's and Sternberg's interpretations of the findings however may differ. To Barber (2005), the results imply that access to formal education is positively associated with opportunities for 'intellectual stimulation' and 'cognitive development'. This notion seems at odds with that of Sternberg (1985/2004) for whom group differences in IQ scores have less to do with variability in cognitive development than with how cognitive ability is demonstrated across cultures.

Sternberg's theory of successful intelligence (Sternberg & Grigorenko, 2000; Sternberg, 2004) asserts that among its three main components, the higher order mental processes of intelligence are universal. The higher order mechanisms characterize how an individual appraises and solves problems and are less likely to differ across cultures. Conversely the other components of intelligence, the actions performed to solve the problem and the ability to learn the actions, may vary cross-culturally (Sternberg, 2004).

For example, the capacity to navigate one's environment (i.e. spatial intelligence) may be manifested differently in the arctic compared to a large Western city (see Sternberg & Hedlund, 2002; Sternberg, 2004). The contrasting environments, each with their own unique configurations and challenges, are thought to differentially shape the knowledge and skills that are acquired therein. It may therefore be equally unfair to assess spatial intelligence in a Londoner using map of the frozen tundra, as it would be to use a city maze puzzle to assess the same skill in an Inuit person. Failure to consider cultural or environmental variables in intelligence measurement is likely to increase the risk of biased estimations of intelligence.

Cultural bias in intelligence measurement. Van De Vijver and Tanzer (2004) define bias as what occurs when differences in scores on a test are not attributable to differences in the measured attribute, but to other unaccounted for group based variables. Reynolds, Lowe and Saenz (1999) assert that mean differences in IQ between cultural groups on any test are not related to differences in ability; but rather to the characteristics of the instrument itself that may be subject to bias, such as the items, norms, language, administration procedures and flaws within the central construct that may favour one group over another. As a consequence, not only are the test score interpretations invalid and not replicable (van de Vijver and Tanzer, 2004), but biased estimates of intelligence can result in unfair labelling, inaccurate behavioural predictions or inappropriate academic interventions (Reynolds, Lowe, Saenz, 1999).

Concerns about bias have served as a rationale for test modification and adaptation in many countries. With regard to US designed tests, studies have suggested that they measure cognitive skills and information that are more likely to be acquired by middle class European-American children (Helms, 1992). For example, a study by Tynes - Jones (2005) found that the 3rd edition of the WISC (US) underestimated psychometric intelligence in a Bahamian population. Other studies have pointed to major differences in IQ scores across standardization samples: Navajo children compared to US children (Tempest, 1998); Canadian vs. US samples (Reynolds, Sanchez & Willson, 1996); and Indian children vs. UK children (Panicker, Hirisave & Subbakrishna, 2006). While differences in IQ scores between groups are not conclusive evidence of cultural bias (Jensen, 1980), the findings do raise the question that there may be major disadvantages to using tests that are normed in another culture. For example, the Raven's Progressive Matrices (Raven, Raven & Court, 2003), use visual stimuli that are commonplace in some cultures while non-existent in others (Benson, 2003). This is not a criticism of existing intelligence tests. Certainly, it would be an impractical undertaking to design a measure that will sample behaviours relevant to all cultures. Such an instrument at the very least will be cumbersome, with exorbitant development costs, and very long administration times. Publishers are unlikely to invest in such tools, opting instead for tests that are relevant to the culture within which the test will be distributed and used. The extent to which a culturally loaded test will be suitable for use in multiple contexts will depend on how well the core construct can be generalized cross culturally. Considering that no two cultures are identical, it would be difficult to conceive of a culture free test (Sternberg & Grigorenko, 2000), but it is not implausible to conceive of an intelligence test that is valid for use within a variety of cultural contexts.

When we choose to use an intelligence test, we are in fact endorsing the definition of intelligence that has been operationalized through the test. Furthermore, before importing a test for use, we must be certain that the measured construct can be generalized to the adoptive culture. American Educational Research Association (AERA) guidelines specify that validity of an imported psychological test should be ascertained before it is used in a target population (AERA, 1999). To establish the validity of a test, a thorough investigation of the central construct is recommended; however, explorations of internal validity are not sufficient. Neither are investigations of criterion or predictive validity. These components are said to reflect but a small subset of that broader network of diverse interrelated constructs (Cronbach & Meehl, 1953; Sternberg & Detterman; 1986; Lissitz, 2009). As prescribed by Flynn (1987) and Sternberg (2004), it is also necessary to explore possible environmental covariates that may explain individual differences in intelligence test scores.

Aetiology of Variability in Intelligence

Genetics and environment. For decades, twin studies have researched the relative impact of genetics and shared environment on variability in measured intelligence, by comparing the similarity in intelligence scores between identical twins and fraternal twins. The central argument is that identical siblings who have identical genomes will show greater similarity in measured intelligence than fraternal twins who share 50% of segregating (variable in humans) genetic information. This proposition is tested by calculating the hereditability coefficient which is the variation in the phenotype that can be attributed to genetics alone (Plomin, Owen & McGuffin, 1994). The same method is also used to estimate the influences of shared and individual-specific environments. Also, adoption studies are conducted to examine the contribution of the genotype of the biological parents, relative to the shared environment that is provided by adoptive parents and siblings (Plomin, Owen & McGuffin, 1994). Shared environment refers to the environment that siblings share within a household that is distinct from other households (Nisbett, Aronson, Blair, Dickens, et al., 2012). This is separate from non-shared (individual-specific) environment which refers to those aspects of environment that contribute to dissimilarities between siblings, such as personal activities, peers, birth place and even random measurement error (Nisbett et al., 2012).

Results demonstrate that a considerable portion of the variance in intelligence can be explained by genes. This finding is not surprising since most agree that all behaviours on which individuals differ are heritable to some extent (Nisbett, et al., 2012). Heritability of intelligence is said to be somewhere between .2 and .8 (Plomin & Spinath, 2002; Deary, Penke & Johnson, 2010; Nisbett et al., 2012). These values depend, in part, on the intelligence domain that is being investigated. For example, the literature indicates that general intelligence is more heritable than both specific ability factors and lower order skills (Kan, Wicherts, Dolan & Van der Maas, 2013). Moreover, hereditability values are said to increase with age (Plomin & Spinath, 2002; Deary, Penke & Johnson, 2010; Haworth, Wright, Luciano, Martin, et al., 2010; Kovas, Voronin, Kaydalov, Malykh, et al., 2013).

Another key finding of twin and adoption studies highlight the gene by environment interaction. Specific environmental variables are said to impact hereditability values associated with intelligence (Neisser, Boodoo, Bouchard & Boykin, 1996; Deary, Penke & Johnson, 2010). Hereditability coefficients were found to be of greater magnitude for groups of high socioeconomic status (SES) than for lower SES groups. Conversely, the contribution of environment to intelligence in lower SES groups has been shown to be greater than in high SES groups. As an example, analyses of SES by hereditability interactions in a sample of twins revealed that at the lowest SES level, shared environment accounted for effectively all of the variation in intelligence (Turkheimer, Haley, Waldron, D'Onofrio &Gottesman, 2003). The contribution of shared environment gradually decreased to zero as SES levels increased. Further evidence of the gene by shared environment interaction emerged from a reanalysis of data from a sample of 839 twin pairs (Harden, Turkheimer & Loehlin, 2007). Results demonstrated equal effects for genes and environment (40% to 40%) in families with the lowest income and lowest education. In the richer, more educated families, genes accounted for 50% of variance in IQ while shared environment accounted for 30%. In another study with 10-year-old twins, Deary, Whiteman, Starr, Whalley and Fox (2004) reported that SES was found to be a key moderating variable in the gene-environment relationship. The results demonstrated that heritability increased with SES from 5% in lower SES groups to 50% in higher SES groups. One clear implication of adoption and twin studies is that in higher SES groups, access to a greater number and quality of intellectually stimulating resources may better ensure that phenotypic expression of genetic intellectual potential is optimized (Nisbett et al., 2012). As such, the findings provide a solid rationale for implementing enrichment programmes as a way of stimulating intellectual development in children. Brooks-Gunn, McCarton, Casey and a team of researchers (1994) studied the outcome of 3 years of an intervention programme, including monitoring, enrolment in early childhood development centres, and parental training, on the cognitive development of premature babies. After three years, results revealed a positive effect for enrichment training on intelligence scores (Mc Carton, Brooks-Gunn, Wallace et al. 1997). Another study, the Carolina Abecedarian Project, compared intellectual ability scores of children who received enrichment strategies with those of a control group (Campbell & Ramey, 1994). Results reflected a statistically significant effect for environmental enrichment training immediately after termination of treatment and then 7 years post treatment.

The notion that environmental enrichment is associated with increased intelligence is expanded upon by Jensen (1997), but from the perspective of a gene by environment covariance. Jensen argues that individuals create enriched environments based on their unique needs and interests. Noting that the heritability of intelligence increases by age, Kan, Wicherts, Dolan and Van der Mass (2013) suggest that genetic potential is maximized because individuals of higher intelligence tend to seek out more intellectually stimulating environments – a process called active gene-environment correlation. Scarr and McCartney (1983) for example, spoke of niche picking which is a tendency for children to gravitate to activities that match their abilities, for example a child with strong verbal abilities may join the debating team, or develop a penchant for poetry. The notion is that as children get older they select environments, influence environmental changes, or actively create experiences that reflect their genetic potential (Scarr & McCartney, 1983; Kovas, Voronin, Kaydalov, Malyhk et al., 2013).

Conversely, some studies have identified a reversal in the direction of the gene/environment interaction (i.e. higher hereditability values for low SES groups; Ashbury, Wachs & Plomin, 2005). Such findings have been explained by an alternative framework, which explains that individuals with a genetic predisposition for low intellectual functioning are likely to be more sensitive to the effects of environmental stressors (Plomin & Rutter, 1998). Some data produced no evidence of interaction (Hanscombe, Trzaskowski, Haworth, Davis, et al. 2012; Kan et al., 2013).

In summary, twin and adoption studies on intelligence have demonstrated that environmental factors, specifically SES, play a role in moderating the extent to which an individual's genetic predispositions are expressed in the phenotype. These findings have been further bolstered by meta-analytic studies that have demonstrated as much as a 12 to 18-point increase in measured intelligence from lower SES homes to higher SES homes (e.g., siblings left with birth parents or children adopted by lower SES parents) (Locurto, 1990; van IJzendoorn, Jutter & Klein Poelhuis, 2005). Additionally, von Stumm and Plomin (2015) reported a 6-point difference in IQ scores between 2-year-old children of low and high SES, with the difference tripling by the time the children were 16 years old.

While the studies did not provide clarity on the mechanisms by which SES contributes to individual differences in intelligence, they have provided a reason to infer that interventions geared towards increasing the level of intellectual stimulation within an environment can have some effect on cognitive development in children. Findings have been largely supportive of this hypothesis. Indeed, intervention studies have demonstrated the positive effects of enrichment programmes on cognitive development in infants. Additionally, quasi experimental studies have revealed positive relationships between quality of home environment and intellectual test performance in children (e.g. Bradley & Casey, 1992; Lee & Barratt, 1993; Liaw & Brooks-Gunn, 1993).

DNA. What are the specific genetic determinants of intellectual ability? Over the years, researchers have identified different genetic markers (single-nucleotide polymorphisms; SNPs) associated with intelligence (Payton, 2009); however, recent studies were not successful at replicating a considerable number of these findings (Chabris, Herbert, Benjamin et al., 2012). Moreover, due to the complex and multi-faceted nature of intelligence, it has been difficult to identify the specific genes that account for individual differences in intelligence (see Payton, 2009; Johnson, Shkura, Langley, Delahaye - Duriez, et al., 2015) – an impossible endeavour without the benefit of very large samples sizes (Nisbett et al., 2012). A recent study by Johnson, Shkura, Langley, Delahaye-Duriez and colleagues (2015) identified 2 genetic networks (M1 and M3) as related to general cognitive ability. Another important finding in this area has been that the genes associated with intelligence are numerous but are of small effect, explaining no more than 0.2% of the variance in IQ scores (Rietveld, Medland, Derringer, Yang, et al., 2013; Plomin & Deary, 2014). Additionally, these intelligence-associated genes are pleiotropic meaning that they influence the development of more than one cognitive ability domain (Plomin & Deary, 2014).

Home environment. Family-related variables such as household income and parental education contribute to intellectual development in children (Mercy & Steelman, 1982). Additionally, environmental enrichment and orderliness have been shown to mediate the relationship between SES and intellectual development (Asbury, Wachs, & Plomin, 2005; Nisbett, et al, 2012). Access to new knowledge via parents or environmental exploration has been associated with intellectual gains in children (Gottfried & Gottfried, 1984; Hart & Risley, 1992). By contrast, children from impoverished environments tend to lack the resources necessary for intellectual growth (Neiss & Rowe, 2000; Petrill, Pike, Price, & Plomin, 2004), and thus can demonstrate intelligence scores as much as 13 points lower than their richer counterparts (Nisbett et al. 2012).

A study on the relationship between IQ and environmental enrichment took place in Romania where consequent to the revolution, the dire developmental and intellectual effects associated with child institutionalization became widely known. At the time, children in residential care suffered from general neglect, rigidity of structure, and a lack of emotional and verbal interaction with their caregivers (Nelson, Zeanah, Fox, Marshall et al., 2007). In this study, the researchers studied the impact of foster care placements on the intellectual functioning of a randomly selected group of infants (Zeanah, Nelson, Fox, Smyke et al., 2003). Measures of intellectual ability revealed that children who were raised in carefully selected and monitored home environments showed average scores of 30+ IQ points greater than the children who remained institutionalized (Nelson et al., 2007).

Research has also shown positive associations between household stability and measured intelligence. Such environments are typically peaceful and quiet (Petrill, Pike, Price & Plomin, 2004). Higher SES settings are argued to be more stable than lower SES surroundings, and this orderliness is proposed to explain a significant proportion of variance in measured intelligence (Petrill, Pike, Price & Plomin, 2004). Furthermore, environmental confusion or chaos (identified as noise, crowding and traffic within the home) has been shown to act as a mediator between SES and lowered intelligence scores (Hart, Pertrill, Deckard & Thompson, 2008).

Parental education and income. The quality of home and family environment is highly dependent on the parent's education and income (Lemos, Almeida & Colom, 2011). What a child learns from their environment depends on the parent's fund of general knowledge and ability to provide enriching experiences (Ceci & Williams, 1997). In other words, in order to facilitate intellectual development in their children, parents must have a sense of what is there to be learnt, and must also have the financial means to provide such exposure. Studies have pointed to a modest relationship between parent education and income, and the child's intellectual levels (Neiss & Rowe, 2000, Ganzach, 2014). Furthermore, genetic vs. shared environment studies demonstrated that the genetic potential for vocabulary development was better expressed in children raised by highly educated parents (Rowe, Jacobson & Van den Oord, 1999). Another study reflected a 19-point difference in IQ between teenagers whose parents had the lowest educational levels and those whose parents were highly educated (Lemos, Almeida & Colom, 2011). Less striking but statistically significant positive correlations were obtained between income and intelligence. Lemos, Almeida and Colom (2011) found that the intelligence scores of low income children were about 7 points below that of children from wealthier families; however, the effect for income was reduced after controlling for parent education.

Schooling. Whereas crystallized ability is argued to be gained through reasoning over time (Cattell, 1971/1987), both declarative knowledge and reasoning ability are hypothesized to be partially influenced by direct instruction (Gottfredson & Saklofske, 2009; Lohman & Lakin, 2009). Therefore, a child's intellectual development may be partly explained by the richness of their formal and informal educational environments **(**Cahan & Cohen, 1989; Rinderman, Flores-Mendoza & Mansur-Alves, 2010). Neisser and colleagues (1996) and Jensen (1997) assert that formal education imparts a broad range of knowledge that promotes the development of intellectual skills. This notion was supported by Barber (2005) who indicated that secondary school enrolment accounted for a significant portion of cross-national variation in IQ scores. One study showed that years of schooling, over and above age, predicted scores on intelligence tests (Cahan & Cohen, 1989). Also, in a Turkish sample, education accounted for a large and significant portion of the variance in cognitive ability test performance (Kudiaki & Alsan, 2008). A study by Ceci (1991) compared IQ scores of same age peers who enrolled in school one year apart. Results demonstrated 5pt higher IQ scores in the children who were enrolled one year earlier. Children's IQ scores have been also shown to decline slightly over long summer breaks (Ceci, 1991). Furthermore, children who drop out of school demonstrate IQ declines of approximately 2pts for every year of high school missed (Ceci & Williams, 1997).

The Outcomes of Intelligence

The previous section explored some of the major contributors to individual differences in intellectual ability. To conclude the discussion, this section will focus on the relationship between intelligence and a number of important life outcomes including occupational success, criminality, health and longevity and academic achievement.

The search for a predictive model of achievement began in the late 19th century with the work of Sir Francis Galton, who theorized that human intellectual ability played a principal role in the prediction of future life success (Reeve & Bonaccio, 2011). In subsequent years, studies have identified a small but significant relationship between intelligence and future health and longevity outcomes (Gottfredson & Deary, 2004; Wraw, Deary, Gale & Der, 2015). Tong, Baghurst, Vimpani and McMichael (2007) suggest that intelligence predicts socioeconomic status and therefore access to a better quality of health care. They also argue that intelligent people are more likely be aware of the consequences of bad health practices and therefore avoid harmful situations. Small but negative correlations between intelligence and criminality have also been discovered (Moffitt, Gabrielli, Mednick & Schulsinger, 1981; Neisser et al., 1996). Neisser et al. (1996) theorize that education acts as a mediator in the relationship between intelligence and criminality. They explain that children of low intellectual ability are less likely to succeed in school, thereby making them more vulnerable to criminal influences.

Compelling findings relate to the associations between intelligence and occupational success, income, and academic achievement. General intelligence is significantly associated with occupational achievement and job performance (Gottfredson, 1997b), with correlations ranging between $r = .30$ and .50 (Neisser et al., 1996). In a meta-analysis of longitudinal studies that explored the relationship between intelligence and a variety of outcomes, Strenze (2007) found that measured intelligence accounted for 43% and 20% of occupational success and income respectively. Schmidt and Hunter (2004) reported that intelligence correlated strongly ($r = .51$) with job performance. It has been suggested that employees with higher intelligence are quicker at learning their jobs and are therefore more likely to excel and be rewarded for their successes (Hunter 1986). With regard to income, measured intelligence is said to explain about 16% of the variance in salary level (Neisser et al., 1996). It has been argued that socioeconomic status acts as a mediator between intellectual levels and future income (e.g. Dubow, Boxer & Huesmann, 2009). However, Nettle (2003) observed an association between measured intelligence and socioeconomic status, with stronger correlations between SES and intelligence in adults than in children, suggesting that intelligence has some explanatory value over and above parental income in determining future income. Moreover, Strenze (2007) revealed that strong predictors of occupational success are academic performance and a combination of socioeconomic indicators – parental education, parental occupation, and family income, but the predictive power of these variables did not exceed that of intelligence.

It is important to note that while IQ finds empirical support as a key contributor to general occupational success, it is not considered to be a major contributor (Nettle, 2003). Education is considered to mediate the relationship between IQ and occupational achievement. This is not surprising because educational background is a key criterion for hiring and also promotion. Additionally, the quality of job relevant knowledge gained prior to employment is probably related to quality of education and is likely to impact job performance.

Among all life outcomes, intelligence explains the highest amount of variance in education (Strenze, 2007). While intelligence has been associated with a variety of outcomes beyond academic performance including health behaviour, job performance and psychological well-being (Hunter & Hunter, 1984; Jensen, 1997; Gottfredson & Saflofske, 2009), it has been argued that life success is in fact a function of education because expertise and general knowledge gained over time increases potential for a multitude of successful life outcomes (Gottfredson, 1997b).

The earliest empirical attempts at investigating the predictive validity of intelligence focused primarily on correlations between intelligence and education. Spearman (1904) found that teachers’ estimates of intelligence correlated significantly with school exam results. Binet (1905) introduced the first intelligence quotient (IQ) tests to identify children who would not likely succeed in mainstream education.

More recently, using data from a large sample of 11-year-old children, Deary, Strand, Smith and Fernandes (2007) found that educational attainment 5 years after intelligence tests were administered correlated highly with measured intelligence (*r* =.81). Typical correlation values between measured intelligence and highest level of educational attainment or by school grades tend to be around 0.54 (Deary & Johnson, 2010). Other data have provided persistent and compelling support for a link between academic achievement and cognitive ability (e.g., Neisser et al., 1996; Jensen, 1997; Gagné & St. Père, 2002; Deary, Strand, Smith, & Fernandes, 2007; Laidra, Pullmann, & Allik, 2007). As an example, meta-analytic studies by Walberg (1984) and later by Gagné and St. Père (2002) have established correlations between cognitive ability and academic performance of .70 and .60 respectively. Strenze (2007) found that measured intelligence accounted for 56% of the variance in academic performance. Also, results of correlational analyses reflected a strong statistically significant relationship between global IQ and all measures of academic achievement (Sattler, 2001; Rohde & Thompson, 2007; Freberg, Vandiver, Watkins & Canivez, 2008). With specific reference to performance on the GCSE examinations, it was found that between 40 and 65 percent of the heritability of academic achievement in this area is explained by intelligence (Krapohl, Rimsfeld, Shakeshaft, Trzaskowski, et al., 2014; Rimfeld, Kovas, Dale & Plomin, 2015).

Conclusion

This chapter explored the historical and contemporary definitions of intelligence and how these paradigms have influenced the design of modern day tests. Also discussed, were the contributions of twin, adoption, intervention and correlational studies to the broad model of intelligence. This chapter attempted to demonstrate that in spite of empirical support for current models, there remains much controversy about the actual definition of intelligence itself. Questions about the nature of intelligence persist as does the debate over what intelligence tests really measure. Issues about the cultural specificity of intelligence and the inability of test developers to construct a universal measure of intelligence were discussed. The chapter argued that because intelligence models are broader than its specific and global abilities, studies of validity must not ignore the network of environmental covariates that give additional meaning to the core construct.

Against this backdrop, this thesis proposes to investigate the cross-cultural validity of the Wechsler Intelligence Scale for Children – 4th Edition (United States version) (WISC-IV (US)) in Trinidad and Tobago (T&T) children. The WISC-IV (US) is an instrument that was designed to assess the cognitive ability of children. It was adopted for use in T&T in 2003. Despite recommendations of the AERA (1999) that all imported tests should be assessed for validity before use, no such study on the WISC-IV (US) has been conducted in T&T, where it has been in use for the past 12 years. It is therefore the aim of this thesis to assess whether the WISC-IV (US) can be considered to provide a reliable and valid estimate of intelligence in T&T children. The thesis also aims to generate an interpretive framework by which the scores of T&T children can be understood.

The following chapter will provide a description of the WISC-IV(US) – its development, content, psychometric properties and uses. The chapter will also provide an overview of its uses in Trinidad and Tobago, as well as the issues that have been raised in regard to the measurement of intelligence in the country. The chapter will then describe the specific research aims of this thesis.

CHAPTER 2

THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN – 4TH EDITION (US) (WISC – IV(US))

The WISC-IV (US) is an individually administered instrument that is used to assess the cognitive ability of children aged 6 to 16 years. The first of the series, the Wechsler Intelligence Scale for Children (WISC), was designed in 1949 specifically for American children aged 5 to 15 years (Flanagan & Kaufman, 2004). Since then there have been 4 further editions. The WISC-IV (US) is the 3rd revision of the test (Wechsler, 2003a). The most recent edition of the test, the WISC-V, was published in 2014.

The WISC-IV (US) is widely used in educational and clinical settings, and contributes towards the identification of intellectual, developmental, behavioural and learning disorders in children (Prifitera, Weiss, Saflokse & Roflhus, 2005). It consists of ten core subtests (Similarities, Vocabulary, Comprehension, Block Design, Picture Concepts, Matrix Reasoning, Digit Span, Letter Number Sequencing, Coding and Symbol Search) and five supplemental subtests (Picture Completion, Cancellation, Information, Arithmetic and Word Reasoning) (Wechsler, 2003b). These subtests are organized into four ability indices (Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI)) and one general composite (Full Scale IQ (FSIQ); Wechsler, 2003b) (See Figure 2.1). The following sections will provide a detailed description of the 10 core subtests that will be examined in this thesis.

WISC-IV Indices	**WISC-IV Core Subtests**	**WISC-IV Supplemental Subtests**
Verbal Comprehension	Similarities	Information
	Vocabulary	Word Reasoning
	Comprehension	
Perceptual Reasoning	Block Design	Picture Completion
	Picture Concepts	
	Matrix Reasoning	
Working Memory	Digit Span	Arithmetic
	Letter Number Sequencing	
Processing Speed	Coding	Cancellation
	Symbol Search	

Figure 2.1. Summary of the WISC-IV core and supplemental subtests from the WISC-IV Technical and Interpretive Manual (Wechsler, 2003b).

The WISC-IV (US) core subtests

Similarities (23 items). This subtest measures abstract reasoning ability and verbal concept formation (Wechsler, 2003b). Test items require examinees to describe how two words are similar (Flanagan & Kaufman, 2004). The Similarities subtest begins with basic verbal reasoning items and ends with items that measure verbal concept formation. Examples of items similar to those found on the Similarities subtest are as follows:

"How are 50 pence and one pound alike?".

"How are happiness and sadness alike?".

This test also measures long term memory, auditory comprehension and expressive language skills (Wechsler, 2003b).

Vocabulary (36 items). Items in the Vocabulary subtest require the examinee to provide the names of pictures and the definitions of words. This test measures word knowledge, verbal concept formation, expressive language, long term memory and fund of information (Wechsler, 2003b). Vocabulary has the highest *g* loading of the 10 core subtests (Prifitera, Weiss, Saflokse & Roflhus, 2005).

Comprehension (21 items). This measure of social knowledge, reasoning and judgement tests the examinee's knowledge of general social concepts (Flanagan & Kaufman, 2004). This is an example of an item similar to those found in the Comprehension subtest:

> *"Why should a person ask permission before borrowing or taking something that belongs to someone else?".*

Questions like the following may also ask the child to comment on social problem solving situations (Wechsler, 2003b):

> *"What should you do if you hear someone screaming for help?".*

This subtest also measures expressive language skills and comprehension (Wechsler, 2003b).

Block Design (14 items). In this subtest, the examinee is presented with multi-coloured blocks and asked to replicate 2 or 3 dimensional models of geometric designs (Wechsler, 2003c; Figure 2.2). Block design tasks measure ability to analyse abstract visual stimuli, visual motor coordination, spatial awareness and organization (Wechsler, 2003b).

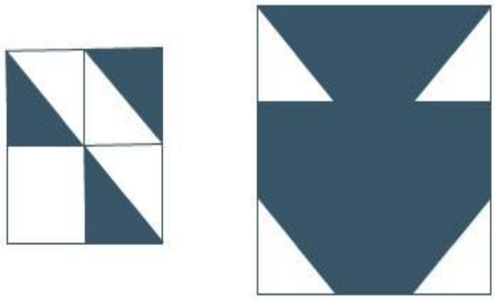


Figure 2.2. 2D geometric designs similar to those found in the Block Design subtest. Adapted from the WISC-IV Stimulus Booklet (Wechsler, 2003c).

Picture Concepts (28 items). This test presents the examinee with either 2 or 3 rows of pictures. They are then asked to select two pictures that go together (Wechsler, 2003b; Figure 2.3). These items measure nonverbal categorical reasoning ability.

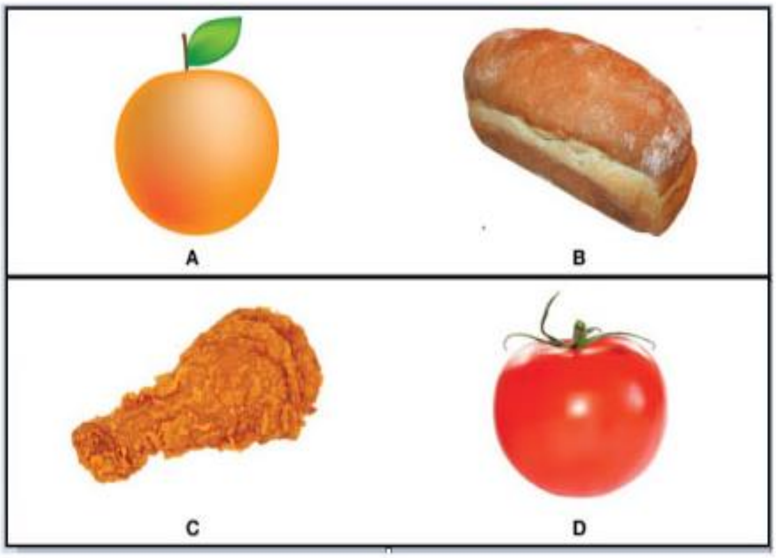


Figure 2.3. Item similar to those found in the Picture Concepts subtest. Adapted from the WISC-IV Stimulus Booklet (Wechsler, 2003c).

Matrix Reasoning (35 items). On this nonverbal reasoning task, examinees are required to select which of five options completes a matrix puzzle (Flanagan & Kaufman, 2004). Matrices items feature recognizable objects as well as abstract shapes and designs (Wechsler, 2003c; Figure. 2.4). This test measures non-verbal abstract problem solving and deductive reasoning ability (Wechsler, 2003b).

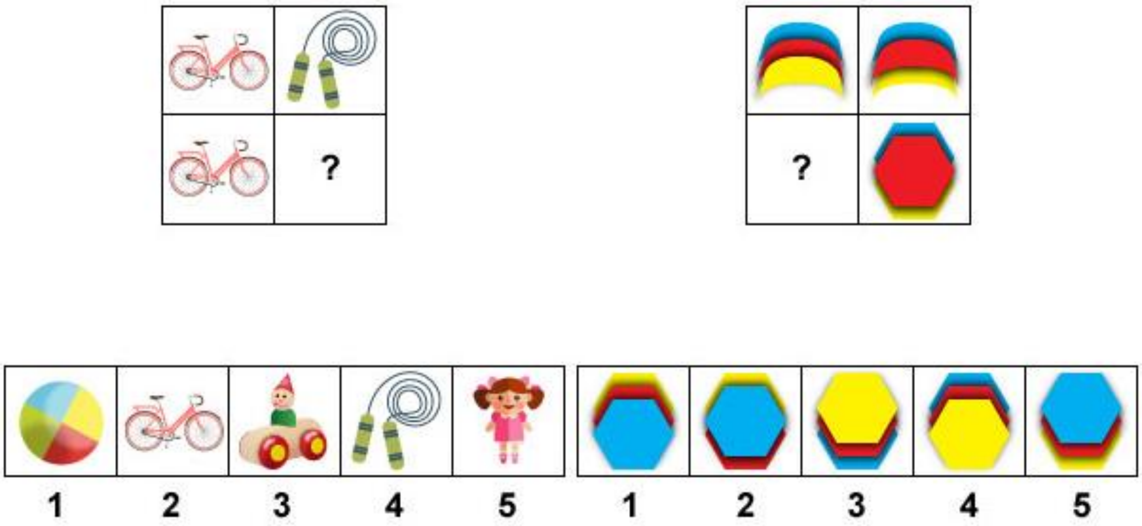


Figure 2.4. Items similar to those found in the Matrix Reasoning subtest. Adapted from the WISC-IV Stimulus Booklet (Wechsler, 2003c).

Digit Span (16 items). This test of working memory is composed of two different kinds of tasks: Digit-Span Forward and Digit-Span Backward (Flanagan & Kaufman, 2004). For both tasks, the examinee listens to number lists of increasing length and is then required to repeat them (Wechsler, 2003b). In Digit-Span Forward the examinee is prompted to repeat the numbers verbatim; but in Digit-Span Backward, the numbers must be repeated backwards (Wechsler, 2003b; Flanagan & Kaufman, 2004). This test measures short term memory, attention, and ability to encode, mentally manipulate and recall auditory information (Wechsler, 2003b).

Letter Number Sequencing (10 items). This task measures attention, concentration, mental manipulation of auditory information and short term memory (Wechsler, 2003b). Examinees listen to lists of randomly presented letters and numbers and first must repeat the numbers in numerical order followed by the letters in alphabetical order (Flanagan & Kaufman, 2004).

Coding. In this subtest, the examinee is shown a key of numbers with their corresponding symbols. Below the key are rows with numbers, but their symbols are missing (Wechsler, 2003b; Figure 2.5). The examinee is then given two minutes to draw the missing symbols below their respective numbers (Flanagan & Kaufman, 2004). Coding measures processing speed, short term retention of visual information, learning ability, attention and visual motor integration (Wechsler, 2003b).

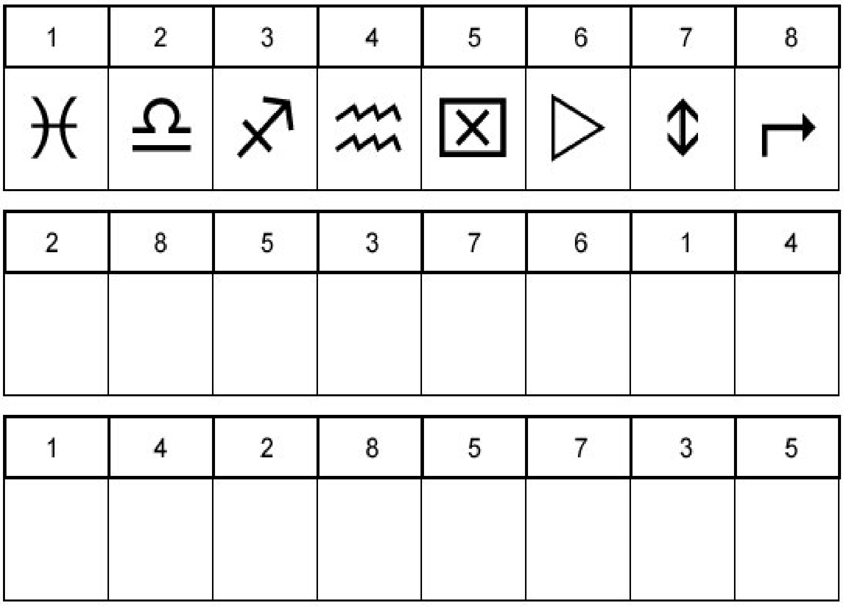


Figure 2.5. Items similar to those found in the Coding subtest. Adapted from the WISC-IV Scoring and Administration Manual (Wechsler, 2003d).

Symbol Search. In this measure of processing speed, each item requires the examinee to search a group of shapes for a target symbol (Flanagan & Kaufman, 2004; Figure 2.6). Examinees must indicate, by checking 'Yes' or 'No', whether or not they see the target symbol within the group ((Flanagan & Kaufman, 2004). Two minutes is the maximum time allowed for completing this task. This test measures processing speed, concentration, visual tracking, visual discrimination and visual motor coordination (Wechsler, 2003b).

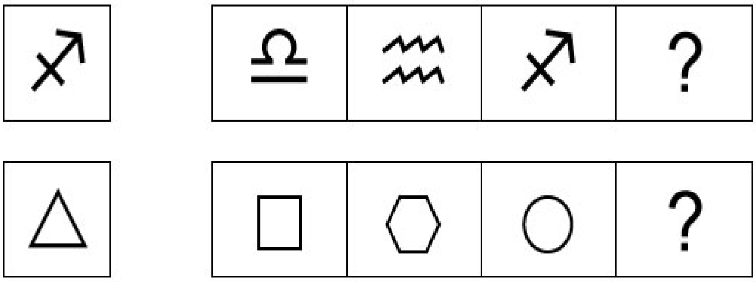


Figure 2.6. Items similar to those found in the Symbol Search subtest. Adapted from the WISC-IV Scoring and Administration Manual (Wechsler, 2003d).

Scoring guidelines for the WISC –IV (US) subtests and composites. WISC-IV (US) composite scores are generally calculated from the 10 subtest scores, but occasionally supplemental tests can be administered as a replacement for core subtests (Wechsler, 2003b). By using the WISC-IV (US) norms tables, each subtest raw score is converted to a subtest scaled score, which ranges from 1 to 19 (M = 10, SD = 3) (Wechsler, 2003b). Within each ability index, the subtest scaled scores are added together before using the conversion tables in the WISC-IV test administration manual (Wechsler, 2003b) to convert them to a composite score. The conversion tables are also used to convert the sum of all the scaled scores to the global IQ score (FSIQ) (Wechsler, 2003b). Index and global IQ scores range from 40 to 160 (M = 100, SD = 15) (Wechsler, 2003b).

Psychometric Properties of the WISC-IV (US)

After item development, revisions and pilot studies, the standardization version of the WISC-IV (US) was used to obtain information about the psychometric properties of the instrument. The test was administered to a stratified sample of 2200 children of ages 6 to 16 years. Sample details are summarized in Chapter 3. Subsamples were also administered additional cognitive ability, adaptive behaviour, memory, personality and academic achievement tests in order to investigate the external validity of the WISC-IV (US).

Reliability. The split half method was conducted for the verbal comprehension, perceptual reasoning and working memory subtests and composites. Test-retest reliability analyses were performed for all subtests and composites. Overall results revealed strong evidence of good reliability. Internal consistency reliability coefficients ranged from .79 to .90 for the subtests, and .88 to .97 for the composites (Wechsler, 2003b). Test- retest reliability coefficients reflected acceptable stability over time (mean interval of 4 weeks between administrations), from .92 for Vocabulary to between .70 and .80 for the remaining tests (Wechsler, 2003b). The FSIQ score demonstrated the most stability of the WISC-IV composites and subtests (Wechsler, 2003b). See Table 2.1 for a summary of split half and test-retest reliability coefficients.

Table 2.1.

Split-Half and Test-Retest Reliability Coefficients (corrected *r*) for WISC-IV (US) Subtests and Composites (Wechsler, 2003b).

	Split-Half Reliability[1]	Test-Retest Reliability[2]
Block Design	.86	.82
Similarities	.86	.86
Digit Span	.87	.83
Picture Concepts	.82	.76
Coding	.85 (estimate)[3]	.84
Vocabulary	.89	.92
Letter/Number Sequencing	.90	.83
Matrix Reasoning	.89	.85
Comprehension	.81	.82
Symbol Search	.79 (estimate)	.80
Picture Completion	.84	.84
Cancellation	.79 (estimate)	.79
Information	.86	.89
Arithmetic	.88	.79
Word Reasoning	.80	.82
Verbal Comprehension	.94	.93
Perceptual Reasoning	.92	.89
Working Memory	.92	.89
Processing Speed	.88 (estimate)	.86
Full Scale IQ	.97	.93

The reliability of the WISC-IV (US) was also investigated in subsequent studies. Ryan, Glass and Bartels (2010) found that reliability coefficients obtained for a sample of 43 elementary students who were retested after 11 months ranged from .26 for Picture Concepts to .81 for Vocabulary and .54 for Processing Speed to .88 for Full Scale IQ. Overall the findings reflected smaller reliability coefficient values compared to the normative sample. Also, because test-retest difference values demonstrated considerable

---

[1] Split-half reliability coefficients were calculated from normative sample data (N = 2200) (Wechsler, 2003b).

[2] Test-retest reliability coefficients were calculated from a subsample of the main dataset (N = 243) (Wechsler, 2003b).

[3] Because of the timed nature of the processing speed tasks, split half coefficients are not appropriate measures of reliability on the Coding, Symbol Search and Cancellation subtests. Instead split-half reliability values were estimated from the test-retest reliability data (Wechsler, 2003b).

variability, stability coefficient confidence intervals were large (e.g. Comprehension: r = .49, 95% CI (.22 to .69); PRI: r = .58, 95% CI (.34 to.75). The authors argued that the lowered stability values reflected a reduction of practice effects associated with a longer interval between test administrations (Ryan, Glass & Bartels, 2010), but results could also be attributed to the small sample size (Watkins & Smith, 2013). In a later study with a larger sample (N = 344), test-retest reliabilities after a 3-year interval ranged from .65 to .82 (Watkins & Smith, 2013). Stability values for this study generally were lower than those of the standardization study, with the exception of the FSIQ, which replicated the values of the Ryan, Glass and Bartels (2010) and the Wechsler (2003b) studies.

Internal validity.

Subtest inter-correlations. Evidence of the internal validity of the WISC-IV (US) was obtained through examinations of the correlational matrices of the 10 core subtests alone and then all 15 subtests. Results revealed statistically significant correlations among all WISC-IV (US) subtests (See Table 2.2). Additionally, within composite subtest correlations were higher than between composite subtest correlations (Wechsler, 2003b). An expanded discussion of the WISC-IV (US) inter-correlation study can be found in Chapter 5.

Factor structure. Wechsler (2003b) articulated a four 1st order factor structure which consisted of the four ability indices and their corresponding subtests (See Figure 2.7). To investigate this model, exploratory and confirmatory factor analyses were performed and they provided empirical support for the 4-factor model. The standardized factor loadings are presented in Table 2.3.

Table 2.2.

Correlational Matrix for WISC-IV subtests – WISC-IV Standardization Study (Wechsler, 2003)

		Sim	Voc	Com	Inf	WR	BD	PCon	MR	PComp	DSpan	LNS	Arith	Cd	SS
	Voc	.74													
	Com	.62	.68												
VCI	Inf	.70	.75	.62											
	WR	.62	.66	.58	.62										
	BD	.50	.48	.36	.48	.44									
	PCon	.45	.42	.40	.40	.41	.41								
PRI	MR	.49	.49	.42	.50	.11	.55	.47							
	PComp	.50	.51	.44	.50	.47	.54	.39	.46						
	DSpan	.39	.42	.36	.40	.35	.35	.30	.38	.26					
WMI	LNS	.47	.50	.43	.48	.42	.38	.36	.42	.34	.49				
	Arith	.56	.59	.52	.62	.48	.55	.42	.54	.41	.47	.51			
	Cd	.28	.31	.30	.31	.25	.36	.29	.34	.44	.23	.30	.36		
PSI	SS	.39	.38	.34	.38	.35	.45	.34	.42	.37	.30	.40	.43	.53	
	Can	.16	.14	.11	.11	.13	.19	.14	.14	.14	.10	.11	.17	.40	.32

Note. Sim = Similarities, Voc = Vocabulary, Com = Comprehension, Inf = Information, WR = Word Reasoning, BD = Block Design, PCon = Picture Concepts, MR= Matrix Reasoning, PComp = Picture Completion, DSpan = Digit Span, LNS = Letter Number Sequencing, Arith = Arithmetic, Cd = Coding, SS = Symbol Search, Can = Cancellation, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index.

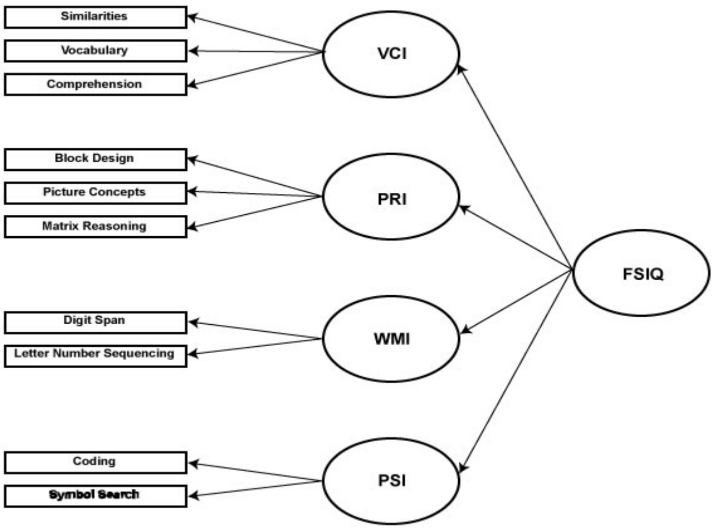


Figure 2.7. WISC-IV (US) current 4 factor structure (Wechsler, 2003b).
Note: VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index, FSIQ = Full Scale IQ.

Table 2.3.

Standardized Loadings of Core Subtests on WISC-IV Composites (All Ages)

(Wechsler, 2003)

	VCI	PRI	WMI	PSI
Similarities	**.74**	.19	-.03	-.06
Vocabulary	**.84**	.02	.03	-.02
Comprehension	**.78**	-.11	.03	.08
Block Design	.01	**.66**	-.02	.08
Picture Concepts	.13	**.45**	.03	.03
Matrix Reasoning	.00	**.69**	.06	.01
Digit Span	.00	.07	**.62**	-.06
Letter-Number Sequencing	.09	-.02	**.62**	.06
Coding	.02	-.01	-.04	**.68**
Symbol Search	-.01	.09	.04	**.65**

Note. VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index.

Additionally, analyses of all 15 subtests also supported the proposed 4 factor model. Wechsler (2003b) also conducted confirmatory factor analyses to compare the 4 factor model with three alternative models as follows:

Model 1: One factor (a general factor measured by the 10 core subtests).

Model 2: Two factors (3 Verbal subtests and 2 Working Memory subtests on the 1st factor, and 3 Perceptual Reasoning and 2 Processing Speed subtests on the 2nd factor).

Model 3: Three factors (3 Verbal subtests on the 1st factor, 3 Perceptual Reasoning subtests on the 2nd factor and 2 Working Memory and 2 Processing Speed subtests on the 3rd factor.

Using indices that are used to assess how well the proposed model fits the data, the 4-factor model was found to demonstrate better fit than any of the three alternative models.

Beyond Wechsler's standardization research, other studies provided support for and additional details about the WISC-IV (US) factor structure (e.g. Flanagan & Kaufman, 2004; Keith, 2005; Keith, Fine, Taub, Reynolds & Kranzler, 2006; Watkins, Wilson, Kotz, Carbone & Babula, 2006; Bodin, Pardini, Burns & Stevens, 2009; Canivez, 2014). Some results revealed consistencies with the CHC model and cross loadings for some subtests. The work of Keith and colleagues (2006) provided a more empirically sound and broader interpretive framework for the WISC-IV. Figure 2.8 summarizes how the WISC-IV (US) core subtests are organized within the CHC framework.

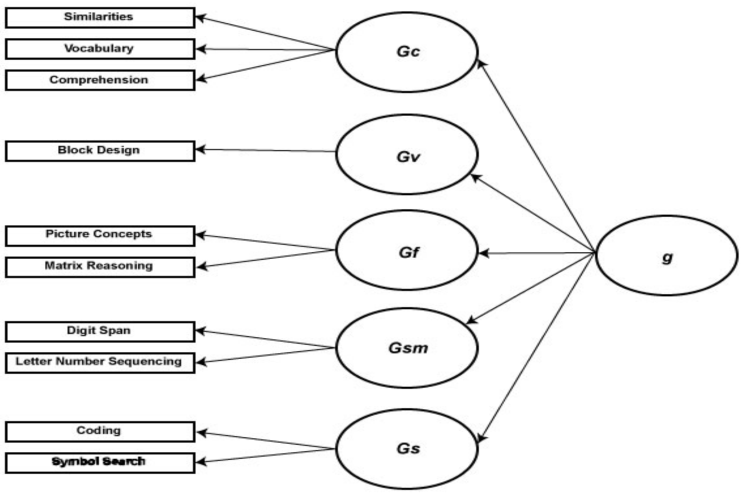


Figure 2.8. WISC-IV (US) core subtests within the CHC framework based on the Keith et al. (2006) study. Note: *Gc* = Crystallized ability, *Gv* = Visual Processing ability, *Gf* = Fluid Reasoning ability, *Gsm* = Short Term Memory, *Gs* = Processing Speed, *g* = General Ability.

From studies of the 15 supplemental subtests, Keith et al. (2006) had identified Block Design as a visual processing factor and Arithmetic as a measure of fluid reasoning. The Similarities subtest was identified as a measure of crystallized ability (*Gc)* *(*Keith et al, 2006) while Flanagan and Kaufman (2004) classified the test as a measure of fluid reasoning ability (*Gf*). Other studies identified Matrix Reasoning as both a *Gf* and visual processing measure (*Gv*) (Carroll, 1993; Lecerf, Rossier, Favez, Reverte & Coleaux, 2010). Another interesting result has been that the Matrix Reasoning task also measures working memory (Salthouse, 1992). Support for Salthouse's findings can be found in studies demonstrating a link between fluid reasoning ability and working memory (Kyllonen, 1994; Fry & Hale, 1996; Engle, Tuholski, Laughlin, & Conway, 1999, Süß, Oberauer, Wittmann, Wilhelm & Schulze 2002; Colom, Rebollo, Palacios, Juan-Espinosa & Kyllonen, 2004).

External validity. Investigations into the external validity of the WISC-IV (US) were done using correlational studies to determine the relationship between WISC-IV

(US) scores and scores on measures of similar or related constructs. Correlations between the WISC-IV (US) and concurrent tests of cognitive ability, academic achievement, memory, adaptive behaviour, intellectual giftedness and emotional intelligence were calculated.

Results demonstrated statistically significant positive correlations between WISC-IV (US) measured intelligence and other Wechsler intelligence scales, such as the WISC –III (Wechsler, 1991) and the Wechsler Preschool and Primary Scales of Intelligence (WPPSI, Wechsler, 2002). Significant positive correlations between WISC-IV (US) scores and memory, giftedness and academic achievement were also obtained. Conversely, there were non-significant correlations between WISC-IV scores and emotional quotient scores (Wechsler, 2003b). The results are summarized in Table 2.4.

Table 2.4.

Correlations between WISC-IV (US) Composite scores and Criterion Measures (Wechsler, 2003b)

	WISC III	WPPSI III	WASI	WIAT II (TA)	CMS (GM)	GRS (IA)	BarOn EQ (TEQ)	ABAS II (GAC)
VCI	.87	.83	.85	.80	.52	.52	.22	.39
PRI	.74	.79	.78	.71	.46	.48	.29	.30
WMI	.72	-	-	.71	.52	.47	.23	.38
PSI	.81	.65	-	.58	.29	.37	.24	.23
FSIQ	.89	.89	.83	.87	.61	.60	.31	.41

Note. VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index, FSIQ = Full Scale IQ, WISC III = Wechsler Intelligence Scale for Children – 3rd Edition, WPPSI III = Wecshler Preschool and Primary Scale of Intelligence – 3rd Edition, WASI = Wechsler Abbreviated Scale of Intelligence, WIAT II (TA) = Wechsler Individual Achievement Test – 2nd Edition (Total Achievement), CMS (GM) = Children’s Memory Scale (General Memory), GRS (IA) = Gifted Rating Scale (Intellectual Ability), BarOn EQ (TEQ) = BarOn Emotional Quotient Inventory (Total Emotional Quotient), ABAS II (GAC) = Adaptive Behaviour Assessment System – 2nd Edition (General Adaptive Composite)

The Cross Cultural Validity of the WISC-IV (US)

Comparisons of IQ test performance among US racial groups have demonstrated that ‘Black’ children typically score about 15 points, and ‘Hispanic’ children – about 9 points lower than ‘White’ children (Weiss, 2003). Over the years, studies on the WISC scales have replicated these findings (e.g. Munford & Munoz, 1980; Arinoldo, 1981; Weiss, 2003). Moreover, after the publication of the WISC-R, a clinical sample of ‘Black’ children demonstrated notably lower mean IQ scores compared to their performance on the WISC (Munford, 1978, Munford, Meyerowitz & Munford, 1980). Examination of the performance of a similar sample of ‘White’ children did not reveal similar decreases in IQ score across tests (Munford, Meyerowitz & Munford,1980). To Munford and colleagues, these discrepancies exposed problems of bias within the Wechsler scales.

There may be some merit to the argument that the WISC-R functions differentially across groups; however, a more thorough exploration of the evidence may weaken this conclusion. First the data actually shows significantly weakened performance for both ‘Black’ and ‘White’ samples on the Similarities and Coding subtests. Even though the difference scores and significance values for the 'Black' sample were larger than the 'White' sample; without effect sizes, it is difficult to determine the relative importance of these inter-test differences, or if these differences are important at all. Also, the authors failed to indicate how similar the 'Black' and 'White' samples were on variables that were likely to impact performance on the tests such as parental education, income or education level. One cannot confidently ascribe cultural bias to the WISC-R without this information.

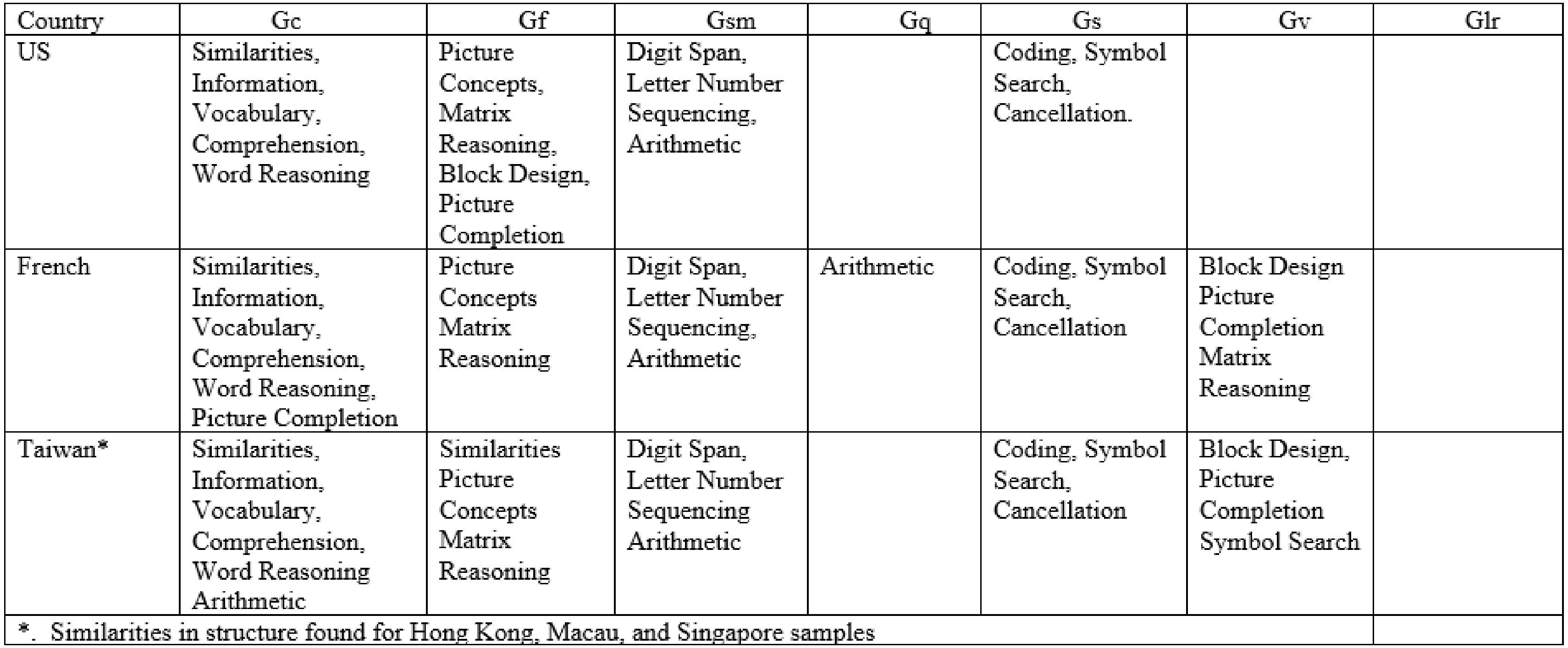
Other researchers also considered observed differences in mean scores to be insufficient indicators of test bias (e.g. Wechsler, 1971, Sandoval, 1979, Ross-Reynolds

& Reschly, 1983). For example, Wechsler wrote a paper in defense of the WISC's cross-cultural validity by referring to the rigorous item bias analyses and reviews that preceded publication of the test (Wechsler, 1971). Sandoval (1979) showed that more thorough investigations of the internal and external properties of the WISC-R actually revealed greater between-group similarities than differences. Ross-Reynolds and Reschly (1983) compared the psychometric properties of the WISC-R among 'Anglo', 'Black', 'Chicano' and indigenous 'Papago' children and found little or no evidence of bias, except in the case of Native American children, for whom the data had produced 'ambiguous' results. Additionally, studies of predictive validity showed that WISC-R and WISC-III scores predicted academic achievement equally well for 'White' children and children of minority groups (Poteat, Wuensch, & Gregg, 1988; Weiss, Prifitera & Roid, 1993; Weiss & Prifitera, 1995). Wechsler (1971) asserted that any racial differences in IQ should not be blamed on the test, but on the environmental disadvantages that have been suffered by minority groups within the US (Prifitera, Weiss, Saflofske & Rolfhus, 2005). Indeed, environmental factors such as income and parental education have demonstrated associations with WISC IQ score differences between groups (Prifitera, Weiss, Saflofske & Rolfhus, 2005). One study used regression analyses to determine how the variance in FSIQ performance that is accounted for by race is impacted by adding key environmental variables to the regression equation (Weiss, Harris, Prifitera, Courville et al., 2006). Results showed that by adding Parental Education, the variance accounted for by race decreased from 4.7% to 2.6%. Also when, income was introduced, the variance attributed to race decreased further to 1.6%.

Further afield, gathering of evidence to investigate the 'cultural fairness' of the WISC extended to international samples. For example, an international study undertaken with the WISC-III sought to determine whether the construct of intelligence upon which the test was designed could be replicated across 12 countries (Georgas, Weiss, van de Vijver & Saklofske, 2003). Results revealed adequacy of fit of the four factor model in all 12 countries including Canada, US, UK, Germany, France, Greece, Sweden, Netherlands, Slovenia, South Korea, Lithuania and Japan (Georgas, Weiss, van de Vijver & Saklofske, 2003). Other international structural invariance studies have also provided support for the factor structure of the WISC-IV (Daseking, Petermann & Petermann, 2007; Chen, Keith, Chen & Chang, 2009; Chen, Keith, Weiss, Zhu & Li, 2010; Fina, Sanchez-Escobedo & Hollingworth, 2010; Lecerf, Rossier, Favez; Reverte & Coleaux, 2010; Nakano, 2011; Watkins, Canivez, James, James & Good, 2013; Reverte, Golay, Favez, Rossier, & Lecerf, 2014/2015) (See Figure 2.7). Cross-cultural structural invariance was also found for the CHC based structure. For example, analyses of the structure of the Taiwanese version of the WISC-IV (Chen, Keith, Chen & Chang, 2009) revealed good fit to the initial four factor structure, however, examination of fit indices revealed better model fit for the broader CHC model. The Similarities, Symbol Search, Matrix Reasoning and Arithmetic subtests demonstrated significant cross-loadings on more than one specific ability area (Chen, Keith, Chen & Chang, 2009) (See Table 2.5 for a summary of these findings). Exploratory factor analyses of the WISC-IV in a sample of French children revealed that Block Design loaded less on fluid reasoning than on visual processing and processing speed (Lecerf et al. 2010, Table 2.5). Similar explorations sought to determine the fit of the French WISC-IV in French-speaking Swiss children (Reverte, Golay, Favez, Rossier, & Lecerf, 2014). Results demonstrated that the Arithmetic test showed up with multiple cross loadings on the Working memory, Verbal Comprehension and Perceptual Reasoning indices. In terms of the CHC model, Arithmetic showed up equally well as a measure of crystallized ability/short-term memory or fluid reasoning/short term memory.

Table 2.5.

WISC-IV factor structure by country (Wechsler, 2003, Wechsler, 2004, Chen et. al, 2009, Lecerf et. al., 2010,)



Country	Gc	Gf	Gsm	Gq	Gs	Gv	Glr
US	Similarities, Information, Vocabulary, Comprehension, Word Reasoning	Picture Concepts, Matrix Reasoning, Block Design, Picture Completion	Digit Span, Letter Number Sequencing, Arithmetic		Coding, Symbol Search, Cancellation.		
French	Similarities, Information, Vocabulary, Comprehension, Word Reasoning, Picture Completion	Picture Concepts Matrix Reasoning	Digit Span, Letter Number Sequencing, Arithmetic	Arithmetic	Coding, Symbol Search, Cancellation	Block Design Picture Completion Matrix Reasoning	
Taiwan*	Similarities, Information, Vocabulary, Comprehension, Word Reasoning Arithmetic	Similarities Picture Concepts Matrix Reasoning	Digit Span, Letter Number Sequencing Arithmetic		Coding, Symbol Search, Cancellation	Block Design, Picture Completion Symbol Search	

*. Similarities in structure found for Hong Kong, Macau, and Singapore samples

The Block Design test correlated significantly with the other tests on the visual processing and processing speed factors, while the Matrix Reasoning task loaded on visual processing and fluid reasoning. Finally, the Symbol Search task, a subtest within the Processing Speed Index loaded on both visual processing and processing speed factors.

Is structural invariance sufficient evidence of validity? Gregoire, Georgas, Saklofske, van de Vijver, Wiezbecki and colleagues (2008) observed that studies of structural invariance were limited to highly industrialized and wealthy northern countries. As such it can be argued that the 12-nation study (Georgas et al.; 2003) and other empirical evidence of cross-cultural structural validity of the WISC-IV could neither be generalized to less industrialized nations nor to countries in the Southern Hemisphere (Gregoire et al., 2008). Additionally, the 12-nation study found national differences in mean scores that were large enough to warrant explanation. For example, on the Digit Span subtest, relatively low scores were observed for Lithuanian children, while South Korean children performed significantly better than all other samples (Georgas et al., 2003). The authors compared the word forms of numbers between countries. Noting that there were longer numerical words in Lithuanian language than in the South Korean language, Georgas et al. (2003) posited a negative relationship between word length and test performance. Georgas and colleagues (2003) rejected the argument that the discrepancies in performance between national samples on the Digit Span subtest can be attributed to differences in ability. Instead they pointed to possible language related bias within the task. The authors argued that comparisons of working memory abilities between countries could not be made without considering linguistic factors (Gregoire et al., 2008). Analyses also demonstrated significantly higher scores for South Korean children and significantly lower scores for Greek children on the Symbol Search task

(Georgas et al., 2003). Authors attributed South Korean performance to their exposure to highly stylized characters in their language system (Gregoire et al., 2008). An explanation could not be offered for the low performance of Greek children.

One possible implication of these findings is that the notion of structural invariance as a robust indicator of cross-cultural validity is weak. Otherwise, we could confidently argue that South Koreans have better working memory than Lithuanians, and better processing speed ability than Greeks. The authors did not test whether South Korean strengths in working memory or processing speed ability can be generalizable beyond verbal measures of working memory and visual motor measures of processing speed. Future studies perhaps can utilize a mixed factorial design in which the relevant matched national samples can be compared on their performance on a variety of working memory (digit span, sentence span, visual-spatial) or processing speed (visual-motor, auditory) tasks.

The assumption of cross-cultural bias appears to be the rationale for many of the international adaptations and standardizations of WISC tests. The WISC-III had been translated into approximately 20 languages, including Chinese, Greek, and French (Georgas, Weiss, Van de Vijver & Saflofske, 2003). The WISC-IV also had been adapted cross-culturally (e.g. WISC-IV (Canadian) (Wechsler, 2004a); WISC-IV (Spanish) (Wechsler, 2004b); WISC-IV (UK) (Wechsler, 2004c); WISC-IV (French) (Wechsler, 2005)). Adaptations are certainly useful for minimizing the negative effects associated with the differential functioning of test items, method bias and construct bias that may accompany tests when they are imported for use elsewhere. For smaller, less developed countries, funding of expensive test modification studies can burden limited financial resources. Care therefore must be taken to use empirical rather than ad hoc approaches for identifying areas of bias or differential functioning in tests. Additionally, alternative strategies such as score adjustments should be considered as a way of addressing issues of bias if they do exist.

The following section will provide a background on the use of the WISC-IV (US) in Trinidad and Tobago, as well as a discussion of some of the issues and problems related to its use. The section also presents a pilot study which investigated the structural validity of the WISC-IV (US) in a sample of referred T&T children, as well as a discussion of the implications of the findings for the development of a cross-cultural interpretive model.

The WISC-IV (US) in Trinidad and Tobago (T&T) – Building a Rationale for a Cross-Interpretive Model

Trinidad and Tobago (T&T) is an independent twin island state in the southern Caribbean. Prior to independence in 1962, both islands were colonized by the Spanish, French and British. Tobago was also colonized by the Dutch and the Courlanders. Traditionally considered as one of the world's developing states, T&T officially achieved 'developed country status' in 2011 (OECD, 2011). T&T's economy is based predominantly on petroleum and natural gas and it has been recognized as a high income country by the World Bank (2015). The population of T&T is about 1.3 million, the largest percentage of which are of African and East Indian descent (Trinidad and Tobago Central Statistical Office (TTCSO); 2011). This multi-ethnic society also is made up of people of Amerindian, Chinese, European, Middle Eastern, and Mixed heritage (TTCSO, 2011). Christianity is the largest religion in the nation, followed by Hinduism, Islam, and other faiths (TTCSO, 2011). T&T's official language is English, although most nationals speak an English, French and African based dialect in informal settings. Many nationals are also familiar with and occasionally use East Indian based colloquialisms.

In T&T, the WISC-IV (US) has been used by T&T based psychologists for more than a decade; however, there has been no prior attempt to establish WISC-IV validity for this population. This is an unfortunate situation in light of widespread concern about the risks associated with using un-validated psychometric instruments. Some concerns regarding the validity of the WISC-IV (US), particularly the Similarities and Picture Concepts subtests, have been raised.

Possible Sources of Content Bias – Similarities and Picture Concepts

The Similarities subtest measures verbal reasoning ability by asking children how two words representing either objects or abstract concepts are alike (Wechsler, 2003b). The Picture Concepts subtest asks the respondents to use their reasoning ability to identify pictures that 'go together' (Wechsler, 2003b). Both tests have been designed to measure how well associations between words or objects are made. Furthermore, how well the respondent performs will depend on the extent to which they can identify the words or objects, their meanings or functions, and any unique physical or abstract characteristics they may possess.

An examination of the Similarities and Picture Concepts tests has revealed items that may function differentially depending on the T&T child's level of exposure to US culture. As an example, an item on the Similarities test asks how an *'apple and a banana'* are similar (Wechsler, 2003b). Apples are available in abundance for purchase in T&T, however they are not grown locally and are therefore relatively expensive. It therefore may be presumptuous to assume that every child in T&T is as familiar with an apple as they are with an orange, mango, or pommerac which are indigenous fruits. Another Similarities question which asks for the connection between *'winter and summer'*

(Wechsler, 2003b) also seems irrelevant in a T&T context where there are only two seasons *'rainy and dry'*.

Similar concerns emerge from exploring the Picture Concepts items that feature objects such as baseball gloves, strawberries, winter sleds, winter gloves and snowmen all of which are disconnected from T&T experiences. It is worth considering that these items may also measure exposure to US culture. The average T&T child should not be expected to know that a sled or snowman are play objects for the winter unless they were taught formally or learned from actual experience. Likewise, their northern counterparts may not necessarily identify a mosquito coil as a means of repelling insects or a pommerac as a red, pear-shaped, tropical fruit unless they have been exposed to this information. One therefore would suppose that had an IQ test contained either of these stimuli, an advantage would be gained based on access to culture specific knowledge.

Implications of Test Bias in Clinical Practice

The WISC-IV (US) is used by psychologists for ruling out developmental and learning disabilities and various clinical disorders (e.g. Autism Spectrum Disorder, Oppositional Defiant Disorder. For example, based on the general diagnostic criteria of the more recent Diagnostic and Statistical Manual 5th Edition (DSM-V; APA, 2013), a child with academic difficulties is rarely diagnosed with a learning disability if their WISC-IV (US) IQ score is lower than the Borderline range (IQ score ≤ 70, 2nd percentile). Such a low IQ score does not exclude intellectual or developmental disorders as possible underlying causes of the academic problem. Of course, these diagnostic interpretations are based on the assumption that the WISC-IV (US) provides a valid and reliable estimate of intelligence in T&T children. Furthermore, it is assumed that the WISC-IV (US) accurately predicts academic performance in this population. But, in the absence of proper validity studies to rule out bias, interpretive errors may be made on the basis of flawed assumptions.

One such error can relate to predicting academic outcomes based on WISC-IV (US) performance. Errors of prediction can occur if the T&T and US education systems use different approaches to academic learning. Indeed, it has been argued that the both countries differ in pedagogical methodology. In T&T, high stakes, highly competitive, ‘one-shot’ examinations are the norm. Teaching approaches are generally subject and exam focused (Rambhajan, 2007; Spence, 2007). Highly specialized teaching methods focusing on rote memorization and repeated practice thus are a lingering feature of the T&T education system (Steinbach, 2012). In contrast, the US education system promotes creativity and critical thinking in education (Garkov, 2002; Kim, 2005). It is therefore possible, that in the US, where reasoning and critical thinking are valued, subtests that measure inductive and deductive reasoning may reveal stronger associations with academic success than in T&T. Conversely, processing speed or memory may explain greater variance in academic performance than reasoning ability in T&T children.

Errors can also impact diagnostic decision making. Consider a scenario in which a child presents with reading difficulties. Through diagnostic testing, the child is assessed to be of borderline ability (e.g. FSIQ = 75, 5th percentile) on the WISC-IV (US), and borderline academic performance (Reading Composite = 70, 2nd percentile) on the Wechsler Individual Achievement Test II (WIAT II; Wechsler, 2001a). Assuming that the WISC-IV(US) truly measures intellectual ability; and using the ability/achievement discrepancy criteria of the WIAT II (Wechsler, 2001b), a discrepancy of 5 would make a diagnosis of Reading Disorder unlikely. However, if performance on the WISC-IV(US) is subject to the influence of an extraneous variable such as environmental exposure, then there is a real possibility that the child has been misdiagnosed. For instance, if the child lives in an isolated community and has little access to the kinds of resources or interactions that can stimulate intellectual development, the FSIQ score may actually underestimate the child's true academic potential. Furthermore, based on the ability/achievement discrepancy, the possible existence of a learning disability would have been missed. As a result, this child will be blocked from much needed specialized interventions. For the purposes of interpretation, the psychologist should be concerned about whether the test actually measures intelligence, and equally important, whether performance on the test is also related to as yet unknown environmental factors (van de Vijver & Tanzer, 2004).

These substantive issues that concern T&T psychologists about the use of the WISC-IV (US) should be explored. To this end, a series of pilot studies were conducted to assess the structural validity of the WISC-IV (US) in a sample of referred T&T students. The following sections will summarize the findings of the study. The results described below raised interesting questions about the influence of environment and exposure on the cross-cultural validity of the WISC-IV (US) in T&T children. A number of hypotheses were derived from the results of the studies, which will be presented at the end of the chapter.

Investigating the Relationship between Environmental Variables and WISC-IV (US) Structural Validity in a Referred Sample of T&T children – A Pilot Study

The data for this study were available from 354 students, selected from a pool of 755 students who applied to the Student Support Services Division (SSSD) of the T&T Ministry of Education (MOE) for concessions for the country's national tests (2008 – 2012). These applicants were either self-referred (applications were initiated by the parent or guardian) or were referred by the SSSD. Participants (N=401), for whom key core subtest information were missing or physical or mental disorders were reported, were removed from the data set. leaving 354 cases for analysis. Demographic and WISC-IV (US) test data were collected from psychological reports that had been attached to the concession applications. The majority of assessments had been conducted between 2006 and 2012 by Masters and Doctoral level psychologists trained in the administration of psychometric tests. A small number of administrations were conducted by supervised trainees of Doctoral level psychologists (See Table 2.6.).

Table 2.6.

Demographic characteristics of a referred sample of T&T children (N=354)

	N	%
Female	130	36.7
Male	224	63.3
6-9 years	113	31.9
10-12 years	196	55.4
13-16 years	45	12.7
Self Referred	163	46.0
SSSD Referred	191	54.0
African Descent	144	40.7
Chinese Descent	3	.8
East Indian Descent	103	29.1
Mixed Descent	41	11.6
White Descent	41	11.6
Not Reported	22	6.2
Central District	60	16.9
North District	129	36.5
South District	164	46.3
Tobago	1	.3
Learning Disability	128	36.2
Mild Mental Retardation	79	22.3
Moderate Mental Retardation	57	16.1
Attention Deficit Hyperactivity Disorder	31	8.8
Speech/Language	31	8.8
Autism Spectrum	4	1.1
No Diagnosis	24	6.8

The following data were available for this sample: gender, age, ethnicity, geographical region, type of referral, and nature of diagnosis. The sample consisted of 224 males and 130 females; ages ranged from 6 to 16 years (M = 10.38, SD = 1.97). The Full Scale IQ (FSIQ) scores ranged from 40 to 135 (M = 74.9, SD = 22.19).

Many of the children were diagnosed with a disability: (38.4%) were diagnosed with either a mild or moderate form of intellectual disability, and (36.2%) were diagnosed with some form of academic learning disability. Others in the sample were diagnosed with attention deficit/hyperactivity disorder (ADHD, 8.8%), autism spectrum disorder (1.1%) and speech-language disorder (SLD, 8.8%), while approximately 6.8% had no diagnosis at the time of referral.

US sample data was obtained from the results of the WISC-IV special group studies[4] (Wechsler, 2003b). The WISC-IV (US) scores for the T&T and US samples were compared. WISC-IV (US) standardization raw data was not used in this study, therefore it was not possible to perform between group ANOVAs or calculate effect sizes. Table 2.7 summarizes the means and SDs for the T&T children diagnosed as learning disabled (N = 128) and US children diagnosed with mixed Reading, Mathematics and Written Expression Learning Disorders (N = 38). T&T children obtained higher scores than their US counterparts on most scales with the exception of the Processing Speed Index (PSI).

When T&T and US children with Mild Mental Retardation (Mild MR)[5] were compared, the T&T sample performed worse on all indices except the Working Memory Index (WMI) (T&T: (M = 71.24, SD = 9.86); US (M = 66.80, SD = 11.10). See Table 2.8. for the means and SDs for T&T and US samples diagnosed with Mild Mental Retardation.

---

[4] At the time of the standardization study, special group validity studies were carried out to determine if the WISC-IV provided valid estimates of ability in groups with various clinical and learning diagnoses. Samples obtained from clinical and educational settings were generally convenience samples and therefore data may not be representative of the respective populations (Wechsler, 2003b)

[5] The DSM-IV TR (APA, 2000) criteria for a diagnosis of Mild Mental Retardation is a FSIQ score between 50-55 and 70 with co-existing deficits in at least 2 areas of adaptive functioning.

Table 2.7.

Comparison of WISC-IV (US) IQ scores (Wechsler, 2003b): Trinidad & Tobago vs. United States children with LD.

	Trinidad & Tobago		United States	
	Mean	SD	Mean	SD
Full Scale IQ	88.94	14.76	87.60	10.60
	(N=128)		(N=38)	
Verbal Comprehension Index	93.22	14.00	89.80	11.40
	(N=128)		(N=39)	
Perceptual Reasoning Index	91.99	16.83	90.10	12.50
	(N=128)		(N=42)	
Working Memory Index	92.92	14.11	89.70	12.30
	(N=128)		(N=41)	
Processing Speed Index	85.74	13.62	90.50	12.60
	(N=128)		(N=38)	

Table 2.8.

Comparison of WISC-IV (US) IQ scores (Wechsler, 2003b): Trinidad & Tobago vs. United States children with Mild Mental Retardation.

	Trinidad & Tobago		United States	
	Mean	SD	Mean	SD
Full Scale IQ	59.58	6.55	60.50	9.20
	(N=79)		(N=56)	
Verbal Comprehension Index	64.66	7.69	67.10	9.10
	(N= 9)		(N=58)	
Perceptual Reasoning Index	64.41	9.47	65.50	10.30
	(N=79)		(N=63)	
Working Memory Index	71.24	9.86	66.80	11.10
	(N=79)		(N=62)	
Processing Speed Index	70.37	9.34	73.00	11.60
	(N=79)		(N=38)	

Comparisons of T&T and US children with Moderate Mental Retardation (Moderate MR)[6] revealed that the US sample performed better on all indices except the Verbal Comprehension Index (VCI) (T&T: (M = 53.38, SD = 7.90); US (M = 52.30, SD = 7.50). See Table 2.9 for the means and SDs for T&T and US samples diagnosed with Moderate Mental Retardation.

Table 2.9.

Comparison of WISC-IV (US) IQ scores (Wechsler, 2003b): Trinidad & Tobago vs. United States children with Moderate Mental Retardation.

	Trinidad & Tobago		United States	
	Mean	SD	Mean	SD
Full Scale IQ	44.80	4.81	46.40	8.50
	(N=56)		(N=47)	
Verbal Comprehension Index	53.38	7.90	52.30	7.50
	(N=56)		(N=55)	
Perceptual Reasoning Index	50.07	5.15	52.50	9.20
	(N=56)		(N=57)	
Working Memory Index	56.96	8.52	57.00	9.50
	(N=56)		(N=53)	
Processing Speed Index	55.82	6.91	58.20	11.00
	(N= 6)		(N=51)	

Between the two ADHD samples. The US sample performed better than the T&T sample on all WISC-IV (US) composites. See Table 2.10 for the means and SDs for the T&T and US samples that were diagnosed with ADHD.

---

[6] The DSM-IV TR (APA, 2000) criteria for a diagnosis of Moderate Mental Retardation is a FSIQ score between 35-40 and 50-55 with co-existing deficits in at least 2 areas of adaptive functioning.

Table 2.10.

Comparison of WISC-IV (US) IQ scores (Wechsler, 2003b): Trinidad & Tobago vs. United States children with Attention Deficit Hyperactivity Disorder.

	Trinidad & Tobago		United States	
	Mean	SD	Mean	SD
Full Scale IQ	90.61	19.61	97.60	14.00
	(N=31)		(N = 82)	
Verbal Comprehension Index	92.81	17.63	99.00	13.60
	(N=31)		(N = 83)	
Perceptual Reasoning Index	94.42	20.06	100.10	14.20
	(N=31)		(N = 89)	
Working Memory Index	94.26	18.38	96.10	15.50
	(N=31)		(N = 89)	
Processing Speed Index	86.42	13.92	93.40	12.60
	(N=31)		(N = 87)	

Examinations of between US and T&T sample means and SDs revealed a trend of lower performance for the T&T samples. An exception was the T&T Learning Disability sample which performed better than the US sample on all but the Processing Speed Index. Also T&T children with mild mental retardation outperformed their US counterparts on the working memory index, and the T&T sample with moderate mental retardation scored higher on the verbal comprehension index.

Broad explanations for performance disparities between groups cannot be offered without knowing if the differences between groups were significant or meaningful. Additionally, even if these statistics were available the findings would not be generalizable because both the US and T&T samples may represent only a small subset of the T&T and US population of children with disabilities. Also, monitoring of adherence to diagnostic standards across samples was not possible for this study. Moreover, the lack of key demographic information about the US samples makes it difficult to point to variables that may explain similarities and differences in scores between the groups.

The main focus of this study was to investigate whether the current factorial structure of the WISC-IV (US) could be generalized to a T&T population, or whether some other structure will emerge in this sample. First, confirmatory factor analyses on the 10 core subtests were done to establish the factor structure and broad ability composition of the WISC-IV (US). Second, model fit indices for the WISC-IV (US) four-factor solution, obtained in the US standardisation sample, were compared with those of alternative single and three-factor models. Third, model fit indices of the four-factor solution were compared in children of differing socioeconomic groups.

The Validity of the Current 4-factor Structure in a sample of Referred T&T children

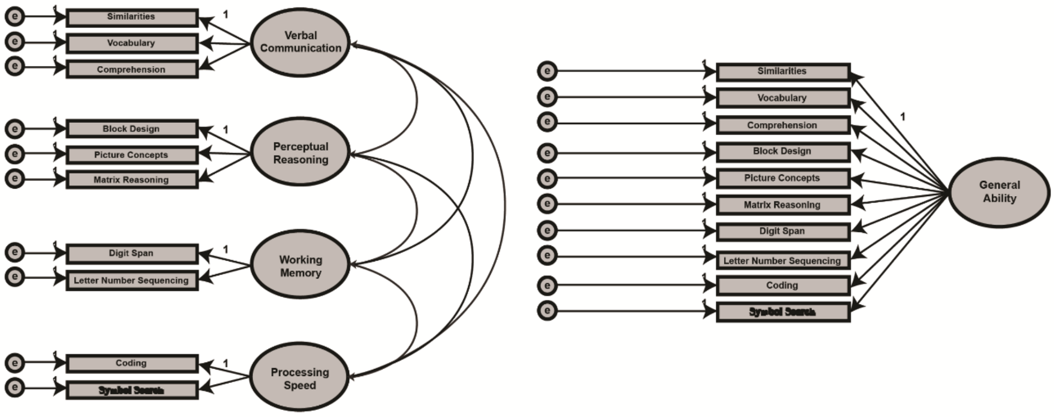
Confirmatory factor analyses (CFAs) were carried out on the main sample using maximum likelihood estimation in AMOS 21 (with SPSS) to compare the current factorial structure of the WISC-IV (US) to alternative models, proposed in the literature and generated as part of the study. Four models were compared:

Model 1. The four-factor current model (VCI, PRI, WMI, PSI) as proposed by Wechsler (2003b) (Figure 2.9).

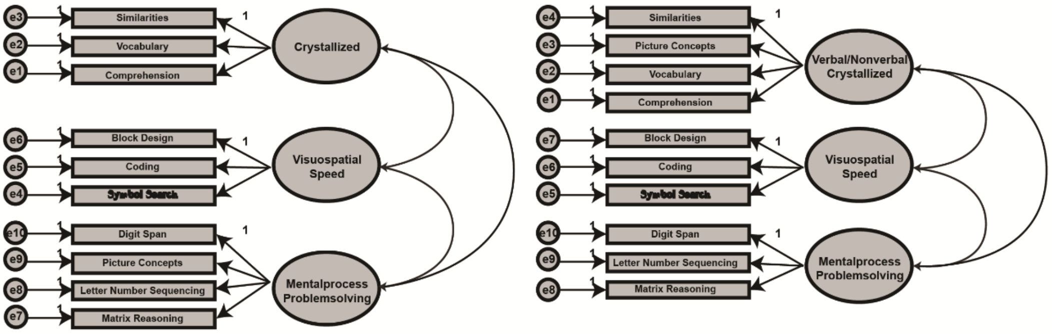
Model 2. A single factor model that loads all ten subtests on to one general ability factor (See Figure 2.9). This model is based on the Spearman (1904) theory that performance on all measures of ability can be explained by general intelligence factor.

Model 3. A three factor model based on the findings of Gregoire et al. (2008) and Lecerf et al. (2010) comprising Crystallized ability (Similarities, Vocabulary, Comprehension); Visuospatial/Speed (Block Design, Coding, Symbol Search), and Mental processing/Problem solving (Picture Concepts, Matrix Reasoning, Digit Span, Letter Number Sequencing) factors (Figure 2.9).

Model 4. This model is based on considerations related to cultural factors, and consists of a Verbal/Nonverbal crystallized ability factor (Similarities, Vocabulary, Comprehension, Picture Concepts), a Visuospatial/Speed factor (Block Design, Coding, Symbol Search), and Mental processing/Problem solving factor (Matrix Reasoning, Digit Span, Letter Number Sequencing) (Figure 2.9).



Model 1 *Model 2*



Model 3 *Model 4*

Figure 2.9. Models 1 to 4: Confirmatory Factor Analysis[7].

---

[7] Model 1 = WISC-IV current 4-factor model with Verbal Comprehension, Perceptual Reasoning, Working Memory and Processing Speed factors (Wechsler, 2003a)
Model 2 = Single general factor model (Spearman, 1904)
Model 3 = *C*rystallized ability, visuospatial/speed, and mental processing/problem solving model (Gregoire et al., 2008; Lecerf et al., 2010)
Model 4 = Verbal/nonverbal crystallized ability factor, a visual spatial/speed factor, and mental processing/problem solving model.

The difference between this model and Model 3 is the presence of the Picture Concepts subtest in the Verbal/Nonverbal crystallized ability factor. The rationale for this model was provided earlier under the section headed "Possible Sources of Content Bias – Similarities and Picture Concepts". With specific reference to Trinidad and Tobago (T&T), the culturally-loaded non-verbal items can be argued to also load on the crystallized intelligence factor, since performance is likely to be influenced by knowledge of US culture. It was therefore considered conceivable and worthy of exploration whether Picture Concepts items would load on crystallized ability for a T&T population.

Results

Absolute fit of each model was evaluated using the normed chi square (χ^2/df). Good model fit is reflected in normed chi-square values below 2 and *p* values over .05 (Bollen, 1989). Considering the size of the sample (N= 354), a significant χ^2 could arise even in cases of good model fit. Thus, additional fit statistics were used to test model adequacy such as the comparative fit index (CFI), the Tucker Lewis index (TLI), the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI) and the root mean square of approximation (RMSEA). CFI, TLI, GFI, and AGFI values in excess of .95; and RMSEA values lower than .06 are indicative of good model fit (Hu & Bentler, 1998, 1999). Since the compared models were non-nested models, the Aikake Information Criterion (AIC) has been offered as a useful method for determining the best fit among competing models (Vrieze, 2012), the best fitting model being the one with the smallest AIC value (Watkins, 2010). AIC values and absolute fit indices revealed that the current four-factor model provided a better fit to the data than the single and three factor alternative models. The results were as follows: lower χ^2 = 63.13 (df = 29, p = .000), χ^2/df ratio = 2.18, AIC = 115.13, CFI = 0.989, TLI = 0.983, GFI = 0.965, AGFI = 0.934 and

RMSEA = 0.058. Based on the results, the current WISC-IV (US) four-factor model as described for the US standardization sample was found to be the best fitting model for the WISC-IV (US) core subtests in this sample, over and above the other proposed models in this study. Table 2.11 presents overall fit indices for models 1 to 4. Figure 2.10 shows that the standardized loadings for the current four-factor model were high and statistically significant for all of the first-order factors (*B*s =.70 – .91, *p*s < .001).

Table 2.11.

Goodness of Fit Indexes for Evaluating Model Adequacy (N= 354)

Model	χ^2	*df*	*p*	χ^2/df	AIC	CFI	TLI	GFI	AGFI	RMSEA
1	63	29	.000	2.18	115.13	0.989	0.983	0.965	0.934	0.058
2	230.55	35	.000	6.59	270.55	0.937	0.920	0.876	0.805	0.126
3	120.40	32	.000	3.76	166.40	0.960	0.972	0.936	0.890	0.088
4	140.86	32	.000	4.40	186.86	0.965	0.951	0.924	0.870	0.098

Note. df = degree of freedom; AIC = Akaike's information criterion; CFI = comparative of fit index; TLI = Tucker Lewis index; GFI = goodness of fit index; AGFI = adjusted GFI; RMSEA ¼ root mean square error of approximation.

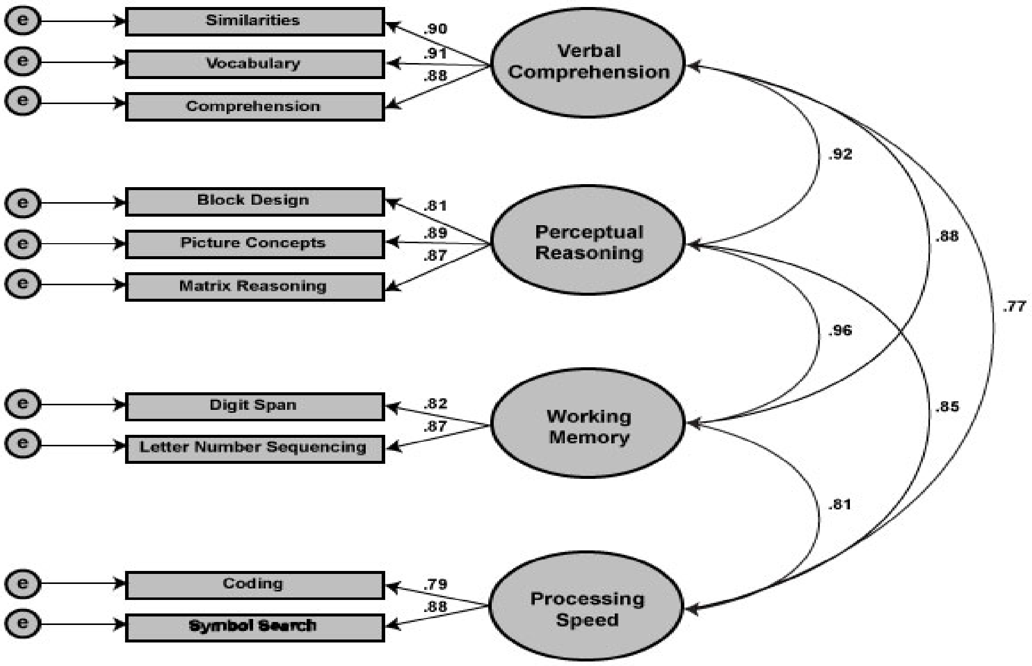


Figure 2.10. Standardized loadings for the current WISC-IV four-factor model.

Model 4, which could be described as the construct bias model of the WISC-IV (US), proposed that the Picture Concepts subtest would be a primary indicator of crystallized intelligence in this sample. The analysis did not provide support for this hypothesis. Lecerf, Rossier, Favez, Reverte, and Coleaux (2010) also tested this assumption in a sample of French children who completed the French version of the WISC-IV. Their proposed model did not demonstrate good fit. These findings suggest that while the Picture Concepts subtest measures some degree of cultural exposure, it may for the most part be a measure of fluid reasoning ability.

Although, the ‘bias model’ (Model 4) did not provide best fit to the data overall, it is possible that the ‘culture’ effects are moderated by specific environmental factors. For example, the loading of specific components onto factors may depend on the levels of SES that affect access or exposure to information. In the 2nd part of the study, SES was explored for its influence on model fit. This was accomplished by examining the invariance between sub-samples through multi-sample confirmatory factor analysis.

Investigating Structural Invariance of the WISC-IV (US) based on SES

In T&T, people of middle to upper socioeconomic status in the society have access to a broad range of US influences through their ability to afford television, internet and travel. Conversely, poorer people are less likely to benefit from such diverse experiences. An appropriate indicator of socioeconomic status, household income, was not available in this sample. Instead, another grouping variable (Referral) which assigned children according to who paid for psychological services (either the Student Support Services Division (SSSD) or parents), was thought to be a useful indicator of socioeconomic status. The policy of the SSSD is to provide assessments to children with learning difficulties who have been identified as needing intervention. With very few exceptions, such children generally come from lower middle/working to poor backgrounds and would not otherwise have access to psychological services[8]. The Self referred children are those whose parents or caregivers were able to afford the £400 to £700 fee typically paid for psychological assessments. The data also revealed that a large portion of this sample of children also attended paid private institutions, which is often an indication of middle to upper social status in T&T.

Multi-sample analysis based on referral category. The SSSD referred sample consisted of 132 males and 59 females, and the Self-referred sample consisted of 92 males and 71 females. The SSSD referred sample ranged from ages 6 to 15 years (M = 10.65, SD = 2.07). For the Self-referred sample, ages ranged from 6 to 16 years (M = 10.07 years; SD = 1.81). FSIQ scores ranged from 40 to 135 (M = 60.39, SD = 15.38) and from 45 to 131 (M = 92.04; SD = 15.87) for the SSSD referred and Self-referred samples respectively. Table 2.12 presents a summary of demographic data by referral category.

Table 2.12.
Key demographic data by referral category for the T&T sample of referred children

	Self Referred (N=163)		SSSD Referred (N=191)	
	N	%	N	%
Female	71	43.6	69	35.2
Male	92	56.4	127	64.8
Learning Disability	90	55.2	38	19.9
Mild Mental Retardation	8	4.9	71	37.2
Moderate Mental Retardation	4	2.5	53	27.7
Attention Deficit Hyperactivity Disorder	27	16.6	4	2.1
No Diagnosis	34	20.8	25	13.1
6 to 9	56	34.4	57	29.8
10 to 12	98	60.1	98	51.3
13 to 16	9	5.5	36	18.9

---

[8] Data on the actual percentage of SSSD referred children whose parents could afford to pay for private testing was not available for this study.

Results

The Self Referred and SSSD referred groups were compared in terms of means and variances. Means and standard deviations of specific and global test scores are provided in Table 2.13, which also presents the mean score comparisons, F statistics, significance values and effect sizes. Results demonstrate that the performance of the self-referred children significantly exceeded that of the SSSD referred children on all WISC-IV composites with moderate effects for referral category on the FSIQ, VCI and PRI indices and small effect sizes on WMI and PSI.

Table 2.13.

IQ score (WISC-IV (US)) by referral category.

	Self Referred		SSSD Referred				
	Mean	SD	Mean	SD	F-value	*p*	Cohen's *d*
FSIQ	92.04	15.87	60.39	15.38	19.02	.000**	.44
VCI	94.56	14.85	65.51	14.48	18.59	.000**	.43
PRI	95.79	16.75	65.23	16.50	17.25	.000**	.42
WMI	95.60	15.49	70.68	14.69	15.51	.000**	.39
PSI	87.52	13.63	69.33	14.42	12.13	.000**	.35

**. $p < .001$

Note: FSIQ = Full Scale IQ, VCI = Verbal Comprehension Index, PRI= Perceptual Reasoning Index, WMI= Working Memory Index, PSI= Processing Speed Index

Table 2.14 presents results of independent analyses. AIC values and absolute fit indices revealed that the current four-factor model provided a better fit to the Self referred sample data than the SSSD referred sample data. In fact, the current four factor model demonstrated good model fit for the Self referred sample and poor model fit for the SSSD referred sample.

Table 2.14.

Goodness of Fit Indexes for Evaluating Model Adequacy.

	χ^2	*df*	*p*	χ^2/df	AIC	CFI	TLI	GFI	AGFI	RMSEA
Self	35.56	29	.187	1.25	87.56	0.991	0.987	0.959	0.921	0.037
SSSD	52.15	29	.005	1.79	104.15	0.978	0.966	0.947	0.899	0.065

Note. Self = Self referred, SSSD = Referred by SSSD, AIC = Akaike's information criterion; CFI = comparative of fit index; TLI = Tucker Lewis index; GFI = goodness of fit index; AGFI = adjusted GFI; RMSEA = ¼ root mean square error of approximation.

The summary of the multi-group analysis and the χ^2 difference test results are presented in Table 2.15. The table shows that the χ^2 values for all models were significant at the p<.001 level with the exception of the unconstrained model (*p* = .007).

Table 2.15.

Goodness of fit indices for Unconstrained and Constrained Models[9]

Model	χ^2	*df*	*p*	χ^2/df	CFI	TLI	GFI	AGFI	RMSEA
1	35.56	58	.007	1.51	0.984	0.975	0.952	0.909	0.038
2	109.00	64	.000	1.70	0.975	0.965	0.942	0.900	0.045
3	120.10	74	.000	1.64	0.974	0.968	0.935	0.903	0.043
4	138.83	84	.000	1.65	0.970	0.968	0.926	0.903	0.043

Note. AIC = Akaike's information criterion; CFI = comparative of fit index; TLI = Tucker Lewis index; GFI = goodness of fit index; AGFI = adjusted GFI; RMSEA = root mean square error of approximation.

---

[9] Model 1: WISC-IV four factor model in which the factor loadings, covariances and variances, and the error variances are freely estimated.
Model 2: The factor loadings are fixed to be equal across samples.
Model 3: The factor variances and covariances are fixed to be equal across samples.
Model 4: The error variances are fixed to be equal across samples.

The unconstrained model was found to attain the lowest χ^2 /df ratio (1.51), and RMSEA (0.038), and the highest values for GFI (.952), AGFI (.909), TLI (.975), and CFI (.984). All considered, the unconstrained model was determined to be the best fitted model. Comparisons between the unconstrained models and the constrained models revealed χ^2 difference values that are statistically significant (See Table 2.16). Such results suggest that the equality constraints do not apply to the two groups. In other words, the findings suggest that the factor loadings, covariances, variances and error variances are not the same for the self-referred and SSSD referred samples.

Having found non-invariance, factors and subtests were examined by testing for invariance of all factor loadings in each WISC-IV index separately, and where non-invariance was found, testing for invariance at the subtest level (Byrne, 2004).

Table 2.16.

Chi Square difference tests for comparing nested models.

Model	χ^2 difference	*df* difference	*P*
1	21.29	6	.002*
2	33.39	16	.007*
3	51.13	26	.002*

**. *p*< .01

Testing for invariant factor loadings related to VCI, PRI, WMI, PSI. Table 2.17 shows that in testing the 4 models related to VCI, PRI, WMI, and PSI, the χ^2 difference values for VCI and WMI were found to be statistically significant and therefore non-invariant across groups ($p<.05$). The next step to identify the non-invariant items within the identified factors produced results that are summarized in Table 2.17. Comparisons demonstrated that the test for invariance related to the factor loadings for the Similarities, Digit Span, and Letter Number Sequencing subtests resulted in χ^2 difference values that were significant (p<.05).

Table 2.17.

Chi Square difference values for comparing Factor and Subtest Loadings

Model	χ^2	*df*	$\Delta\chi^2$	Δ*df*	*p*
Model 1. Unconstrained	87.71	58			.007*
Model 2. Factor loadings on PRI constrained equal	90.72	60	3.01	2	.222
Model 3. Factor loadings on VCI constrained equal	101.55	62	13.84	4	.008*
Model 4. Factor loadings on WMI constrained equal	100.06	62	12.36	4	.015*
Model 5. Factor loadings on PSI constrained equal	93.37	62	5.66	4	.226
Model 6. Model 2 with factor loadings on Similarities constrained equal	101.49	61	13.79	3	.003*
Model 7. Model 2 with factor loadings on Comprehension constrained equal	90.93	61	3.23	3	.358
Model 8. Model 2 with factor loadings on Vocabulary constrained equal	92.30	61	4.59	3	.204
Model 9. Model 2 with factor loadings on Digit Span constrained equal	100.59	62	12.88	4	.012*

*p<.05. Note. PRI = Perceptual Reasoning Index, VCI = Verbal Comprehension Index, WMI = Working Memory Index, PSI = Processing Speed Index,

Implications of the Pilot Study

The means comparisons showed that children in the self-referred sample performed better than children in the SSSD referred sample on all WISC-IV(US) indices. While the results seem to replicate studies that have shown a statistically significant relationship between income and IQ (e.g. Weiss, Harris, Prifitera, Courville et al., 2006; Lemos, Almeida & Colom, 2011), some caution in interpretation is necessary.

An examination of the demographic data for both samples revealed that more children in the SSSD referred group were diagnosed with mental retardation (64.9%) than in the self-referred group (7.4%). Additionally, more children in the self-referred group were diagnosed with a learning disability. This means that at least 55.2% of the self-referred children's IQ (compared with 19.9% of SSSD sample) equalled or exceeded a score of 70. These findings may reflect selection issues that should not be ignored. It is possible that SSSD referrals focus on children with the most severe of learning difficulties. In such a case, the SSSD sample may not be truly representative of lower income children with learning difficulties.

If the sample is a representative sample, then score discrepancies are hypothesized to reflect the effects of environmental deprivation. Children from lower income families are more likely to suffer from delays in brain and intellectual development as a result of a lack of nutrition and health care, as well as poor environmental stimulation and stability (Neiss & Rowe, 2000; Petrill, Pike, Price, & Plomin, 2004; Weiss, Harris, Prifitera, Courville et al., 2006, Hart, Petrill, Deckard, & Thompson, 2007). Conversely, children from well-off families are more likely to have access to the physical, intellectual and academic resources necessary for intellectual growth (Bradley & Casey, 1992; Lee & Barratt, 1993; Liaw & Brooks-Gunn, 1993). If it is true that the referral categories are indeed good proxy variables for household income, then the results of the between group comparisons of the pilot study provide a good rationale for investigating the relationship between income and measured intelligence in T&T children.

In another analysis of the pilot data, comparisons of model fit indices showed that the factor structure of the WISC-IV (US) can be replicated in the sample of self-referred students, but not in the sample of SSSD referred students. Such results seem to provide empirical evidence in support of the WISC-IV construct bias hypothesis. Although independent analyses alone were insufficient to determine invariance between self-referred and SSSD referred samples, simultaneous analyses provided further evidence. The χ^2 difference tests revealed that the factor patterns across the two samples were non-invariant. Overall, the results suggest that the WISC-IV (US) may be a valid instrument for use in certain groups of Trinidadian children but not others.

Further tests for invariance identified the Similarities, Digit Span, and Letter Number Sequencing tests as non-invariant across groups. These findings may be related to potential differences in quality of education between the two groups. The Similarities test uses a word analogy format which requires definition finding and word categorization. Based on the findings of cross-cultural validity studies, both steps require reasoning ability (e.g. Flanagan & Kaufman, 2004; Chen, Keith, Chen & Chang, 2009) as well as a broad general knowledge (crystallized ability) (Keith et al., 2006). Additionally, with regard to the Working Memory subtests, there also has been evidence of a link between Digit Span performance and quality of education (Otrosky – Solis & Lozano, 2006).

As has been discussed in Chapter 1, the link between intelligence and education has been demonstrated empirically (Cahan & Cohen, 1989; Rinderman, Flores-Mendoza & Mansur-Alves, 2010). Also, crystallized ability, which is said to be gained through the long term application of reasoning ability (Cattell, 1971; 1987), is also the result of direct instruction (Gottfredson & Saflofske, 2009; Lohman & Lakin, 2009). Therefore, quality of education is expected to impact a child's intellectual development (Cahan & Cohen, 1989; Ceci, 1991; Rinderman, Flores-Mendoza & Mansur-Alves, 2010). We know less, however, about how quality of education influences how individual subtests within an intelligence test are organized into factors. Take for example, a hypothetical test that measures fluid reasoning ability. In order to solve items correctly the examinee must be familiar with simple mathematical concepts such as Perimeter, Circumference or Area. If all examinees were familiar with these concepts, then there is very good reason to attribute individual differences in performance to differences in reasoning ability. Conversely, if there is considerable variability among examinees in their knowledge of these math concepts, then the items can load on both reasoning ability and crystallized ability. With specific regard to the results of the pilot, it is believed that Similarities and Working Memory subtests measure different constructs in the Self referred and SSSD referred groups because of the differences in quality of education between samples. Therefore, future work will investigate how quality of education in a T&T sample impacts the way subtests are organized within the WISC-IV(US).

The influence of parental education and household income on measurement non-invariance between high and low SES groups is another worthy area for future study in this thesis. Access to new knowledge learnt from parents or via the resources parents provide within the shared environment (i.e. internet, television, educational games, exploration and travel) has been associated with intellectual gains in children (Gottfried & Gottfried, 1984, Hart & Risley, 1992). Additionally, what a child learns from the parent depends on the parent’s fund of general knowledge. Studies have pointed to a modest relationship between parent education and the child’s intellectual levels (Neiss & Rowe, 2000; Lemos, Almeida &Colom, 2011; Ganzach, 2014). Furthermore, genetically informative studies by Rowe, Jacobson and Van den Oord (1999) demonstrated that the genetic potential for vocabulary development was better expressed in children raised by highly educated parents. This thesis will therefore examine how intelligence is related to parental education and household income, as well as the influence of parental education and household income on the fit of the WISC-IV model of intelligence in T&T children.

Another variable that was not examined in the study, US exposure, is considered useful for future research. Earlier it was hypothesized that access to US information through television, internet or travel may impact performance on the test. In light of the results of the multi-sample analysis and the fact that high SES children are more likely to have access to US culture, the thesis will investigate how US exposure influences model fit in a T&T sample.

In addition to understanding how environmental variables influence the structure of the WISC-IV, this thesis is also concerned about the predictive validity of the WISC-IV (US) in T&T children. Of particular concern is the relationship among environmental variables (specifically schooling), IQ and academic achievement in these children. The literature provides clear evidence that human intellectual ability has a principal role in the prediction of future achievement (Reece & Bonaccio, 2011), but achievement is a product not merely of individual characteristics, but also of environmental factors (Winne & Nesbit, 2010). In psychoeducational testing, diagnoses of learning disabilities are considered when a child’s academic performance falls significantly below scientifically established expectations based on their ability (APA, 2000). Criteria based on environmental variables such as quality of schooling lack this clarity, and so, in spite of their proven associations with academic achievement in developing countries (Gamoran & Long, 2007), educational variables play a less explicit role in diagnostic decision making. It is worth considering that a schooling factor potentially can be large enough to influence diagnostic decisions. This is plausible in a scenario where schooling can be shown to lower the effect of intelligence on academic performance. It is anticipated that the mediating effect of schooling on the intelligence/academic achievement may be observed in T&T, where there is a long history of inequity and competition within the education system. This topic will be discussed in more depth in Chapter 6. The hypothesis that school membership can outweigh intellectual ability in predicting scores on the national exam will be tested in this thesis.

In light of the above, the aims of this thesis are as follows:

1. To study the performance of T&T children on the WISC-IV (US) by first making comparisons between the T&T and US samples, and also by making within sample comparisons of performance on the various WISC-IV (US) subtests and composites.
2. To investigate the internal structure of the WISC-IV (US) based on T&T sample data, by using both exploratory and theory driven methods of analysis.
3. To investigate the relationship between parent education, income, schooling and US exposure on WISC-IV (US) measured intellectual ability in T&T children.
4. To investigate the relationship between intelligence, formal education and academic achievement in T&T children.
5. To assess the validity of a cross-cultural interpretive path model for understanding the performance of T&T children on the WISC-IV.

The interpretive model that is illustrated in Figure 2.11, diagrammatically summarizes the main hypotheses of this thesis. It features the antecedent variables, the central construct and the outcome variable. Antecedent variables are presented in the model as parental education, family income, quality of formal education, and direct exposure to US culture. The central construct is the WISC-IV (US) structural model of intelligence that best fits the sample data. The main outcome variable of interest, academic performance (measured by performance on T&T's national exams) is also presented in this diagram.

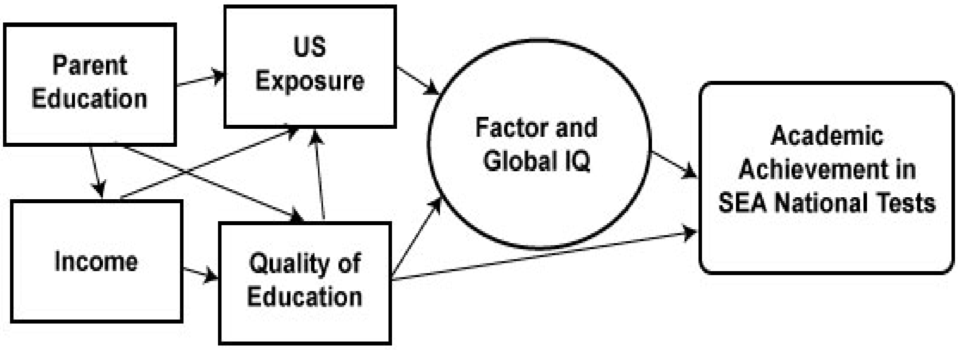


Figure 2.11. Simple path diagram for cross cultural interpretive model for WISC-IV (US) measured intelligence in T&T children.

Eight studies employing a wide array of statistical methods were conducted identify an interpretive framework that best explains the variance in the WISC-IV (US) performance of a sample of T&T children. The following chapter will describe the methodology employed in all eight studies that were conducted to meet the aims of this thesis. The chapter will provide descriptions of the sample, research procedures and instruments, as well as the statistical methods that were used in each study.

CHAPTER 3

METHODS

This chapter describes the methods that were used to assess the cross-cultural validity of the WISC-IV (US) in T&T children. Specifically, data was collected to investigate the fit of a cross-cultural interpretive model of WISC-IV measured intelligence. The first section of this chapter outlines the participant recruitment process. It also provides a description of the sample and compares the sample demographics with T&T population data. A description of all materials that were used for data collection is provided in the second section. The third section discusses the data collection procedures and statistical methods that were used in the studies. To conclude, the fourth section provides a description of the remaining chapters of this thesis.

Participants

Recruitment. The data for this study were collected from 11 and 12-year-old primary and secondary school T&T children who volunteered to participate during an ongoing recruitment process. Students were recruited through the help of the Ministry of Education of T&T (MOE) and school principals who gave the research team permission to collect data within their respective schools. Other participants were recruited verbally, with flyers or by email. Consent for individual testing was sought from the participants' parents or guardians, who were also asked to complete demographic questionnaires that were attached to the consent forms and submitted on the day of testing. See Appendix A for the Information Sheet, Consent Form and Demographic Questionnaire. Additional data about participant and school academic performance were sourced from various departments within the MOE. Based on the parental reports, participants were excluded if they presented with any clinical diagnoses that may potentially hamper the reliability of the WISC-IV (US) scores, such as depression, anxiety or psychotic disorder. Participants were also omitted if they reported a serious current medical condition, a motor dysfunction affecting fine motor skills; a visual impairment that prevents performance on visual/spatial tasks; or any hearing impairments that may negatively impact performance on verbal tasks.

Description of the Sample. Data were collected from a total of 211 children who were preparing to write or had just written the Secondary Entrance Assessment (SEA) national exams. Listwise deletion was applied to 6 cases that were missing key core subtest information. Additionally, two participants who were not T&T nationals were removed from the dataset, leaving 203 cases for analysis. The remaining sample consisted of 121 males and 82 females. The participants were divided almost equally by age (11 years = 51.2%; 12 years = 48.8%).

The sample fairly matched the population demographics reported by the Trinidad and Tobago 2011 Population and Housing Census Demographic Report (2011) on Gender, but not on the variables of Ethnicity or Parental Education. The majority of the sample were of African descent (36.9%) followed by those of mixed ethnicity (23.2%). Children of East Indian descent were underrepresented in the sample, while there were larger percentages of European and Chinese descent compared to national population data. Demographic data are summarized in Table 3.1 and comparisons with population data are presented in Table 3.2.

Table 3.1.

Key demographic data for the T&T sample of 11 to 12-year-old children, N = 203

	N	%
Gender		
Female	82	40.4
Male	121	59.6
Age		
11 years	104	51.2
12 years	99	48.8
Ethnicity		
African	75	36.9
Chinese	7	3.4
East Indian	39	19.2
European	33	16.3
Middle Eastern	2	1.0
Mixed	47	23.2
Parental Education		
Did not complete Primary	3	1.5
Primary School	15	7.4
O’Levels	51	25.1
A’Levels	21	10.3
Associates/ Technical	37	18.2
Bachelor	46	22.7
Masters	17	8.4
PhD	11	5.4
Other	2	1.0
Learning Disability	51	25.1
No Disability	150	73.9
No response	2	1.0
US Exposure		
None	26	12.8
Some	108	53.2
Moderate	52	25.6
High	16	7.9
No answer	1	0.1
Public School	109	53.7
Private School	94	46.3
School Ranking		
Excelling	91	44.8
Within Expectations	103	50.7
At Risk	9	4.4

Table 3.2.

Comparisons of sample vs population data by Gender, Ethnicity, and Parental Education

	Sample %	Population %
Gender		
Male	59.6	51.1
Female	40.4	49.9
Ethnicity		
African	36.9	34.2
Chinese	3.4	0.3
East Indian	19.2	35.4
European	16.3	0.6
Middle Eastern	1.0	0.1
Mixed	23.2	22.8
Parental Education		
No Primary	1.5	*
Primary	7.4	29.8
Secondary	35.4	43.5
Associates/Technical	18.2	6.2
Tertiary	36.5	8.4

Note: * no data available

Materials

Demographic Questionnaire. The demographic questionnaire was designed to obtain key data about the participants: Gender, Age, Name of School, Ethnicity, Parent Education, Parent Income, Learning Disability, and Exposure to US culture (see Appendix A).

Parental education. The questionnaire asked the informant to indicate the highest level of education achieved by each of two main caregivers in the household: Did not complete primary school, completed primary school, GCSE O' Levels, A' Levels, Associates Degree, Bachelor, Masters, and PhD. This information was recorded in the data tables as a five-level categorical variable: 1= Did not complete primary school; 2 =

Completed primary school; 3 = Completed O' Levels, 4 = Completed A' Levels and Associates Degree; 5 = Completed Bachelor degree or higher. For data analysis, a separate variable was created and used as the parent education variable. To create this variable, the following rules were followed:

1. If there was only one parent or caregiver in the home, then that person's level of education was used regardless of the level of education of the non-custodial parent or parents.
2. If two or more main caregivers were in the home the values for parents with the highest level of education was used for analysis.

US Exposure. This item was used as a way of categorizing the participants on level of exposure to US culture (Appendix A). The respondent was required to choose, on the basis of a brief description for each category, one of the following: 1 = Very Little Exposure; 2 = Some Exposure, 3 = Moderate Exposure; and 4 = Very Much Exposure.

Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) (US). The WISC-IV is an individually administered instrument for assessing the cognitive ability of children aged 6 to 16 years. The WISC-IV comprises 15 subtests which are organized into 4 ability indices and are used to produce 5 composite scores: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), Processing Speed Index (PSI) and the Full Scale IQ (FSIQ). The WISC-IV (US version) was standardized on a sample of 2,200 children that was representative of the US population based on data from the March 2000 census. Further information about the standardization procedures can be obtained from the WISC-IV Technical manual

(Wechsler, 2003a). Information about the reliability and validity of WISC-IV (US) can be found in Chapter 2.

WISC-IV (US) subtest raw scores are converted, using age based normative tables, to scaled scores which range from 1 to 19 (M = 10, SD = 3; Wechsler, 2003b). Within each of the four specific ability composites, scaled scores are added together and the sum is converted to an Index Standard Score which ranges from 40 to 150 (M = 100, SD = 15). Also, the scaled scores of all 10 subtests are summed and then used to calculate a Full Scale IQ (FSIQ) (Wechsler, 2003b). More information about WISC-IV indices, subtest and scoring procedures are presented in Chapter 2.

The Secondary Entrance Assessment (SEA). The SEA is a national examination for children between the ages of 11 and 14 who are preparing to enter secondary school in T&T. The examination assesses a student's mastery of the primary school curriculum in Language Arts, Mathematics, Creative Writing, Physical Education, Art and Music. A broader discussion of the SEA can be found in Chapter 6. Because analyses focused on participants' performance on Language Arts and Mathematics, a description of the two subject areas will be provided here.

SEA Language Arts. The Language Arts items measure skills such as vocabulary, spelling, comprehension, grammar, punctuation, and graphic representation (MOE, 2014). Students are expected to show how well they have grasped the mechanics of language. Also they must demonstrate how well they can read, summarize key points, understand, and make inferences from written text. The test also assesses skills in written expression (MOE 2014). A copy of the SEA Language Arts 2013 (MOE, 2013) examination paper can be found in Appendix B.

SEA Mathematics. The 46 item Mathematics exam consists of three sections. It measures skills in math concepts, calculation, algorithmic thinking and problem solving

(MOE, 2014). Items test knowledge of numbers, geometry, statistics, measurement and money. Students are awarded for any correct steps taken towards solving the problems and are therefore encouraged to show their notes and workings on the exam paper (MOE, 2014). A sample of the Mathematics 2014 (MOE, 2014) examination paper is provided in Appendix C.

Scores in Language Arts and Mathematics range from 1 to 100, but are given weights when calculating the total score. Weights are assigned according to the emphasis of each subject in the curriculum (Mathematics = 5 and Language Arts = 2; MOE, 2014).

The SEA scores were found to display strong negative skew. This means that the data clustered towards the higher end of performance on the Mathematics and Language exams. This is not surprising in light of the negative skew of the Full Scale IQ distribution. Transformation techniques were unsuccessful, so both SEA variables were eventually transformed into five level categorical variables: 1 = 0-30 marks; 2 = 31-50 marks; 3 = 51-70 marks; 4 = 71 -85 marks; and 5 = 86-100 marks.

Academic Performance Index (API). School Performance scores were obtained using the MOE's Academic Performance Index (API). The API is a metric that is based on the National Test scores of individual children. Within a particular school, each child is assigned a weighting representing his or her performance on the subject that is tested. The top ranking Level 4, which demonstrates excellent performance is assigned a weighting of 1.4. Subsequent lower levels from 3 to 1 are assigned weightings of 1.0, 0.6, and 0.2 respectively; and students who are absent for the test are given a weighting of 0 (DeLisle, Smith, Lewis, Keller et al., 2009). Second, the percentage of students achieving each performance rating within their school is determined. The API score of the school for each subject is calculated by multiplying each weighting achieved by the percentage of students achieving the weighting. The products for all weightings are then summed

(DeLisle, Smith, Lewis, Keller et al., 2009). For example, if half of the students are excelling and half are performing at a weighting of .6 on Mathematics, the school receives a score of 100 points ((1.4 x 50) + (.6 x 50)) in that subject.

It follows that a school's performance can fall between 0 and 140 points on any given subject; and because there are 4 subject areas in total, the final points for all areas are added together to arrive at an API score of between 0 and 560 points (DeLisle, Smith, Lewis, Keller et al., 2009).

Procedure

Data collection was done over a period of 25 months between October 2012 and September 2014. Tests were administered by author of the study who has over 20 years of experience in WISC-III and WISC-IV administration, and by 4 trained graduates of the University of the West Indies, Trinidad and Tobago, Clinical Psychology Masters Degree programme. The trained graduates had at least one year of taught courses as well as practical training and work experience in the administration of the WISC-IV (US). The author performed over 1/3 the test administrations and scoring.

Training. Prior to the start of data collection, the author provided additional training to the trained graduates in order to review standardized testing procedures and scoring and to clarify guidelines for data collection. The author trained testers to follow standardized guidelines of test administration as have been set out in the manual of the WISC-IV (US) (Wechsler, 2003b) with no deviation. After the training session was completed, the testers were debriefed and provided with a list of participants. Each tester was assigned to either a school or a private office in which they would test the participants. The testers were advised that all administrations should occur in one session that could last between 45 and 120 minutes, with 5 minute breaks every 45 minutes during the session.

Each tester was required to make arrangements with the school prior to each testing day so that they can confirm which students would be available. Testing took place within both the morning and afternoon sessions. Principals were asked to provide a quiet and private room for testing, that was well lit, air-conditioned and with minimal distractions or disturbances. Those administrations that were done in a school setting were paused during recess or lunch breaks when the school was noisiest. Private office testing was done in a clean cool room with minimal distractions. Participants' responses were recorded verbatim on WISC-IV assessment protocols. Protocols were scored and submitted with other additional testing materials in a sealed envelope.

Interrater agreement. Twenty of the assessment protocols were double-scored by different testers. The subtests selected for double scoring were also used in the Wechsler (2003b) inter-scorer agreement studies and consisted of items that required open ended responses (Similarities, Vocabulary and Comprehension). Testers were directed to transfer verbatim responses to an unscored protocol which was then scored by another tester. Inter-scorer agreement was calculated using Cronbach's alpha. As in the standardization studies, interrater reliability coefficients were calculated using the intra-class correlation method recommended by McGraw and Wong (1996) on SPSS 21. Inter-scorer reliabilities were as follows: .98 for Similarities, .99 for Vocabulary, .97 for Comprehension. These values are highly comparable with those obtained by Wechsler (2003b) (.98 for Similarities, .98 for Vocabulary and .95 for Comprehension).

Data Protection and Ethical Considerations. Prior to testing, children were informed about the study and the testing procedures. They were informed that they were free to withdraw if they were no longer willing to participate at any point in time. Parents and participants were informed that all information obtained were confidential; and once data input was complete, their names would be removed from any spreadsheets. They were also reassured that assessment materials would be discarded carefully at the end of the study; and additionally, a link with a report on the findings of the study will be emailed to all parents upon completion.

Key ethical concerns regarded the susceptibility of the participant to some psychological upset, anxiety or extreme discomfort during testing. As such, participants with psychological disorders such as depression or anxiety were excluded from the study. Furthermore, testers were reminded to stop testing if the participant displayed any signs of distress (e.g. crying).

Data Analysis and Chapter Summary

A number of statistical techniques were employed to meet the aims of this thesis. A pilot study was conducted in Chapter 2 to assess the validity of the WISC-IV four factor model in a sample of referred T&T students. Independent and multi-sample analyses helped identify the best fitting model and also determined model equivalence between high and low SES subsamples. Based on the arguments provided and the results of the pilot study, the author designed and conducted 8 additional studies. Chapter 4 will describe the results of a series of ANOVAs that were used to compare the mean performance of the sample on the different WISC-IV subtests and composites. In Chapters 5 and 6, model testing was accomplished with the use of exploratory and confirmatory factor analyses; multiple regression and logistic regression to separately explore key components of the model. In Chapter 7 structural equation modelling and logistic regression were done to investigate the fit of the full interpretive model.

Chapter 4: Descriptive statistics and comparisons of means. Chapter 4 addressed the first aim of this thesis by making comparisons between the T&T and US standardization samples. Means and standard deviations of the WISC-IV subtest and composite scores are presented. Skewness and kurtosis values were calculated, and frequency distributions were examined to assess for normality. ANOVAs were performed to compare WISC-IV composite mean scores within the samples. Results were summarized as F statistics, *p* values and effect sizes in a separate table. Probability values equal to or lower than .05 were considered to be indications of a significant difference between subtest mean scores.

Using data obtained from the WISC-IV technical manual (Wechsler, 2003b), the author compared the T&T and US standardization sample on the 10 core subtests scores. Because the US standardization raw data were not used in this study, comparisons of mean scores and standard deviations were done by calculating between- sample difference scores and discussing any trends.

Chapter 5: Assessing the validity of the current four-factor model. The 2nd aim of the thesis was addressed in Chapter 5 through 3 studies that focused on assessing the structural validity of the current 4 factor structure of the WISC-IV. To this end, inter-correlation analysis and exploratory and confirmatory factor analyses were performed.

Inter-correlation analysis. Analyses of the zero order correlations among the WISC-IV (US) 10 core subtests were conducted in Study 2. Inter-correlation matrices were visually examined to determine how the subtests were organized.

Exploratory factor analysis (EFA). In Study 3, EFA was done using SPSS 21 to summarize into factors the correlational matrix of the 10 core subtests. Two approaches were used. The first method which included principal axis factoring with oblique rotation was used by Wechsler (2003b) in the standardization study to confirm the 4 factor structure of the WISC-IV (US). Oblique rotation is recommended when factors are hypothesized to be inter-correlated (Kline, 2000). The second approach, the Schmid & Leiman (1957) orthogonalization method, was used by Watkins et al. (2006) in the study which investigated the hierarchical factor structure of the WISC-IV (US). This higher order analysis which was conducted for this thesis by using R Psych package, examines the correlational matrix of the factors to extract a higher order factor (Schmid & Leiman, 1957, Watkins et al., 2006). The results of the first order EFA were presented in a pattern matrix table of the factor loadings of each subtest. The results of the higher order EFA were summarized in tabular format as the loadings of each subtest on all factors, the variance in each subtest that is explained by the factors, the communality values, and the uniqueness values. These results for the T&T sample were compared to the results of the Watkins et al. (2006) study.

Confirmatory factor analysis (CFA). Confirmatory factor analyses (CFAs) were carried out in Study 4, by using maximum likelihood estimation in MPlus 6 to determine which of the alternative models fit the sample data best. The analyses compared the current 4-factor model with single, two and three oblique factor models, and 4 higher order models.

The proportion of variance in each of the 10 subtests, attributable to the first order and higher order factors as well as subtest specific variance and error variance were calculated. Absolute fit of each model was evaluated using the normed chi square (χ^2/df). Additional fit statistics were used to test model adequacy such as the Aikake Information Criterion (AIC), the comparative fit index (CFI), the Tucker Lewis index (TLI), the goodness of fit index (GLI), the adjusted goodness of fit index (AGFI), the standardized root mean square residual (SRMR), and the root mean square of approximation (RMSEA) with confidence intervals and probability values (Floyd and Widaman, 1995; Vrieze, 2012). Good model fit is reflected in normed chi-square values below 2 and *p* values over .05 (Bollen, 1989), CFI, TLI, GLI, and AGFI values in excess of .95, SRMR lower than .08, and RMSEA values lower than .06 (Hu & Bentler, 1998 & 1999; Hopper, Coughlan & Mullen; 2008).

Since the compared models were non-nested models, chi-square difference testing is not generally recommended, in spite of the fact that it has been used in the past to determine if alternative models result in a significant increase in model fit (Bentler & Satorra, 2010). Instead, the Aikake Information Criterion (AIC) has been offered as a useful method for determining the best fit among competing models (Vrieze, 2012), the best fitting model being the one with the smallest AIC value (Watkins, 2010).

Chapter 6: The relationship between environmental factors and performance on the WISC-IV (US). Chapter 6 is concerned with the 3rd and 4th aims of this thesis. First the chapter seeks to identify the environmental variables associated with the WISC-IV (US) performance in T&T children. In Study 5, analyses of the zero order correlations between environmental variables and the 5 WISC-IV (US) composite scores (VCI, PRI, WMI, PSI and FSIQ) were performed. Five multiple regression analyses, one for each of the following WISC-IV dependent variables were conducted. Results were presented in tables which summarized the variance explained by each model, as well as the F statistic. Additionally, probability, *Beta* and t values for each variable in the model were summarized.

The relationship between IQ and environmental variables and performance on the SEA. Study 6 of the chapter examined the relationship between academic achievement and the independent variables of global ability and quality of formal education. With categorical dependent variables, logistic regression was chosen as the most appropriate method for analysis. Logistic regression is recommended for use when seeking to predict a discrete outcome (Tabachnick & Fidell, 2013). In this case, two sequential multinomial logistic regression analyses were conducted using SPSS 21 to investigate the predictors of one of five levels of academic performance in SEA Language Arts and SEA Mathematics

Unlike multiple regression, logistic regression analyses assess the goodness of fit of the model by comparing the hypothesized model to a constant with no predictors. As such, probability values for the χ^2 should reflect a statistically significant difference between the compared models ($p \leq .05$) (Tabachnick & Fidell, 2013). The tables present regression coefficients, Wald statistics, standard error, degrees of freedom and odds ratios with confidence intervals for all predictors. Pseudo R^2 values will also be reported.

Chapter 7: Investigating the fit of the cross-cultural interpretive model.

Chapter 7 describes a study assessing the fit of a cross-cultural interpretive model of the WISC-IV (US) which is the 5th research question put forward in this thesis. In Study 7, Structural Equation Modelling (SEM) was performed for the purpose of assessing the fit of the WISC-IV (US) cross cultural interpretive model. Analyses were done using MPlus 6. SEM is highly recommended for exploring a model that includes combined structural and measurement components (Tabachnick & Fidell, 2013). SEM uses multiple techniques employing CFA and regression to explore the path model hypothesis. For this particular model, CFAs were performed to investigate the structural validity of the WISC-IV (US) model, and regression analyses are used to determine the relationship between the measured environmental variables and the latent ability factors (Muthén & Muthén, 1998 – 2010).

For the CFA, comparisons of alternative non-nested models were done. Model fit was determined using maximum-likelihood estimation: χ2 and probability values; CFI; TLI; RMSEA with confidence intervals, SRMR and AIC values. The extended model was also assessed using the model fit indices, however, to identify the most parsimonious model, the full model was constrained gradually to its most simplified form. Then using chi square difference testing (χ2 diff) testing, each model was compared to the preceding, more complex, model to determine whether adding additional parameters would explain the data significantly better than a more simplified model.

In Study 8, as an additional attempt to assess the validity of the cross cultural interpretive model, global ability factor scores were calculated and entered into the models which were then assessed for fit. First, CFAs were done for four competing models. Factor scores were saved using MPlus6 Regression Method (Muthén & Muthén, 1998-2010) and then entered into four separate regression models as predictors of SEA Language Arts and SEA Mathematics. The regression models were assessed using simple logistic regression in SPSS 21 and compared for best fit.

The final chapter, Chapter 8, provides a summary of the findings of this thesis and discusses the implications for future research and practice.

Summary

The aim of this thesis is to assess the validity of a cross cultural interpretive model of the WISC-IV (US) in a sample of Trinidad and Tobago (T&T) children. Nine studies were undertaken. The pilot study was described in Chapter 2 and the results of the additional 8 studies will be described in chapters 4, 5, 6 and 7. Chapter 4 will describe the performance of the T&T sample on the WISC-IV (US) and make comparisons with the US standardization sample. Chapter 5 will present the results of factor analyses done to investigate the validity of the current 4 factor model of the WISC-IV (US). Chapter 6 will investigate the external validity of the WISC-IV (US) in two parts. First regression analyses will explore the relationship between environmental variables (such as socio-economic status, schooling, and US exposure) and WISC-IV (US) performance. Second logistic regression analyses will identify the environmental and latent variable predictors of academic achievement in the sample. In Chapter 7, structural equation modelling will be used to determine the fit of a proposed cross-cultural interpretive model of WISC-IV measured intelligence. Chapter 8 will provide a summary and discussion of the findings of this thesis.

CHAPTER 4

EXAMINING THE PERFORMANCE OF THE T&T SAMPLE

This chapter presents the results of analyses that were conducted to examine the performance of the T&T sample on the WISC-IV (US). Frequency distributions were inspected for normality and the mean scores and standard deviations of all 10 core subtests and 5 ability composites were summarized. Mean scores of the T&T sample were compared with corresponding data from the WISC-IV (US) standardization sample. Additionally, within sample comparisons of means were done to determine if there are any meaningful inter-subtest or inter-composite discrepancies in performance.

Data Screening

Pre-analysis data screening for accuracy, missing values, outliers and normality was performed using SPSS 21. Three subtests, Pictures Concepts, Vocabulary, and Letter Number Sequencing (LNS); and three composites, Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI) and Full Scale IQ (FSIQ) displayed significant negative skew. After identifying and recoding outliers as was recommended by Field (2009), *z*-scores for Picture Concepts, Letter Number Sequencing, VCI and PRI approached the upper limit of significance (*z*-score = 3.29, p<.001) for skewness but did not exceed it. Platykurtic distributions, which are distributions that are flatter than the normal curve, were identified for all but the Block Design, Digit Span, Coding and Symbol Search subtests, but their z score values were non-significant for kurtosis.

Researchers are cautioned against putting too much value on *z*-scores when sample sizes exceed 200. It is recommended that visual inspections of frequency distributions as well as absolute skewness and kurtosis values should guide decisions about sample normality (Field, 2009; Kim, 2013). Absolute skewness and kurtosis values of >2.1 and >7.1 respectively reflect a significant departure from normality (West, Finch & Curran, 1995). Based on all the aforementioned criteria, all subtest and index frequency distributions are considered to be satisfactory for normality. Values for skewness and kurtosis with their respective standard errors are presented in Table 4.1. Frequency distributions for WISC-IV (US) subtests are presented in Figure 4.1, and the distributions for the composites comprise Figure 4.2.

Table 4.1.

WISC-IV (US) IQ scores and skewness and kurtosis values for Trinidad & Tobago (N=203) sample (11 – 12 years)

	Mean	SD	Skewness	SE	Kurtosis	SE
Block Design	8.4	3.0	-.22	.17	.02	.34
Similarities	8.6	3.4	-.13	.17	-.37	.34
Digit Span	9.0	3.2	.38	.17	.42	.34
Picture Concepts	9.6	3.3	-.55	.17	-.41	.34
Coding	7.0	2.7	.15	.17	-.12	.34
Vocabulary	8.9	3.2	-.49	.17	-.30	.34
Letter Number Sequencing	8.9	3.0	-.55	.17	-.28	.34
Matrix Reasoning	9.0	3.2	-.39	.17	-.15	.34
Comprehension	8.0	3.0	-.39	.17	-.00	.34
Symbol Search	8.3	3.3	-.24	.17	-.22	.34
Verbal Comprehension Index	90.8	16.7	-.51	.17	-.01	.34
Perceptual Reasoning Index	93.7	15.6	-.52	.17	-.29	.34
Working Memory Index	93.5	15.4	-.21	.17	.49	.34
Processing Speed Index	87.3	15.1	.10	.17	.13	.34
Full Scale IQ	89.8	15.5	-.40	.17	-.30	.34

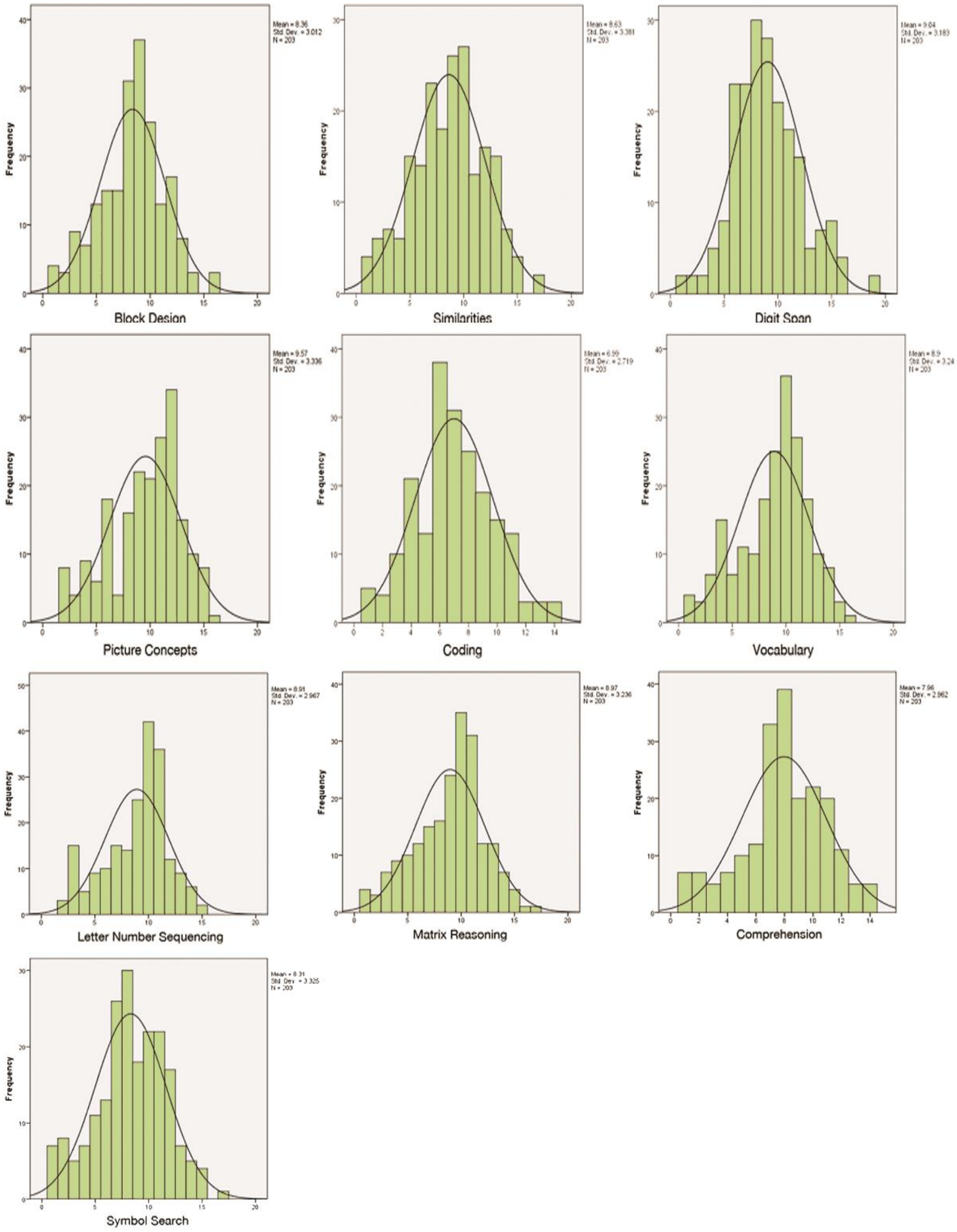


Fig 4.1. Frequency distributions of WISC-IV (US) core subtests Block Design, Similarities, Digit Span, Picture Concepts, Coding, Vocabulary, Letter Number Sequencing, Matrix Reasoning, Comprehension, Symbol Search.

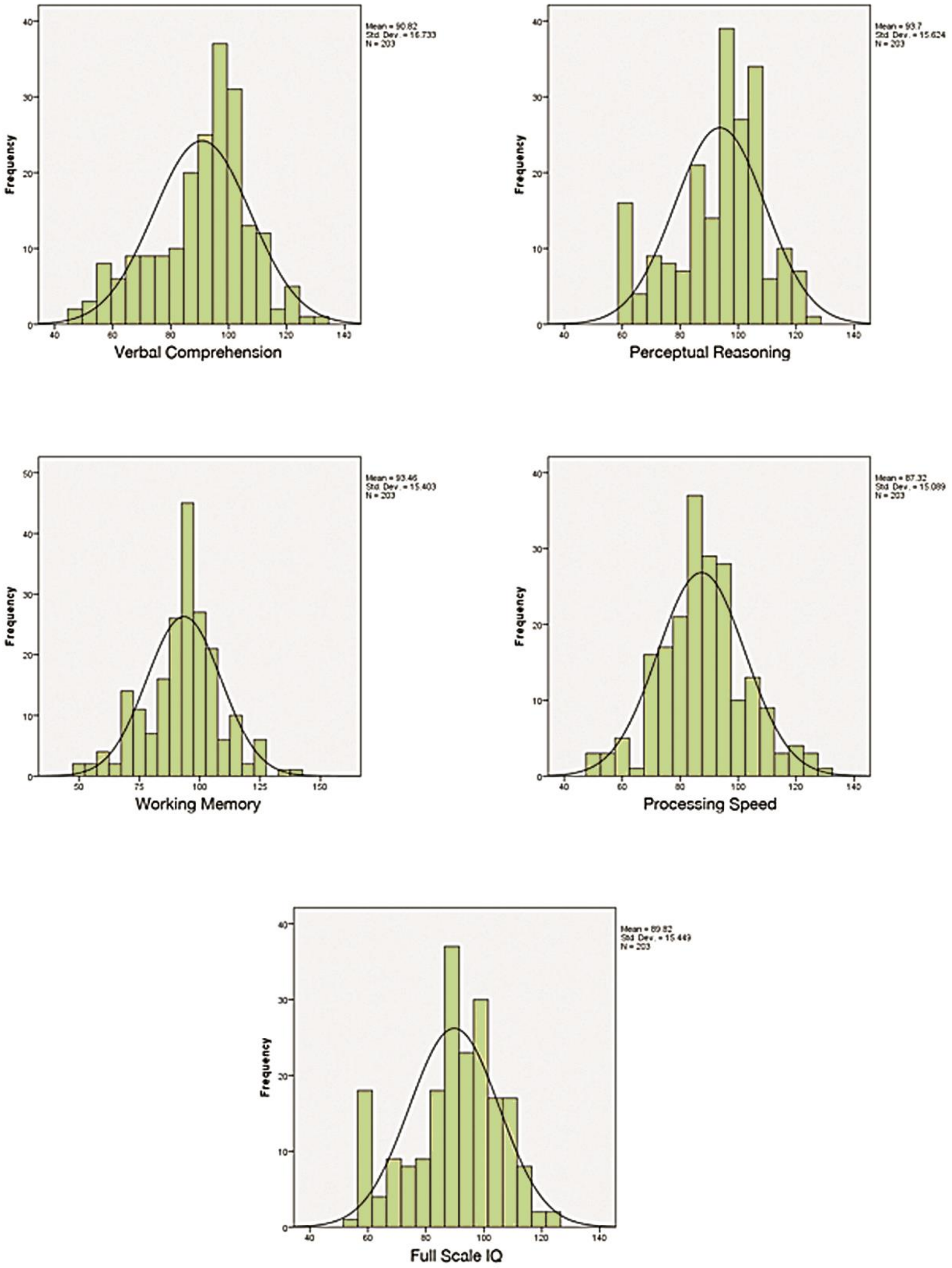


Fig 4.2. Frequency distributions of WISC-IV (US) composites: Verbal Comprehension, Perceptual Reasoning, Working Memory, Processing Speed and Full Scale IQ.

Study 1

The T&T sample was stratified according to age. WISC-IV (US) means and standard deviations were then computed for each group and compared with the means and standard deviations of 11 and 12 year olds of the US standardization sample. Analyses of variance between groups and calculations of effect sizes were not possible in this case because WISC-IV (US) standardization raw data was not available for this study.

Repeated Measures ANOVAs were then performed with SPSS 21 to compare the performance of the participants across the WISC-IV (US) subtests and composites. Mauchley's tests of sphericity were performed to determine whether the variances of the differences between all WISC-IV subtests and composites were equal. The result of violations of assumptions of sphericity is reduced power, and so corrections are applied in order to reduce the probability of Type II error.

Results

Between Group Comparisons of WISC-IV (US) performance - (T&T vs US children)

Demographic data for the T&T sample are presented in Tables 3.1 and 3.2 of Chapter 3. The WISC-IV (US) standardization sample was equally divided on the basis of gender (Wechsler, 2003b). Exact frequency values for the demographic variables of Parental Education and Ethnicity were not available for the full US sample; however approximate frequencies are presented graphically in the WISC-IV (US) Technical and Interpretive Manual (Wechsler, 2003b). Additional demographic data for the US sample by age, ethnicity and parental education, age, gender and parental education, and age, gender and ethnicity also can be found in the manual (Wechsler, 2003b).

Means and standard deviations of the WISC-IV (US) subtests and composites of the T&T sample and the US standardization sample, are presented for the 11 and 12-year old subgroups in Tables 4.2 and 4.3, respectively.

Table 4.2.

WISC-IV (US) IQ scores: Trinidad & Tobago (N=203) sample and United States standardization sample (N = 200) (11 years)

	Trinidad & Tobago		United States	
	Mean	SD	Mean	SD
Block Design	8.6	3.0	10.1	3.1
Similarities	9.1	3.5	9.9	2.9
Digit Span	9.2	3.4	9.9	3.1
Picture Concepts	9.5	3.4	10.2	2.8
Coding	7.1	2.5	10.2	3.0
Vocabulary	9.2	3.4	10.0	2.8
Letter Number Sequencing	9.0	3.0	10.1	3.0
Matrix Reasoning	9.0	3.1	10.0	3.0
Comprehension	8.2	3.1	9.9	2.8
Symbol Search	8.5	3.3	10.1	2.8

Table 4.3.

WISC-IV (US) IQ scores: Trinidad & Tobago (N=203) sample and United States standardization sample (N=200) (12 years)

	Trinidad & Tobago		United States	
	Mean	SD	Mean	SD
Block Design	8.1	3.0	10.0	3.1
Similarities	8.1	3.2	10.0	2.9
Digit Span	8.9	3.0	9.9	3.0
Picture Concepts	9.6	3.2	10.1	3.1
Coding	6.9	2.9	9.9	3.0
Vocabulary	8.9	3.0	10.1	3.1
Letter Number Sequencing	8.7	3.3	10.0	3.1
Matrix Reasoning	9.0	3.4	10.1	2.9
Comprehension	7.7	2.8	10.2	3.1
Symbol Search	8.1	3.4	9.9	3.0

The tables demonstrate that for both age groups, T&T children obtained lower scores than their US counterparts on all subtests. With the exception of the Coding subtests, mean differences did not exceed 1 Standard Deviation. The biggest gaps in performance between groups were on the Coding and Comprehension subtests; while the smallest disparities in performance were on the Picture Concepts and Matrix Reasoning subtests.

Within-Group Comparisons of Means in the T&T Sample

Repeated Measures ANOVAs were conducted to determine whether the differences among subtest and index mean scores are statistically significant. Results demonstrated statistically significant differences in mean performance among WISC-IV (US) subtests as well as among the indices. The Mauchley's test of sphericity revealed a violation of the sphericity assumption, $\chi^2 = 28.63$, $p < .001$. Greenhouse-Geisser corrections therefore were employed to reduce Type II error.

Subtest mean score comparisons. The results of the main ANOVA revealed that there were significant differences among the 10 core subtests, *F* (6.93, 1399.12) = 20.65, $p < .001$ (See Table 4.4). Post hoc contrasts demonstrated that Coding mean scores were significantly lower than all other subtests with effect sizes (Cohen's *d*) ranging from medium to large (.35 to .85; See Table 4.5).

Table 4.4

Comparisons of Subtest Means using Repeated Measures ANOVA – Trinidad & Tobago sample (N = 203)

	SS	*df*	MS	F	*P*	η_p^2
Subtests	933.40	6.93	134.76	20.64	.000	.09
Error	9131.50	1399.12	6.53			

Note: SS = Sum of Squares, *df* = degrees of freedom, MS = Mean Square, F = F statistic, *p* = probability values; η_p^2 = partial eta squared.

Table 4.5.

Statistically Significant Pairwise Comparisons of 10 core subtests – Trinidad & Tobago sample (N=203)

				95% CI			
		Mean Diff	Std. Error	Lower Bound	Upper Bound	*p*	Cohen's *d*
Coding							
	Block Design	-1.37	.23	-1.83	-.92	.000	.48
	Similarities	-1.64	.26	-2.16	-1.13	.000	.50
	Digit Span	-2.06	.25	-2.56	-1.56	.000	.70
	Picture Concepts	-2.58	.25	-3.07	-2.09	.000	.85
	Vocabulary	-1.92	.23	-2.39	-1.45	.000	.67
	LetterNumberSeq	-1.93	.23	-2.38	-1.47	.000	.69
	Matrix Reasoning	-1.98	.24	-2.45	-1.51	.000	.66
	Comprehension	-.98	.22	-1.41	-.54	.000	.35
	Symbol Search	-1.33	.19	-1.71	-.95	.000	.44
Picture Concepts							
	Block Design	1.21	.21	.79	1.62	.000	.38
	Similarities	.94	.21	.53	1.35	.000	.26
	Digit Span	.52	.24	.05	.99	.031	.52
	Vocabulary	.67	.22	.24	1.10	.003	.20
	LetterNumberSeq	.67	.20	.25	1.06	.002	.21
	Matrix Reasoning	.60	.22	.17	1.03	.007	.18
	Comprehension	1.61	.21	1.12	2.02	.000	.51
	Symbol Search	1.26	.26	.75	1.77	.000	.38
Comprehension							
	Similarities	-.67	.16	-.99	-.34	.000	.20
	Digit Span	-1.08	.23	-1.54	-.63	.000	.35
	Vocabulary	-.94	.15	-1.24	-.65	.000	.32
	LetterNumberSeq	-.95	.19	-1.33	-.57	.000	.32
	Matrix Reasoning	-1.01	.21	-1.42	-.59	.000	.33
Symbol Search							
	Digit Span	-.73	.26	-1.25	-.22	.005	.22
	Vocabulary	-.59	.23	-1.05	-.13	.012	.19
	LetterNumberSeq	-.60	.25	-1.09	-.11	.017	.19
	Matrix Reasoning	-.66	.26	-1.17	-.14	.013	.20
Block Design							
	Digit Span	-.69	.24	-1.15	-.22	.004	.22
	Vocabulary	-.54	.22	-.97	-.12	.013	.17
	LetterNumberSeq	-.61	.21	-.96	-.15	.008	.18
	Matrix Reasoning	-.61	.19	-.99	-.22	.002	.19
Vocabulary	Similarities	.28	.14	.01	.54	.042	.04

Comprehension was the 2nd weakest area of performance for the sample. Differences between Comprehension and the remaining subtests ranged from small to moderate. The sample performed best on the Picture Concepts test. Small to medium effects were observed for other significant mean score differences and are summarized in Table 4.5.

Index mean score comparisons. Results of the main ANOVA revealed significant differences among the four ability indices, *F* (2.73, 550.77) = 16.59, *p*< .001 (See Table 4.6). Tests of simple effects showed that the sample performance on the VCI and PSI composites was significantly lower than on the WMI and PRI composites. Effect sizes for differences between PSI and the WMI and PRI composites were moderate (.40 and .41 respectively). Other effect sizes were small (See Table 4.7).

Table 4.6

Comparisons of Index Means using Repeated Measures ANOVA – Trinidad & Tobago sample (N = 203)

	SS	*Df*	MS	F	*P*	η_p^2
Composites	5384.32	2.73	1974.77	16.59	.000	.08
Error	65556.18	550.77	119.03			

Note: SS = Sum of Squares, *df* = degrees of freedom, MS = Mean Square, F = F statistic, *p* = probability values; η_p^2 = partial eta squared.

Table 4.7.

Statistically Significant Pairwise Comparisons of 4 composites – Trinidad & Tobago sample (N=203)

				95% CI			
		Mean Diff	Std. Error	Lower Bound	Upper Bound	*P*	Cohen's *d*
VCI							
	PRI	-2.88	.87	-4.60	-1.16	.001	.18
	WMI	-2.64	.96	-4.53	-.75	.006	.16
	PSI	3.50	1.13	1.27	5.73	.002	.23
PSI							
	PRI	-6.38	1.13	-8.62	-4.15	.000	.41
	WMI	-6.14	1.15	-8.42	-3.87	.000	.40

Note. VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index.

Discussion

Explaining Differences in Mean IQ Scores

Examinations of frequency distributions revealed a clustering of data in the upper end of the range of IQ scores. These findings are not surprising in light of the disproportionate number of higher SES participants that volunteered for this study. The smaller proportion of very low scoring outliers resulted in mean scores that were considerably lower than the median values. Even after recoding the outliers, the T&T sample subtest and index averages were below the established means for the WISC-IV (US) (10 and 100 respectively). Also, the T&T children performed below their US counterparts on all WISC-IV (US) subtests and composites. In the absence of raw standardization data, these between group discrepancies in performance could not be analysed for significance; however, the findings are considered to be worthy of discussion as they mirror those of previous studies on cross-cultural differences in WISC performance (See Chapter 2).

The biggest gap between national samples was on the Coding subtest which is a measure within the Processing Speed Index (PSI). International studies have identified significant differences between national samples on processing speed (e.g. Neubauer & Benischke, 2002; Cores, Vanotti, Eizaguirre, Fiorentini, et al., 2015). Performance discrepancies may reflect differences in the way cultures value time and speed. This argument is persuasive, especially in light of findings that work speed and walking pace is slower in less developed and collectivistic societies as well as in warmer climates (Levine & Norenzayan, 1999). Levine and Bartlett (1984) also found that people in more populated cities walked at a faster average pace than in less populated cities. To the author's knowledge, the relationship between walking speed and processing speed is not known. However, investigations into the correlations between measures of walking speed and processing speed is an important topic for research. At this time, 'pace of life' explanations for the T&T sample's Coding performance will only be speculative. Further exploration of the 'pace of life' hypotheses is beyond the scope of this thesis, but future research in this area is encouraged. Studies specifically should examine the extent to which environmental and socio-cultural variables such as population density, average temperature, developmental indices and collectivism account for cross national differences in processing speed scores.

In addition to the between sample analyses, the results of the within sample subtest mean score comparisons identified Coding as the area of most weakness for the T&T sample. Difference scores between Coding and the other 9 core subtests were significant with moderate to large effect sizes. These findings may be explained by examining the composition of the T&T sample, in which 1/4 of the cases reported a diagnosis of at least one learning disability. In the absence of the relevant population statistics, it is not clear how well the T&T sample frequencies and population frequencies match on this variable. However, with the prevalence of diagnosed learning disabilities in the US, Canada and the UK between 3 and 5 percent (Statistics Canada, 2007; Public Health England, 2013; National Center for Learning Disabilities, 2014); one can hypothesize that the T&T sample is more compatible with clinical or referred samples than with the general population. If this is the case, then lower performance on measures of processing speed can be expected in this sample, similar to the lower performance on the WISC in children diagnosed with ADHD, Autism or Traumatic Brain Injury (eg. Calhoun & Mayes, 2005; Mayes & Calhoun, 2006; Solanto, Gilbert, Raj, Zhu, et al., 2008; Goldstein, Allen, Minshew, Williams, et al., 2008; Allen, Thaler, Donohue & Mayfield, 2010).

In order to test the learning disability hypothesis, within group subtest means comparisons were undertaken after removing the 51 learning disability cases from the dataset. Table 4.8 provides the subtests means for the T&T samples before and after removing the learning disability cases. Also, means for the US standardization sample and a sample of US children with mixed Reading, Mathematics and Written Expression learning disabilities (WISC-IV Special Group Studies; Wechsler, 2003b) are provided in Table 4.8.

Examinations of subtest means show that the T&T ‘learning disability’ and ‘no learning disability’ samples performed similarly on the Coding subtest. Additionally, both the T&T learning disability and no learning disability samples had lower Coding scores than the US learning disability and standardization samples respectively. In the absence of raw data for the US samples, these differences were not tested for significance.

Table 4.8

Sample Means Comparisons of T&T sample with Learning Disabilities and US Learning Disability sample, and T&T sample without Learning Disabilities with US Standardization Sample.

	T&T - LD (N = 203)	US - LD (N = 34)	T&T - NLD (N = 152)	US - STD (N= 2200)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Block Design	8.1 (3.0)	8.5 (2.5)	8.4 (3.0)	10.0 (3.1)
Similarities	8.1 (3.2)	7.9 (2.3)	8.7 (3.5)	10.0 (2.9)
Digit Span	8.9 (3.0)	7.9 (2.6)	9.1 (3.2)	9.9 (3.0)
Picture Concepts	9.6 (3.2)	8.7 (3.0)	9.7 (3.5)	10.1 (3.1)
Coding	6.9 (2.9)	8.0 (2.4)	7.0 (2.7)	9.9 (3.0)
Vocabulary	8.9 (3.0)	8.0 (2.2)	8.9 (3.3)	10.1 (3.1)
Letter Number Sequencing	8.7 (3.3)	8.6 (3.0)	8.9 (3.2)	10.0 (3.1)
Matrix Reasoning	9.0 (3.4)	8.0 (2.2)	9.0 (3.4)	10.1 (2.9)
Comprehension	7.7 (2.8)	8.7 (2.5)	8.0 (3.0)	10.2 (3.1)
Symbol Search	8.1 (3.4)	8.6 (2.8)	8.1 (3.4)	9.9 (3.0)

Note: T&T – LD = T&T sample with learning disabilities, US – LD = US sample with learning disabilities, T&T – NLD = T&T sample without learning disabilities, US – STD = US standardization sample.

The results of the Repeated Measures ANOVA revealed that even after removal of the learning disability cases, there were statistically significant differences among the 10 core subtests (F (6.85, 1020.61) = 16.34, $p< .001$; Table 4.9)[10]. Moreover, post hoc contrasts found that the mean of the Coding subtest was significantly lower than the other subtests. Similar with the full sample, effect sizes ranged from medium to large (.33 to .86). Results therefore do not support the learning disability hypothesis. Table 4.10 summarizes the between subtest mean difference values, standard error, p values and effects sizes for comparisons made between Coding and other subtests.

---

[10] Sphericity assumption violated: $\chi^2 = 28.63$, $p < .001$

Table 4.9

Comparisons of Subtest Means using Repeated Measures ANOVA – Trinidad & Tobago sample without Learning Disability cases (N = 152)

	SS	*df*	MS	F	*p*	η_p^2
Subtests	734.54	6.85	107.24	16.34	.000	.10
Error	6698.16	1020.61	6.56			

Note: SS = Sum of Squares, *df* = degrees of freedom, MS = Mean Square, F = F statistic, *p* = probability values; η_p^2 = partial eta squared.

Table 4.10

Statistically Significant Pairwise Comparisons of Coding with 9 core subtests – Trinidad & Tobago sample (N=203)

				95% CI			
		Mean Diff	Std. Error	Lower Bound	Upper Bound	*p*	Cohen's *d*
Coding							
	Block Design	-1.37	.27	-2.25	-0.48	.000	.48
	Similarities	-1.61	.31	-2.65	-0.57	.000	.51
	Digit Span	-2.03	.29	-3.01	-1.04	.000	.68
	Picture Concepts	-2.67	.29	-3.62	-1.71	.000	.86
	Vocabulary	-1.89	.28	-2.81	-0.96	.000	.63
	LetterNumberSeq	-1.88	.27	-2.78	-0.98	.000	.64
	Matrix Reasoning	-1.99	.28	-2.93	-1.05	.000	.65
	Comprehension	-0.95	.26	-1.80	-0.10	.012	.33
	Symbol Search	-1.09	.23	-1.85	-0.32	.000	.36

The sample's performance on Comprehension was the other main area of weakness – with scores that were significantly lower than the Similarities, Digit Span, Picture Concepts, Vocabulary, Letter Number Sequencing and Matrix Reasoning subtests. Typically, a poor performance on the Comprehension subtest is indicative of weak verbal reasoning ability, comprehension and expressive language skills (Wechsler, 2003b), as well as poor social knowledge and judgement (Flanagan & Kaufman, 2004). While the results may represent a true weakness in this aspect of verbal comprehension ability, it is possible that the responses obtained on the comprehension task reflect social conventions that are specific to T&T. Performance discrepancies between US and T&T samples may therefore be explained by differences in normative culture between groups. In US culture, the ability to assert oneself verbally is valued and nurtured in children. This might be why American children are observed to be quite outspoken and assertive. Conversely, T&T children are raised to be more reticent and inhibited in their interactions, particularly when they are in a formal context or in the presence of adults or authority figures. T&T children also are more likely than their American peers to demonstrate verbal restraint in oral testing scenarios. Since the WISC-IV (US) verbal comprehension question and answer format rewards a more open-ended type of response, it is not surprising that T&T children performed poorer than the US sample on this test. Interestingly on the other measure of verbal reasoning, Similarities, which employed a more close-ended response format, the performance of the T&T sample was considerably better. Qualitative examination of the response patterns of T&T children to open ended questions may help identify the factors associated with performance of T&T children on the Comprehension subtest.

On the Picture Concepts subtest, the analyses revealed that the sample's performance significantly exceeded that of the other 9 subtests with small to large effect sizes. The mean score on the test appeared to be almost comparable with the US standardization sample although this could not be determined statistically. Earlier it was predicted that the American content of the Picture Concepts test will negatively influence performance in children with little access to US culture. However, the findings did not support this hypothesis. Instead, the apparent 'cultural loading' of the Picture Concepts items did not appear to have as deleterious an impact on performance as expected.

Perhaps the effect of US Exposure was weakened as a result of selection bias. Because the sample consisted of a larger portion of high SES participants, it is posited that overall, the sample is likely to have greater than normal access to US information and therefore perform better on the test. To test this hypothesis, the T&T sample was stratified by US Exposure, following which Repeated Measures ANOVAs were conducted. Comparisons between Picture Concepts and the other subtests were made for each of the US Exposure groups. The Sphericity assumption was met only for the Very Little/None Exposure subsample, $\chi^2 = 59.03$, $p < .07$. Violations of sphericity assumptions were found for all other subsamples (Some: $\chi^2 = 163.86$, $p < .001$; Moderate: $\chi^2 = 120.09$, $p < .001$; Very Much: $\chi^2 = 115.29$, $p < .001$), and Greenhouse-Geisser corrections were applied. Results demonstrated that significant differences in performance across subtests were found for US Exposure as a main effect (See Table 4.11).

Table 4.11.

Comparisons of Subtest Means using Repeated Measures ANOVA – Trinidad & Tobago US Exposure subsamples.

	SS	*Df*	MS	F	*p*	η_p^2
Very Little/None (N=26)						
Subtests	148.14	9	16.46	3.45	.001	.12
Error	1073.46	225	4.77			
Some (N=108)						
Subtests	579.67	6.68	86.74	12.28	.000	.10
Error	5050.13	715.06	7.06			
Moderate (N=52)						
Subtests	235.13	5.56	42.31	5.91	.000	.10
Error	2029.98	283.43	7.16			
Very Much (N=16)						
Subtests	176.23	4.01	43.90	3.45	.004	.12
Error	708.18	60.21	11.76			

Note: SS = Sum of Squares, *df* = degrees of freedom, MS = Mean Square, F = F statistic, *p* = probability values; η_p^2 = partial eta squared.

However, post-hoc contrasts found significant differences between Picture Concepts and other subtests only within the samples that had Some and Moderate Exposure to US Culture (Table 4.12). Additionally, difference values and effect sizes were larger for the Moderate Exposure sample than the Some Exposure sample.

Table 4.12.

Statistically Significant Pairwise Comparisons with Picture Concepts: Trinidad & Tobago US Exposure subsamples

				95% CI			
		Mean Diff	Std. Error	Lower Bound	Upper Bound	*p*	Cohen's *d*
Some Exposure (N=108)							
Picture Concepts	Block Design	1.29	.31	0.26	2.33	.002	.40
	Coding	2.57	.33	1.45	3.68	.000	.83
	Comprehension	1.57	.27	0.67	2.46	.000	.49
	Symbol Search	1.57	.36	0.38	2.77	.001	.46
Moderate Exposure (N=52)							
Picture Concepts	Block Design	1.71	.37	0.44	2.99	.001	.66
	Coding	2.85	.48	1.19	4.50	.000	.83
	LetterNumberSeq	1.40	.36	0.17	2.64	.012	.62
	Comprehension	1.54	.39	0.18	2.89	.012	.71

These findings support the notion that T&T children with higher US exposure perform better on the Picture Concepts subtest. There is an exception however in the very exposed sample, where there was no significant difference in performance between Picture Concepts and the other subtests. The lack of significant effect for US exposure at the highest level of the variable may due to the small size of the subsample. Another explanation is that there may be a non-linear relationship between the two variables. In other words, the effect of US exposure tends to taper off as children become more exposed, so that there is no difference in US knowledge between children with moderate levels and children with very high levels of exposure. Multiple regression analyses which will be further discussed in Chapter 6, are expected shed further light on the relationship between US Exposure and intelligence.

Conclusion

The present study evaluated the performance of the T&T sample on the WISC-IV (US). Distribution of scores, as well as means and SDs were computed separately for 11 and 12 year olds; and then compared to the performance of 11 and 12 year olds of the US normative sample. The study also compared the sample's performance among different subtests and composites. The results suggested that the sample of T&T children tested in this study performed below similarly aged US children on all tests. Additionally, moderate differences between the processing speed index and other composites were found. A speculative explanation that was offered for these processing speed differences was the so called 'pace of life' differences between people living in hot and cold environments, between those living in more and less densely populated cities, and between people from more and less developed countries. This hypothesis cannot be tested in this thesis, but it is considered a question worthy of future examination.

Within sample analyses demonstrated that the Comprehension subtest mean score was significantly lower than the other subtest mean scores. The small to moderate differences between Comprehension and the other subtests were argued to point to potential method bias issues for the WISC-IV (US). It was suggested that unique cultural values and socio-cultural expectations about the way children and adults communicate may have influenced the performance on this test. Future research should examine how cultural norms on verbal expression relate to performance on verbal tests with an open-ended response format.

A major limitation for means comparisons is selection bias, and this may have impacted the findings of this study. As an example, the within sample analyses revealed that the largest discrepancy in mean scores was between Coding and the 9 other core subtests. It was posited that the finding may be explained by the large percentage (25.1%) of children diagnosed with learning disabilities in the sample. To test this hypothesis, analyses were redone after removing the learning disability cases from the dataset; however, the results did not demonstrate any major reduction in between subtest difference scores or effect sizes.

To further illustrate the potential impact of selection bias on means comparisons, the sample's tendency towards relatively higher IQ scores on the Picture Concepts test, was argued to be associated with the disproportionately large number of higher SES participants in the study. Good performance on the Picture Concepts test was hypothesized to be influenced by the extent to which the children were exposed to US information. It was also argued that higher SES children were more likely to have access to US information through television, internet or travel. To address this question, the sample was stratified according to degree of US exposure, following which means comparisons between Picture Concepts and the other subtests were conducted. Results supported the argument that US exposure and Picture Concepts performance were positively associated although this effect was not observed at the highest level of US exposure. Results seemed to point to a non-linear relationship. Further tests of the association of WISC-IV (US)performance with US exposure in this sample will reported in Chapter 6.

CHAPTER 5

INTERNAL VALIDITY OF THE WISC-IV(US) IN THE T&T SAMPLE

This chapter presents and discusses the findings of three studies that were conducted to assess the internal validity of the WISC-IV (US) in a sample of T&T children. Study 2 examined the inter-subtest correlations to determine the size, direction and significance of relationships among the 10 core subtests of the WISC-IV (US). Study 3 employed two different techniques of exploratory factor analysis, oblique rotation and orthogonal rotation, to determine how the 10 core subtests are grouped based on the data from this sample. Study 4 used confirmatory factor analyses to compare the fit of the Wechsler (2003b) 4 factor model with alternative models. The results of these analyses contribute to the development of a framework for interpreting the scores of T&T children on the WISC-IV (US).

Study 2:

An Examination of the Inter-correlations among Core Subtests of the WISC-IV (US)

Though a large number of cross-cultural validity studies on the WISC-IV (US) have focused on how the correlation and covariance matrices aggregate into factors, few have reported on preliminary examinations of the correlation matrix. As with multiple regression analyses, examinations of the correlation matrix are considered to be useful for hypothesis building prior to Confirmatory Factor Analysis and therefore it is given some attention in this thesis.

In the WISC-IV (US) standardization study, it was hypothesized that all subtests will correlate to varying degrees with each other (Wechsler, 2003b). The strength of the relationships was predicted to depend on whether the subtests belonged to similar or different factors (Wechsler, 2003b). As an example, verbal comprehension subtests were expected to correlate more highly with each other than with working memory, perceptual reasoning or processing speed subtests. Also, Wechsler (2003b) predicted that tests with significant *g* loadings (e.g. Block Design, Similarities and Vocabulary) would inter-correlate highly. Furthermore, '*g* -loaded' tests within the same 1st order factor, were expected to correlate more highly with each other than with other tests in the same factor (Wechsler, 2003b). Picture Concepts was hypothesized to correlate highly with verbal comprehension subtests because verbal problem solving strategies were thought to aid performance on this perceptual reasoning subtest (Wechsler, 2003b).

Wechsler's investigations provided support for these hypotheses. The pattern matrix revealed weak to strong subtest inter-correlations overall with strong within composite correlations (*rs* between .10 and .75; See Table 5.1) (Wechsler, 2003b). These findings have been largely replicated in cross cultural studies. In a validation study for Mexico's WISC-IV, subtest inter-correlations that were significant and ranged from small to large (*rs* between .18 and .76), were also found (Fina, Sanchez-Escobedo & Hollingworth, 2012). Additionally, a Colombian study on the psychometric properties of the WISC-IV (Spanish) (Wechsler, 2004b), found significant and positive correlations among subtests, however these ranged from moderate to strong (.51 to .85) (Mejias-Contreras & Albarracin-Rodriguez, 2013).

For the Mexican and Colombian studies, within composite correlations were not consistent with the findings of the US standardization study. For the Mexican sample, Block Design correlated more strongly with Vocabulary (.60) than with Picture Concepts (.50). This finding may be explained by the mutual *g* loadings of Block Design and Vocabulary. Also, Picture Concepts correlated more strongly with Block Design than with Similarities (.51) and Vocabulary (.51). This finding may support Wechsler's (2003b) assumption that Picture concepts is verbally mediated. In the Colombian sample, Block Design demonstrated stronger correlations with Vocabulary (.72) as well as with Similarities (.78); and Picture Concepts correlated more strongly with Similarities (.68) than with Matrix Reasoning (.61). Also, the Coding and Letter Number Sequencing correlation (.76) exceeded that of Coding and Symbol Search (.70), while Letter Number Sequencing demonstrated a higher correlation coefficient with Vocabulary (.75) than with Digit Span (.70). The latter result suggests that level of education may play a role in the child's ability to understand instructions on this test and also to complete items. This point will be discussed more thoroughly later in the chapter.

The Wechsler study demonstrated moderate to strong correlations among the *g* loaded subtests (Block Design, Similarities and Vocabulary) (Wechsler, 2003b); and as was shown above, these findings were supported by the other cross-national data, except that highly *g* loaded Vocabulary and Comprehension were more strongly correlated than Vocabulary and the other strongly *g* loaded test, Similarities, in the Mexican and Colombian samples (Fina, Sanchez-Escobedo & Hollingworth, 2012; Mejias- Contreras & Albarracin-Rodriguez, 2013). Also perceptual reasoning subtests and subtests of the verbal comprehension and the working memory indices were shown to correlate moderately in all three studies (Wechsler, 2003b; Fina, Sanchez-Escobedo & Hollingworth, 2012; Mejias-Contreras, Albarracin-Rodriguez, 2013). The overlap between the verbal comprehension and perceptual reasoning subtests were attributed to their high *g* loadings (Fina, Sanchez-Escobedo & Hollingworth, 2012). Also, the moderate correlation between perceptual reasoning and working memory have found support in studies demonstrating the positive link between working memory and performance on measures of reasoning (Salthouse, 1992; Kyllonen, 1994; Fry & Hale, 1996; Engle, Tuholski, Laughlin, & Conway, 1999, Süß, Oberauer, Wittmann, Wilhelm and Schulze (2002).

Additional findings from the US, Mexican and Colombian studies that reflected strong associations among the processing speed tests and also moderate correlations with Block Design (Wechsler, 2003b; Wechsler, 2004b; Mejias-Contreras & Albarracin-Rodriguez, 2013), may be explained by a visual processing as a well as a motor skills factor (Wechsler, 2003b).

Analyses

To assess whether this pattern of results can be replicated in the T&T sample, bivariate correlation analyses of the 10 core subtests were conducted on SPSS 21 using the scaled scores of the T&T sample. The T&T sample which consisted of 11 and 12-year-old children was first stratified according to age. Correlation coefficients were then obtained for both samples; and using Box's M in SPSS 21, the correlation matrices for both groups were assessed for equality. The results indicated that pattern matrices of the two groups were not significantly different (p = .38), therefore the subtest correlational analyses could be conducted for the full sample. Bivariate correlation analyses were also conducted on the 5 ability indices for the full sample. Because index scores are calculated from converted scores and are therefore not age based, it was considered unnecessary to calculate composite correlations separately for each age group. Based on the results of the WISC-IV(US) inter-correlation analyses (Wechsler, 2003b), the following hypotheses were tested:

1. All 10 core subtests will correlate significantly with each other;
2. Within composite inter-correlations are higher than between composite inter-correlations;
3. Subtests with high *g* loadings, Block Design, Similarities and Vocabulary (Wechsler, 2003a), will demonstrate high inter-correlations;

4. Similarities and Vocabulary will correlate more highly with each other than with the other Verbal Comprehension subtest (Comprehension);
5. Picture Concepts will correlate highly with both verbal and perceptual reasoning tests;
6. The WISC-IV (US) composites will correlate significantly with each other. Based on the inter-subtest correlations of the US, Mexican and Colombian studies, the weakest correlation is expected to be between the processing speed index and the other WISC-IV (US) indices.

Hypotheses 2, 3, 4, 5 and 6 will be tested by visual examinations of correlation coefficients. Statements about the magnitude of one correlation coefficient relative to another are therefore not supported by statistical analyses of significance in this study, but are used for hypothesis building.

Results

Pearson correlation coefficients for the core subtest pattern matrix are presented in Table 5.1. Table 5.1 also includes the correlation coefficients from the Wechsler (2003b) study in order to make comparisons with the T&T sample. Results show that, consistent with findings discussed earlier, significant and positive subtest inter-correlations (*p* <.01) were observed in the T&T sample. Additionally, inter-correlations appear to support the grouping of subtests into the current WISC-IV (US) 4-factor structure.

Table 5.1.
Correlational Matrices of WISC-IV subtests for the T&T and US Standardization Samples Combined

		T&T*			US*			T&T			US			T&T		US		T&T	US
		Sim	Voc	Com	Sim	Voc	Com	BD	PCon	MR	BD	PCon	MR	DS	LNS	DS	LNS	Cd	Cd
	Voc	**.83**			**.74**														
VCI	Com	**.74**	**.77**		**.62**	**.68**													
	BD	.52	.52	.46	.50	.48	.36												
PRI	PCon	.61	.55	.56	.45	.42	.40	**.56**			**.41**								
	MR	.57	.62	.54	.49	.49	.42	**.61**	**.55**		**.55**	**.47**							
	DS	.44	.45	.42	.39	.42	.36	.42	.45	.40	.35	.30	.38						
WMI	LNS	.56	.60	.58	.47	.50	.43	.52	.58	.62	.38	.36	.42	**.52**		**.49**			
	Cd	.27	.36	.40	.28	.31	.30	.35	.34	.37	.36	.29	.34	.26	.35	.23	.30		
PSI	SS	.41	.49	.47	.39	.38	.34	.37	.39	.36	.45	.34	.42	.35	.36	.40	.40	**.60**	**.53**

All T&T correlations are significant at the p<.01 level. US sample correlations are significant (Wechsler, 2003b).

Note. Sim = Similarities, Voc = Vocabulary, Com = Comprehension, BD = Block Design, PCon = Picture Concepts, MR= Matrix Reasoning, DS = Digit Span, LNS = Letter Number Sequencing, Cd = Coding, SS = Symbol Search, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index.

Consistent with the 3rd hypothesis, moderate correlations were observed between the *g* loaded verbal tests and Block Design; and as expected, the *g* loaded tests of the Verbal Comprehension Index (Similarities and Vocabulary) correlated more strongly with each other than with Comprehension. However, some differences from the US sample were also revealed. For example, in the T&T sample, both Matrix Reasoning and Picture Concepts correlated equally well with the verbal comprehension and the other perceptual reasoning subtests.

Letter Number Sequencing correlated less strongly with Digit Span (T&T sample: *r* = .52; US sample: *r* = .49) than with all the verbal comprehension and 2 perceptual reasoning subtests (T&T sample: *r* = .56, .60, .58, .58, .62; and US sample: *r* = .47, .50, .43, .36, .42 for the Similarities, Vocabulary, Comprehension, Picture Concepts and Matrix Reasoning subtests respectively). Index score inter-correlations generally replicated the findings of the WISC-IV (US) standardization study, except in some cases, where the T&T sample correlations were stronger than the US sample (See Table 5.2).

Table 5.2.

Correlational Matrices of WISC-IV Composites for the T&T and US Standardization Combined sample (N=203)

	T&T				US			
	PRI	WMI	PSI	FSIQ	PRI	WMI	PSI	FSIQ
VCI	.70	.64	.49	.89	.62	.56	.43	.85
PRI		.65	.45	.88		.52	.51	.86
WMI			.42	.79			.40	.76
PSI				.65				.70

All T&T correlations are significant at the p<.01 level. US sample correlations are significant (Wechsler, 2003b).

Note. VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index.

Discussion

Subtest Inter-Correlations

The findings of the study mostly replicated the results of the Wechsler (2003b) inter-correlation analyses. As Wechsler predicted, all subtests correlated with each other suggesting that they may measure a common higher order factor. The data also supported the hypothesis that the subtests potentially can be grouped into four factors. The T&T sample data provided support for the hypothesis that the Picture Concepts test would be moderately correlated with the verbal comprehension subtests.

The *g* – loaded tests (Block Design, Similarities and Vocabulary) also demonstrated moderate to strong inter-correlations (*r*s between .44 and .84). However, the analyses of the T&T data also revealed some interesting divergences (discussed in the following sections) from the findings of the US standardization study.

A key finding was that of a strong positive association of the Picture Concepts, Matrix Reasoning and Letter Number Sequencing subtests with the verbal comprehension subtests. The strength of the relationships between these non-verbal and verbal subtests either matched or exceeded within factor correlations. There are several possible explanations for these patterns of correlations.

One argument can be that the tests all tap into fluid reasoning ability. Indeed, working memory and fluid reasoning abilities have been linked (Salthouse, 1992; Kyllonen, 1994; Fry & Hale, 1996; Engle, Tuholski, Laughlin, & Conway, 1999; Süß, Oberauer, Wittmann, Wilhelm & Schulze, 2002). Similarities, has also been found to load on a fluid reasoning factor (Flanagan & Kaufman, 2004; Chen et al, 2009). A strong correlation between Picture Concepts and Similarities in the T&T study is particularly interesting. In fact, the relationship between this pair of variables is stronger than the relationship between Picture Concepts and the other tests of perceptual reasoning. Similar but not as striking findings were demonstrated in the Mexican and Columbian studies where correlations between Picture Concepts and Similarities exceeded that of Matrix Reasoning and Block Design respectively. Such an association may be explained by a unique sort of associative or categorical reasoning ability that is required for both tests. Similarities involves linking of words while Picture Concepts requires the linking pictures (Wechsler, 2003b). Both also require verbal strategizing – an ability to reason verbally how items can share similar features or purposes – to facilitate problem solving on tasks.

Take for example a Similarities type item that asks how happiness and sadness are alike (See the example in Chapter 2). A child can use the strategy of picturing a happy person or a sad person, from which their response can be ‘A facial expression’. For this response the child may be awarded a score of one. However, if the child uses a strategy of finding a verbal definition for each word and then finding the commonality between the words (e.g., “They are emotions”), then a 2-point response is more likely. Likewise, when faced with a Picture Concepts type item, for example, pictures of the following foods: an orange, a tomato, a slice of bread, and cooked chicken (See Figure, 2.3), a child may experience difficulty unless they have knowledge of specific verbal food categories. If a child can categorize the objects into food groups such as fruits, breads, and meats they are more likely to achieve the correct response – that the orange and the tomato go together.

Another factor, an underlying education cluster, may contribute to the strong correlations between Picture Concepts, Matrix Reasoning, Letter Number Sequencing and the verbal comprehension. In other words, an education or exposure related variable is proposed to mediate performance on these subtests. In the pilot studies (reported in Chapter 2) it was hypothesized that Picture Concepts loaded on a crystallized ability factor, potentially mediated by the child’s exposure to US information. The Picture Concepts subtest which contains stimuli that may be more accessible to an American child, might place a T&T child at a considerable disadvantage especially if that child is unfamiliar with US culture. This hypothesis was not supported by the pilot data which did not support a verbal/nonverbal crystallized factor; however, the argument that Picture Concepts is linked to an education or exposure factor is believed be of merit and worthy of further examination.

Formal education has been shown to predict performance on the Digit Span test of working memory (Otrosky – Solis & Lozano, 2006; Kosmidis, Zafiri, & Politimou, 2011). Word span and digit span performance have been associated with levels of literacy and formal schooling (Kosmidis, Zafiri, & Politimou, 2011). While these findings may not be generalizable to other measures of working memory, it is reasonable to assume that the formal academic skills associated with Digit Span performance (number knowledge) (Kosmidis, Zafiri, & Politimou, 2011) may also influence performance on the Letter Number Sequencing test (alphabet and numerical knowledge). Similar arguments were put forward to explain the non-invariance of the Similarities, Digit Span, and Letter Number Sequencing tests across self-referred and SSSD-referred samples of the pilot studies in Chapter 2. The influence of an education related factor or covariate on Picture Concepts, Matrix Reasoning, Letter Number Sequencing and the verbal comprehension subtests is therefore worthy of further investigation through multiple regression and factor analyses.

Composite Inter-Correlations

In support of Hypothesis 6, analyses which focused on composite inter-correlations revealed that the verbal comprehension, perceptual reasoning and working memory indices were more strongly correlated with each other than with the processing speed index. This finding replicates the results of studies that demonstrate that processing speed subtests load the least strongly on *g* (Keith et al., 2006; Chen, Keith, Chen & Chang, 2009). The processing speed index therefore may be less suitable than the other composites to calculate a full scale IQ (FSIQ) score. The General Ability Index (GAI) of the WISC-IV (US) is a recommended, and for many psychologists an established replacement for the FSIQ if the Working Memory and Processing Speed indices are considered to contribute to a less reliable estimate of the child's true potential (Raiford, Weiss, Rolfhus & Coalson, 2005; Saklofske, Zhu, Coalson, Raiford, & Weiss, 2010). A similar approach could be useful for a sample such as this, if it is determined through factor analysis that the subtests of the processing speed composite load very weakly on a general factor. The following sections report the relevant analyses that were conducted to investigate how well the processing speed subtests as well as the index as a whole load on a general factor. The analyses will also help determine how the data is organized in the sample. Specifically, the investigations will seek to determine how much of the variance in the performance on the Picture Concepts, Matrix Reasoning, Letter Number Sequencing, and the verbal comprehension tests can be attributed to fluid reasoning, education/exposure or both.

Study 3:

Exploratory Factor Analysis of the WISC-IV (US) Core Subtests in a T&T Sample

Factor Analysis is a statistical method by which a number of observed variables can be reduced to a smaller number of fairly independent factors (Floyd & Widaman, 1995; Tabachnick & Fidell, 2007). These factors represent underlying latent constructs that summarize the correlations among a group of variables (Floyd & Widaman, 1995; Kline, 2000; Tabachnick & Fidell, 2007). In the case of the WISC-IV (US), factor analysis reduces the data from the core and supplemental subtests into a smaller set of latent ability factors– Verbal Comprehension (VCI), Perceptual Reasoning (PRI), Working Memory (WMI), Processing Speed (PSI) and General Intelligence (FSIQ). These ability factors are understood theoretically to cause the observed behaviours. Thus, a person who obtains a very high score on the FSIQ composite is seen as someone of very high general intelligence.

Factor analysis is concerned with what is known as common variance which is the variance that is shared among the variables in the matrix (Kline, 2000). Common variance is distinguishable from variance associated with the unique properties of each subtest (unique variance) and that which is related to error (Kline, 2000). By analysing the common variance, the procedure extracts factors that best summarize linear combinations of the variables within the matrix (Kline, 2000). The relationship between the variables and their corresponding factors can be represented as coordinates plotted relative to X and Y axes. The X and Y axes represent the factors, and the coordinates are the factor loadings. The results of the first solution may prove problematic in interpretation because the relationship of the observed variables with one factor or another may overlap considerably (Tabachnick & Fidell, 2007). In other words, a variable may demonstrate salient loadings (> .3) on more than one factor. To solve this problem, the analyses must produce solutions in which the relationships of each variable to a specific factor is more clearly defined. This is accomplished by rotating the factor axes (Tabachmick & Fidell, 2007).

Two types of rotation are available to the analyst. One accounts for correlations among the factors (oblique rotation), such as in IQ tests where intellectual abilities are expected to be related to each other in some way. The other rotation assumes that the factors are independent or uncorrelated (orthogonal rotation). The number of factors that can be extracted is equivalent to the number of variables in the matrix, however the factors that account for the biggest chunk of variance among the variables are the most important. This degree of importance is generally represented by eigenvalues which represent how well variables correlated within the matrix. When the variables are better related, the eigenvalues are greater than or equal to one; however, less strict criteria propose an eigenvalue cut off of .70 (Field, 2009).

The resulting linear combinations summarize the relationship among the items that make up the factor. A linear combination is defined mathematically as $Y = b_1X_1 + b_2X_2 + b_3X_3 \ldots b_nX_n + \varepsilon_1,$ where Y is the factor, b is the factor loading, X is the variable and ε is error (Field, 2009). As an example, the Similarities, Vocabulary and Comprehension subtests make up the VCI factor of the WISC-IV(US). Each of these three VCI subtests correlate differently with the factor. The strength of the relationship between the factor and the subtest is the factor loading (Field, 2009). To calculate the factor score, different weights based on factor loadings can be applied in order to compute the scaled scores which are then summed (Di Stefano, Zhu &Mindrila, 2009; Field 2009).

This method provides the basics of factor analysis using an exploratory approach; however, in assessing validity, two different factor analytic approaches can be used: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA which is used here, employs a blind approach, in which the pattern matrix is explored freely without prior specifications and guidelines about underlying factors (Tabachnick & Fidell, 2007). In psychometric measurement, test items are selected according to how well they are expected to measure the construct of interest. EFA assesses validity by revealing the nature of the test's underlying structure thereby confirming whether the developers have adequately operationalized the construct (Tabachnick & Fidell, 2007). The following analyses assess the internal validity of the WISC-IV (US) in a sample of

T&T children, using two methods of EFA designed to identify the underlying factor structure of the WISC-IV (US) in this sample.

Two key studies by Wechsler (2003b) and Watkins (2006) used different EFA methods to assess the validity of the WISC-IV (US) and obtained differing results. In the standardization study, exploratory factor analyses used principal axis factoring with oblique rotation to confirm the 4 factor structure of the WISC-IV (US) (Wechsler, 2003a). The Watkins (2006) study was concerned with uncovering the hierarchical structure of the WISC-IV (US). These aims were accomplished with the Schmid & Leiman (1957) orthogonalization method. This higher order analysis examines the correlational matrix of the subtests to extract a higher order factor (Schmid & Leiman, 1957, Watkins, 2006; Watkins et al., 2006). After being residualized of the variance explained by the higher order factor, the variance explained by the 1st order orthogonal factors is then extracted (McClain, 1996).

For the purposes of these analyses, EFAs utilizing two types of rotation - the oblique method employed by Wechsler (2003b), and the orthogonal method employed by Watkins (2006) were used to assess the validity of the 4 first order factor structure for the WISC-IV as proposed by Wechsler (2003b) and then the hierarchical structure of the WISC-IV as evidenced by Watkins (2006). For this study, the observed scaled scores of the full sample of 203 T&T children were used. The sample was not stratified for age in order to increase the power of the analysis.

Analyses

Principal Axis Exploratory Factor Analysis (EFA) was performed with R 3.2.1(Psych package) on the scaled scores of the 10 core subtests of the WISC-IV (US). Data screening analyses resulted in a Kaiser-Meyer-Olkin Measure of Sampling

Adequacy coefficient (KMO) of .90, and a χ^2 (45) = 1149.29 (*p* <.001) on Bartlett's Test of Sphericity. KMO values were more than .81 for individual subtests. All values reflected adequate sample size and a non-random correlational matrix (Tabachnick & Fidell, 2007; Field, 2009). Guidelines for retaining factors are varied, however; using both Eigenvalues >.70 (Joliffe, 1986 cited in Field, 2009) and the visual scree test (Cattell, 1966 cited in Field, 2009), it was first determined that 3 criteria can be extracted. However, with the eigenvalue for the fourth factor at .666, and in the interest of replicating the Wechsler (2003b) study, a decision was made to include a four factor extraction in the series of EFAs. Therefore, one, two, three and four factors respectively were extracted using oblique rotation with Promax (k=4).

Results

The results of the first order EFAs are presented in Tables 5.3, as a pattern matrix of the factor loadings of each subtest with their communality values. The 4 factor model demonstrated some indicators of good model fit and accounted for 65% of total variance (χ2 (11) = 14.89, *p*=0.19; BIC= -43.55; TLI= .98; RMSEA=.05 (90% CI: .00; .09); RMSR=.01).

Good model fit was also observed for the 3 factor solution which explained 62% of total variance (χ2 (18) =26.16, *p*=0.10; BIC= -69.47; TLI= .98; RMSEA=.05 (90% CI: .00; .08); RMSR=.02). Both the two factor and one factor solutions demonstrated poor fit to the data. Loadings of the subtests on factors, and communality values are shown in Tables 5.4.

Table 5.3.

Goodness of Fit Indexes for Evaluating Model Adequacy for first order EFAs (N= 203)

Model	χ^2	*df*	*p*	TLI	BIC	RMSR	RMSEA (90%CI)
1 factor	174.06	35	.000	0.88	-11.9	0.07	0.07 (.12 - .16)
2 factor	102.95	26	.000	0.88	-35.19	0.05	0.12 (.09 - .15)
3 factor	26.16	18	.10	0.98	-69.47	0.02	0.05 (.00 - .08)
4factor	**14.89**	**11**	**.19**	**0.99**	**-43.55**	**0.01**	**0.05** **(.00 - .09)**

Note. *df* = degree of freedom; TLI = Tucker Lewis index; BIC = Bayesian Information Criterion; RMSR= root mean square residual; RMSEA ¼ root mean square error of approximation.

Table 5.4.

Factor Loading and Communalities – 3 factor solution

	Verbal	Non Verbal	Speed	h^2
Sim	**.83**	.18	-.11	.83
Voc	**.81**	.10	.05	.85
Com	**.67**	.11	.13	.70
BD	-.09	**.77**	.04	.54
PCon	.11	**.65**	.00	.54
MR	.03	**.75**	.00	.60
DS	.05	**.52**	.05	.34
LNS	.08	**.74**	-.04	.60
Cod	-.14	.08	**.80**	.60
SS	-.08	-.08	**.77**	.64

(χ2 (18) = 26.16, *p* = 0.10; BIC = -69.47; TLI = .98; RMSEA = .05 (90% CI: .00; .08); RMSR = .02)

Note: h^2 = Communality or portion of variance explained by the factors

Sim = Similarities, Voc = Vocabulary, Com = Comprehension, BD = Block Design, PCon = Picture Concepts, MR = Matrix Reasoning, DS = Digit Span, LNS = Letter Number Sequencing, Cod = Coding, SS = Symbol Search.

Factor inter-correlations are summarized in Table 5.5. The three factor solution indicated that the 10 core subtests were organized with salient loadings on 3 factors, tentatively identified as Verbal, Nonverbal, and Speed. The Nonverbal factor consisted of subtests associated with the WISC-IV (US) PRI and WMI factors combined. No salient cross-loadings were identified in this solution. Factor correlation coefficients ranged from .54 for Verbal and Speed to .75 for Verbal and Nonverbal.

Table 5.5.

Factor Correlations: Three Factor Solution

	Verbal	Nonverbal
Nonverbal	.75	-
Speed	.54	.60

The four factor solution demonstrated a replication of the factor structure of WISC-IV(US) (Wechsler, 2003b) (See Table 5.6). For Picture Concepts, cross loadings were noted, with the highest pattern coefficient on the factor consistent with the Working Memory Index. Table 5.7 summarizes the correlations among the four factors, which ranged from moderate to high (.53 to .78), thus supporting an underlying higher order factor.

Table 5.6.

Factor Loading and Communalities – 4 factor solution

	VCI	PRI	WMI	PSI	h²
Sim	**.90**	.04	.06	-.11	.83
Voc	**.89**	.08	-.07	.04	.85
Com	**.72**	.00	.06	.13	.70
BD	-.03	**.58**	.16	.06	.53
PCon	.14	**.28**	**.35**	.01	.53
MR	.06	**.91**	-.12	.00	.73
DSpan	-.07	-.14	**.85**	.02	.49
LNS	.08	.29	**.49**	-.04	.61
Cod	-.13	.14	-.07	**.81**	.62
SS	.12	-.14	.09	**.75**	.64

$\chi2$ (11) = 14.89, *p* = 0.19; BIC = -43.55; TLI = .97; RMSEA =.04 (90% CI: .00; .09); RMSR =.01. Note: h² = Communality or portion of variance explained by the factors VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index, Sim = Similarities, Voc = Vocabulary, Com = Comprehension, BD = Block Design, PCon = Picture Concepts, MR = Matrix Reasoning, DS = Digit Span, LNS = Letter Number Sequencing, Cod = Coding, SS = Symbol Search.

Table 5.7.

Factor Correlations: Four Factor Solutions

	VCI	PRI	WMI
PRI	.73	-	
WMI	.76	.80	-
PSI	.55	.55	.57

Note: VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index.

In order to explore the higher order structure of the raw score data, the *Schmid-Leiman Orthogonalization (SL)* procedure (Schmid & Leiman, 1957) was performed using R (Psych package). This procedure was employed to assess the higher order structure of the WISC-IV(US) by extracting a higher order factor directly from the observed variables. The second phase of analyses is concerned with the common variance that has been residualized of the variance accounted for by the higher order factor (Watkins, 2006; Watkins et al., 2006). The first order coefficients represent the unique variance that is accounted for by the first order factors which are uncorrelated with each other and the higher order factor (Watkins, 2006; Watkins et al., 2006).

Using oblique rotation techniques, the 4 factor solution was observed to demonstrate the best fit of 4 models that were analysed; therefore, SL orthogonalization procedures were conducted to extract a higher order factor with four 1^{st} order factors. The resulting solution is summarized in Table 5.8. An additional analysis was done to determine if factor loadings for the T&T and US samples were similar. The Tucker Coefficient of Congruence (*rc;* Tucker, 1951), was calculated for the matrices. The resulting *rc* was .94 indicating that the loadings for the two samples were similar to a high degree and almost equal (Jensen, 1998; Lorenzo-Seva & ten Berge, 2006).

The results of the SL analyses revealed that all 10 subtests contributed a considerable amount of variance to the higher order factor. Pattern coefficients ranged from .39 to .88 and are considered to be salient loadings on the higher order factor. After being residualized of what was accounted for by the higher order factor, variances ranged from .39 for Similarities and Vocabulary to .81 for Coding. The subtests with the highest loadings on the general factor were Similarities, Vocabulary, Comprehension, Picture Concepts, Matrix Reasoning and Letter Number Sequencing. Coding displayed the lowest loading on a general factor although its loading on the processing speed factor was considerable.

Table 5.8.

Factor Loading and Communalities – 4 factor for the T&T Sample.

	G		F1		F2		F3		F4		h²	u²
	b	Var	*b*	Var	*b*	Var	*b*	Var	*b*	Var		
Sim	**.78** (.71)	.61	**.45** (.43)	.20							.82	.18
Voc	**.78** (.72)	.61	**.51** (.49)	.26							.88	.12
Com	**.73** (.61)	.53	**.38** (.46)	.14							.69	.31
BD	**.66** (.65)	.44							- (.29)		.50	.50
PCon	**.71** (.56)	.50			**.30**	.09			(.20)		.60	.40
MR	**.74** (.68)	.55							**.66** (.30)	.44	.99	.01
DS	**.56** (.55)	.31			**.25** (.34)	.06					.38	.62
LNS	**.72** (.63)	.52			**.22** (.33)	.05					.58	.42
Cod	**.44** (.47)	.19					**.71** (.47)	.50			.71	.29
SS	**.51** (.58)	.26					**.54** (.45)	.29			.57	.43
Total Var	45.2%		6.0%		2.0%		7.9%		4.4%		67.2 %	38.2 %
Common Var	67.1%		8.9%		2.9%		11.8%		6.5%		100	

$\chi2$ (11) = 11.75, p = 0.38; BIC = -46.69; RMSEA =.02 (90% CI: .00; .08); RMSR = .02

Note: *b* = Standardized loading, F1 = Factor 1, F2 = Factor 2, F3 = Factor 3, F4 = Factor 4, Sim = Similarities, Voc = Vocabulary, Com = Comprehension, BD = Block Design, PCon = Picture Concepts, MR = Matrix Reasoning, DS = Digit Span, LNS = Letter Number Sequencing, Cod = Coding, SS = Symbol Search.

US Sample loadings (Watkins, 2006) are presented in brackets.

After higher order variance was accounted for, negligible communality values for Block Design were observed. Additionally, Matrix Reasoning displayed a communality value of 1 after the orthogonalization. The general factor (*g*), explained overall more variance in the 10 core subtests than the first order factors, with one exception. The factor that is associated with Processing Speed items explained 50% of the variance in Coding compared with the general factor which explained 20%. Additionally, the first-order factors accounted for between 2.9% (F2) and 11.8% (F3) of common variance and 2.0% (F2) to 7.9% (F3) of total variance. The higher order factor accounted for 67.1% of common variance and 45.2% of the total variance.

Discussion

The results of the SL orthogonalization analyses supported a general factor and 4 first order factor solution. The salient loadings of all subtests on the higher order factor provides compelling evidence that the tests all contribute at varying levels to a higher order, general intelligence factor. Furthermore, general intelligence seems to explain almost over 60% of the common variance. These findings are compatible with those of the Watkins (2006) and Watkins and colleagues (2006) studies, in which 71.3% and 75.7% respectively of the common variance was found to be attributable to general intelligence. Similar findings were obtained from higher order EFAs performed on the10 core subtests of the WISC Spanish (2004b) that were administered to 500 Spanish speaking children. Results showed that *g* accounted 73.2 % of the common variance (McGill & Canivez, 2016). Also, similar to the results of Watkins et al. (2006), Watkins (2006) and McGill and Canivez (2016), the highest loadings on *g* were shared by Similarities and Vocabulary, followed by Matrix Reasoning, Comprehension, Letter Number Sequencing, Block Design, and in all but 1 study (Watkins, 2006), Picture

Concepts. The results of the EFA of the present study lend considerable support the notion that the strong correlations among the verbal comprehension subtests and Picture Concepts, Matrix Reasoning and Letter Number Sequencing may be explained by their strong loadings on a general intelligence factor. Compared to the results of the Watkins study, the impact of the general intelligence factor is stronger in the T&T sample than the US standardization sample.

The Coding subtest was least associated with a general intelligence factor (*b* = .44), a result that is compatible with the Watkins, (2006), Watkins et al. (2006) (*b* = .45) and McGill and Canivez (2016) (*b* = .39) findings. This suggests that Coding may not be a very effective measure of general intellectual ability. An additional finding was that the Block Design subtest retained little common variance after being residualized of the variance accounted for by the general factor. This finding suggests that in this model, Block Design is more reliable as measure of g than perceptual reasoning ability. Additionally, because it did not load on the perceptual reasoning factor, Block Design may have measured a fifth factor such as visual processing or spatial ability. This hypothesis is supported by the LeCerf et al (2010) and Reverte et al. (2014) studies which also found Block Design loaded less on a perceptual reasoning factor than on a visual processing factor.

Picture Concepts, Working Memory and the Verbal Comprehension subtests –An Education/Exposure Related Cluster?

The first order EFAs on the T&T sample data performed extractions of one, two, three and then four factors using an oblique rotation method in order to replicate the Wechsler (2003b) standardization study. Results provided support for the four-factor model theorized by Wechsler(2003b). The Picture Concepts subtest loaded simultaneously on the factors associated with Perceptual Reasoning and Working Memory, with the loading for working memory exceeding perceptual reasoning. The results suggest that for the T&T sample, performance on the Picture Concepts test was more closely related to the construct that is measured by the working memory subtests. Additionally, the higher order EFA showed that the subtests clustered into 4 1st order orthogonal factors: (1) a verbal crystallized factor consisting of Similarities, Vocabulary and Comprehension; (2) an un-named factor consisting of Picture Concepts, Letter Number Sequencing and Digit Span; (3) a Processing Speed factor with Coding and Symbol Search; and a Perceptual Reasoning Factor with Matrix Reasoning.

What is this un-interpreted Working Memory/Picture Concepts factor? Wechsler (2003b) described the Digit Span and Letter Number Sequencing subtests as measures of auditory short term memory, attention and concentration. Studies have also pointed to the possible influence of education on working memory subtests (Otrosky – Solis & Lozano, 2006; Kosmidis, Zafiri, & Politimou, 2011). Additionally, the results of the inter-correlation study showed that Picture Concepts was more strongly correlated with Letter Number Sequencing than either of the two other subtests of the Perceptual Reasoning Index (Block Design and Matrix Reasoning). Both Picture Concepts and Letter Number Sequencing demonstrated correlations with the Verbal Comprehension subtests that mostly exceeded their within factor correlations.

Prior to the pilot studies, it was hypothesized that Picture Concepts tapped into crystallized ability which is developed through formal and informal education. This hypothesis was assessed by testing the fit of a model which included Picture Concepts as part of the verbal/nonverbal crystallized ability factor with Similarities, Vocabulary and Comprehension. This model did not fit the data. Interestingly after stratifying the sample based on a proxy variable for SES, and investigating the structural invariance of the model between the groups, the findings pointed to both the working memory and Similarities subtests as responsible for the factorial non-invariance between samples. It was suggested that both the Similarities and Working Memory tests reflected differences in formal education and general exposure between the groups.

On the basis of the findings of strong inter-correlations among the verbal tests, the Letter Number Sequencing test and the Picture Concepts test, a nonverbal education related cluster was hypothesized. This cluster is argued to reflect a set of crystallized abilities that have been developed through either formal or tacit learning, but are not expressed verbally. This is not reasoning ability as it depends on the fund of information gathered over time. In other words, while Picture Concepts was designed to measure a verbally mediated ability to form associations between pictures (Wechsler, 2003b), it also may tap into the child’s picture vocabulary as well as their knowledge of a wide range of objects and their respective functions. Such knowledge will not be conveyed verbally, but will act to mediate between the child’s latent reasoning ability and their observed responses on the test.

As Picture Concepts is proposed to be a measure of acquired knowledge and exposure to US information, similarly, Letter Number Sequencing is considered to be a measure of the child’s learnt ability to organize numbers and letters in specific order. As such, it is proposed that the Picture Concepts and Letter Number Sequencing tests may measure nonverbal crystallized ability in addition to categorical reasoning and working memory respectively. This hypothesis will be tested with Confirmatory Factor Analysis (CFA).

Study 4:

Confirmatory Factor Analysis of the WISC-IV (US) Core Subtests in a T&T Sample

As opposed to the purely exploratory approach of EFA (Floyd & Widaman, 1995; Barrett, 2007), Confirmatory Factor Analysis (CFA) is predicated on a theory (Tabachnick & Fidell, 2007; Jackson, Gillaspy Jr. & Pure-Stephenson, 2009). CFA assesses a clearly defined hypothesized model to determine how well it summarizes the data (Floyd & Widaman, 1995; Tabachnick & Fidell, 2007; Jackson, Gillaspy Jr. & Pure-Stephenson, 2009). Though both methods are interested in latent structure, using CFA to supplement EFA analyses is recommended because the statistical methods used for each may result in different conclusions about the data (Barrett, 2007).

CFA is a subset of a larger body of analytic techniques called Structural Equation Modelling (SEM) which statistically analyses models used to describe the interaction among one or more independent variables and dependent variables (Tabachnick & Fidell, 2007). SEM also simultaneously analyses a collection of smaller models within a larger model (Barrett, 2007). SEM will be discussed in more detail in Chapter 7.

CFA assesses the fit of the 'measurement model' which is a subset of the larger model (Tabachnick & Fidell, 2007). The measurement model describes the relationship among observed variables and latent constructs (Tabachnick & Fidell, 2007). The information obtained by conducting CFAs can be useful for adjusting and refining the measurement model (Floyd & Widaman, 1995). In the case of the WISC-IV (US), CFA is useful for determining whether the Wechsler (2003b) four 1st order factor model or the Watkins (2006) direct hierarchical model provides the best fit to the data. CFA does not only investigate the fit of the model to the data, but also uses the maximum likelihood estimation to compare the fit and parsimony of the main hypothesis to alternative models to determine which one explains the data best (Floyd & Widaman, 1995). It also tells us if a significant amount of variance remains after the factors are accounted for (Barrett, 2007).

Model fit is the extent to which the population covariance implied from the parameters of the model and the covariances of the actual sample are the same (Barrett, 2007). The residuals (discrepancy between samples) is summarized by the χ^2 goodness of fit statistic; or otherwise, a null hypothesis test, where the smaller values indicate smaller discrepancies between models. A well-fitting model is one in which sample covariances and population covariances are equal. The researcher therefore wants to know if the residuals are greater than what will occur by chance, therefore, a significant χ^2 reflects poor model fit (Bollen, 1989). Because power increases with sample size, it will not be surprising to find large and significant χ^2 for very large samples. In such cases, it is recommended that goodness of fit indices should also be used to evaluate model fit (Barrett, 2007).

To determine model fit, chi square goodness of fit is the recommended statistic. Bollen (1989) recommended that chi square values below 2 and *p* values over .05 are indicators of good model fit; however as was stated prior, increasing the sample size increases the sensitivity of the analyses to differences between implied population and sample covariances. It therefore is not unusual to see chi-square values exceed these thresholds. As such, model evaluation can be accomplished though the examination of fit indices such as the Aikake Information Criterion (AIC), goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), comparative fit index (CFI), the Tucker Lewis index (TLI), the standardized root mean square residual (SRMR), and the root mean square of approximation (RMSEA) (Floyd & Widaman, 1995; Vrieze, 2012). Thresholds for good model fit have been suggested by Hu and Bentler (1998, 1999) and by Hopper, Couglan and Mullen (2008) to be equal to or more than .95 for CFI, TLI, GLI, and AGFI, lower than .06 for RMSEA, and lower than .08 for SRMR. Additionally, the model with the smallest AIC value is deemed to be the best fitting model (Hooper, Couglan & Mullen, 2008; Watkins, 2010). Barrett (2007) argued for caution in using model fit indices over the χ^2 statistic in CFA, suggesting that indices thresholds have been used in an ad hoc manner with no theoretical rationale for the indices that are used or the thresholds that are adopted as criteria for good model fit.

To further assess the internal validity of the WISC-IV(US), the Wechsler study used CFA to assess the fit of an a-priori model which consisted of the four 1st order factors. The analyses were done on the 10 core subtests in one study and the 10 core and 5 supplemental tests in another study (Wechsler, 2003b). To assess the validity of the hypothesized model, Wechsler (2003b) compared the 4 factor model to 3 competing models:

1. A one factor model in which the 10 subtests loaded on a general factor;
2. A two factor model (Similarities, Vocabulary, Comprehension, Digit Span and Letter Number Sequencing on one factor, and Block Design, Picture Concepts, Matrix Reasoning, Coding and Symbol Search on the other factor); and
3. A three factor model (Similarities, Vocabulary, Comprehension on one factor, Block Design, Picture Concepts, Matrix Reasoning on another factor; and Digit Span, Letter Number Sequencing, Coding and Symbol Search on the third factor.

Table 5.9 summarizes the goodness of fit statistics for the CFAs on the total standardization sample. The results demonstrated that the 4 factor model demonstrated good model fit and also best fit compared to the alternative models (Wechsler, 2003b). The Wechsler study (2003b) did not use CFAs to investigate the higher order structure of the WISC-IV or at least did not report on the findings. This approach was surprising especially since general factor is included as part of the scoring structure of the WISC-IV (Chen et al., 2009). With regard to the scoring structure, studies have found support for the Wechsler proposed model, in which the higher order factor (measured as FSIQ) is indirectly measured by the 10 core subtests (Keith, 2005; Bodin, Pardini, Burns & Stevens, 2009; Watkins, 2010).

Table 5.9.

Goodness of Fit Statistics for Confirmatory Factor Analysis of 10 Core Subtests – Wechsler (2003b) Standardization Study

	Goodness of Fit Indices				
	χ^2	*df*	χ^2/df	AGFI	RMSEA
Null Model	8965.24	45	199.23		
Model 1	1376.95	35	39.34	.83	.132
Model 2	687.80	34	20.23	.90	.094
Model 3	497.91	32	15.56	.93	.081
Model 4	131.62	29	4.54	.98	.040

Note: χ^2 = chi square, *df* = degrees of freedom, χ^2/df = normed chi square, AGFI = adjusted goodness of fit index, RMSEA = root mean square of approximation.

Confirmatory factor analyses also have demonstrated good fit for the direct model proposed by Watkins (2006); and in fact, superior fit compared to the other models (Watkins, 2010). These findings have found additional empirical support (Golay, Reverte, Rossier, Favez & Lecerf, 2012; Devena, Gay & Watkins, 2013; Canivez, 2014). Additionally, Watkins (2010) found that the global factor explained most of the common variance of the WISC-IV (US) subtests (75% of common variance and 48% of total variance). Additionally, the four factors each accounted for less than 10% of either common or total variance. The findings of the study were considered to provide important implications for the interpretation of test scores in real life practice.

It was felt that the Wechsler established scoring structure did not adequately describe the performance of the standardization sample based on the data. Also, Watkins (2010) argued that the Wechsler indirect hierarchical scoring structure is difficult to interpret without knowing the exact loadings of the variables on the general factor. The Watkins (2010) study highlighted the importance of CFA analyses to help provide an alternative interpretive framework from which performance on the WISC-IV (US) may be understood. Such alternative approaches will be discussed in Chapters 7 and 8.

Analyses

The present study used CFA to investigate the validity of the four first order factor structure of the WISC-IV (US) that was proposed by Wechsler (2003b) in the sample of 11 and 12-year-old T&T children. Confirmatory factor analyses (CFAs) were carried out using maximum likelihood estimation in MPlus 6 to compare the fit indices of the model to those of 7 alternative models. Two of the alternative models were identified by Keith (2005) and Watkins (2006), while the other three were suggested by the author of this study to explain the findings already contained in this thesis. The tested models are as follows:

Model 1. A one factor model (10 core subtests measure a general factor).

Model 2. A two oblique factor Model: Verbal (Vocabulary, Similarities, and Comprehension); and Nonverbal (Block Design, Picture Concepts, Matrix Reasoning, Digit Span, Letter Number Sequencing, Coding and Symbol Search).

Model 3. A three oblique factor Model: Verbal (Vocabulary, Similarities, and Comprehension); Nonverbal (Block Design, Picture Concepts, Matrix Reasoning, Digit Span, Letter Number Sequencing) and Processing Speed (Coding and Symbol Search).

Model 4. The four oblique factor current model: VCI, PRI, WMI, PSI theorized by Wechsler (2003b) (Figure 2.9, Chapter 2).

Model 5. A two level indirect hierarchical model with a higher order general factor and 4 correlated 2nd order factors: VCI, PRI, WMI, PSI (endorsed by Keith, 2005).

Model 6. A two level direct hierarchical model consisting of an overarching general factor measured by the 10 core subtests, and then four 1st order specific ability factors that are orthogonal to each other as found in the Watkins (2006/2010) studies.

Model 7. This model was based on the inter-correlation and EFA findings described earlier in this Chapter. This model which was found to demonstrate good fit, consists of a higher order general factor measured from the 10 core subtests; a Verbal Crystallized ability factor (Similarities, Vocabulary, Comprehension), a Non Verbal Crystallized factor (Picture Concepts, Digit Span, Letter Number Sequencing); Processing Speed (Coding, Symbol Search), and a Perceptual Reasoning factor (Matrix Reasoning and Block Design). After the extraction of the higher order factor, the results of the orthogonal higher order EFA saw negligible communality values for Block Design. However, in light of its high correlation with Matrix Reasoning (.61), it was retained in this theoretical model.

Model 8. This model replicates the first order structure of Model 7; however, the general factor is indirectly measured by the core subtests via the 4 specific ability factors. In this model, and like Model 5, the general factor is a summary of the 4 first order factors, Verbal Crystallized ability, Non Verbal Crystallized ability, Perceptual Reasoning and Processing Speed.

Results

Absolute fit of each model was evaluated using the normed chi square (χ^2/df) and model fit indices AIC, CFI, TLI, RMSEA and SRMR. Table 5.10 presents the model fit statistics for models 1 to 8. The table shows that for the oblique first order factor models, each model fits better than the previous model. Additionally, the one and two factor models did not demonstrate adequate fit to the data. The oblique three and four factor models demonstrated good fit, but neither model appeared to fit better than the other.

Table 5.10.

Goodness of Fit Indexes for Evaluating Model Adequacy (N= 203)

Model	χ^2	*df*	*p*	χ^2/df	AIC	CFI	TLI	SRMR	RMSEA
1	170.68	35	.000	4.88	9424.41	0.88	0.85	0.06	(CI90%)
2	98.06	34	.000	2.88	9353.80	0.94	0.93	0.05	.14 (.12 to .16)
3	45.10	32	.06	1.41	9304.83	0.99	0.98	0.03	.09 (.07 to .12)
4	41.13	29	.07	1.42	9306.86	0.99	0.98	0.03	.05 (.00 to .07)
5	43.25	31	.07	1.40	9304.98	0.99	0.98	0.03	.05 (.00 to .08)
6	37.63	30	.16	1.25	9301.37	0.99	0.99	0.02	.04 (.00 to .07)
7	41.77	31	.09	1.35	9303.50	0.99	0.99	0.03	.04 (.00 to .07)
8	41.72	31	.10	1.35	9303.45	0.99	0.99	0.03	.04 (.00 to .07)
									.04 (.00 to .07)

Note. χ^2 = chi square, *df* = degrees of freedom, χ^2/df = normed chi square, AIC = Akaike's information criterion, CFI = comparative of fit index, TLI = Tucker Lewis index, SRMR = squared root mean residuals, RMSEA= ¼ root mean square error of approximation.

Comparing the four 1st order factor indirect and direct hierarchical models (Models 5 and 6 respectively), both models demonstrated good fit, however the direct model provided a better fit to the data. These findings replicate the results of the Watkins (2010) study in which the direct model demonstrated better fit than both the indirect hierarchical and first order oblique factor models. In models 7 and 8, the observed variables were organized into 4 factors (Verbal Crystallized, Nonverbal Crystallized, Perceptual Reasoning, and Processing Speed), with a higher order general factor. In Model 7 the general factor was directly measured by the 10 core subtests, and in Model 8, the general factor was measured by the 4 first order factors. Between the two, the indirect model demonstrated a better fit to the data. Model 8 demonstrated a better fit than Model 5 but not Model 6. Model 6, the Watkins (2006) defined, 4 factor direct higher order factor structure provided the best fit for the T&T sample compared to the other models. Table 5.11 summarizes the WISC–IV (US) subtest standardized loadings and variance estimates based on the direct hierarchical model. Residual variances, communalities and variance explained values are also summarized in the table.

Table 5.11.

Standardized Loadings, Variances, Total and Common Variance Explained, Communalities, and Uniqueness values for the CFA of the direct hierarchical model.

	G		VC		PR		WM		PS		h²	u²	*p*
	b	Var	*b*	Var	B	Var	*b*	Var	*b*	Var			
Sim	**.78**	.61	**.41**	.17							.77	.23	.000
Voc	**.79**	.62	**.55**	.30							.91	.08	.002
Com	**.75**	.56	**.33**	.11							.68	.33	.000
BD	**.67**	.45			**.54**	.29					.74	.26	.006
PCon	**.74**	.55			**.11**	.01					.56	.44	.000
MR	**.74**	.55			**.22**	.05					.60	.40	.000
DS	**.58**	.33					**.31**	.10			.43	.57	.000
LNS	**.77**	.60					**.22**	.05			.65	.35	.000
Cod	**.46**	.21							**.87**	.76	.96	.03	.707
SS	**.54**	.29							**.41**	.17	.46	.54	.000
Total	47.7%		5.8%		3.5%		1.5%		9.3%		67.6%		
Comm.	70.6%		8.6%		5.2%		2.2%		13.8%		100.0%		

Note: *g* = general intelligence, VC = Verbal Comprehension, PR = Perceptual Reasoning, WM = Working Memory, PS = Processing Speed, h² = communality, u² =uniqueness, *p* = probability value, Sim = Similarities, Voc = Vocabulary, Com = Comprehension, BD = Block Design, PCon = Picture Concepts, MR = Matrix Reasoning, DS = Digit Span, LNS = Letter Number Sequencing, Cod = Coding, SS = Symbol Search, Total = Total variance explained, Comm. = Common variance explained.

The results of the CFA indicated that the general factor accounted for 70.6% of common variance and 47.7% of total variance. The Verbal Comprehension factor accounted for 8.6% of common variance and 5.8% of total variance. The Perceptual Reasoning factor accounted for 5.2% of common variance and 3.5% of total variance and the Working Memory and Processing Speed factors accounted for 2.2% and 13.8% respectively of common variance and 1.5% and 9.3% respectively of total variance. The results demonstrate that the higher order factor explained more of the common and total variance than the factor scores. Interestingly an examination of the residual variances for the 10 core subtests revealed that with the exception of the Coding and Vocabulary subtests, a significant portion of subtest variance remained unexplained. Over 90% of the variance for the Vocabulary and Coding subtests were explained by the factors in the model. The Tucker Coefficient of Congruence (*rc;* Tucker, 1951), was calculated to assess the equality of the matrices for the direct hierarchical models that were obtained in this study and the Watkins (2010) study. The resulting *rc* (.98) suggested that the loadings for the T&T and US standardization samples were statistically equal (Jensen, 1998, Lorenzo-Seva & ten Berge, 2006).

Discussion

Confirmatory factor analyses conducted on a sample of T&T 11 and 12-year-old children provided support for several alternative measurement models. The Wechsler (2003b), four 1st order oblique factor model demonstrated good fit to the data. Additionally, the Keith (2005) indirect hierarchical model, featuring four 1st order oblique factors and a general intelligence factor which is measured by the specific ability factors also found support in the data. These findings give interpretive validity to the Wechsler scoring structure featuring VCI, PRI, WMI and PSI as summary scores of their respective subtests, and then the FSIQ as a summary score of the 4 ability composites.

The Watkins (2006) direct hierarchical model also demonstrated good fit, and the best fit of all the tested models. Cross-cultural support for the direct hierarchical model was recently found with the WISC-IV (Spanish; Wechsler, 2005) (McGill & Canivez, 2016). The direct and indirect hierarchical models which featured 1st order verbal crystallized, non-verbal crystallized, processing speed and perceptual reasoning factors also fit the T&T sample data well. Fit indices for both of these models placed them second and third respectively to the Watkins model. If found to be replicable, the results will increase the options available for interpretation of the WISC-IV(US) scores of T&T children. These options will be discussed briefly here and in more detail in Chapter 8.

The Wechsler (2003b) indirect model is essentially the current scoring structure of the WISC-IV (US) (Keith, 2005). The 10 core subtest scores are used to calculate the four ability indices (VCI, PRI, WMI, PSI) from which the general intelligence score (FSIQ) is computed (Wechsler, 2003b). Since the FSIQ is a summary score of the 4 indices and not the 10 core subtests, severe discrepancies among index scores may decrease the validity of the FSIQ as an estimate of general intelligence (Saklofske, Rolfhus, Prifitera, Zhu & Weiss, 2005). In such a case, interpretation of the individual ability scores may be the best alternative (Saklofske, Rolfhus, Prifitera, Zhu & Weiss, 2005). Despite arguments against profile analysis (see Watkins, Glutting, Lei, 2007; Freberg, Vandiver, Watkins & Canivez, 2008; Watkins, 2010), there is more to be gained in clinical practice by using this approach, especially if the design of interventions require identification of areas of cognitive strength and weakness (Kubiszyn, Meyer, Finn, Eyde, et al. 2000; Weiss, Prifitera, Holdnack, Saklofske, Rolfhus & Coalson, 2006). Additionally, the pattern of scores among the ability indices can provide useful information for diagnostic decision making. For example, significant weaknesses in processing speed and working memory have been associated with diagnoses of Attention Deficit Hyperactivity Disorder and Specific Learning Disability (Thaler, Bello & Etcoff, 2013; Cornoldi, Giofre, Orsini & Pezzuti, 2014; Fenollar-Cortés, Navarro-Soria, González-Gómez, García-Sevilla, 2015). Also, elevated scores in Similarities and Matrix Reasoning, and lowered scores in Comprehension and Processing Speed have been associated with Autism Spectrum Disorders (Calhoun & Mayes, 2005; Mayes & Calhoun, 2006; Oliveras-Rentas, Kenworthy, Roberson-III, Martin & Wallace, 2012).

The direct higher order factor is argued to be more useful than specific ability scores (Watkins, 2003 & 2010), whether or not there is significant index score scatter. Indeed, studies have demonstrated that FSIQ remained a strong predictor of academic achievement whether or not there was considerable variability among index scores (Watkins, Glutting, Lei, 2007; Freberg, Vandiver, Watkins & Canivez, 2008, Rowe, Kingsley, & Thompson, 2010; Watkins, 2010).

With regards to models 7 and 8, it is difficult without conjecture to give meaning to these models. The clustering together of Picture Concepts and the Working Memory subtests was argued in this thesis to reflect an underlying influence of exposure and education. The extent to which a hypothesized non-verbal crystallized ability cluster is valid in a real world context can only be tested by examining the linear relationship between this factor and the hypothesized environmental variables. Future studies should use structural equation modelling to examine the relationship between antecedent variables of education and US exposure and the non-verbal crystallized ability factor.

Conclusion

This study was limited by methodological setbacks related to selection bias. Therefore, further research is necessary before it can be determined whether the results generalise beyond the subset of the population used in this study. The following chapter will report on a study that was conducted to investigate the relationship between WISC-IV (US) performance and key antecedent variables: schooling, income, US exposure and parental education. Additionally, the results of an examination of the relationship between academic achievement and the following independent variables: WISC-IV (US) measured IQ, income, parental education and schooling on in this sample will be presented. Results are hoped to contribute to a broader interpretive model which will be assessed in Chapter 7.

CHAPTER 6

THE EXTERNAL VALIDITY OF THE WISC-IV(US)

This chapter will present and discuss the findings of studies that were conducted to explore the external network of the WISC-IV (US). To meet the aims of this chapter, Study 5 investigated the relationship between performance on the WISC-IV (US) and antecedent environmental variables. It was hypothesized that parental education, US exposure, school type and school performance will show statistically significant associations with WISC-IV (US) performance in this sample of T&T children. Study 6 investigated the relationships among intelligence, school performance and academic achievement. It is hypothesized that intelligence and school performance will explain a significant portion of variance in academic achievement in this sample.

Study 5

The Relationship Between Environmental Variables and WISC-IV (US) Performance

The WISC-IV (US) Technical and Interpretive manual (Wechsler, 2003b) conceptualizes intelligence as a multi-levelled factorial model within a network of inter-correlated variables See Figure 6.1). Figure 6.1 provides a summary of the broad model which has been adapted from the Wechsler standardization studies (Wechsler, 2003b). The Wechsler model consists of the 10 core subtests, four 1st order ability factors, a higher order general factor and a number of external variables. These external variables include academic achievement, adaptive behaviour, memory, attention, and intellectual giftedness. A summary of the correlations between FSIQ and these external variables is provided in Table 2.4 of Chapter 2. Missing from the network are antecedent biological and environmental variables, but these are perhaps implied features of the Wechsler's framework. Using the data from T&T children, this thesis aims to shed some further light on the nature of this network, including both antecedent and outcome factors.

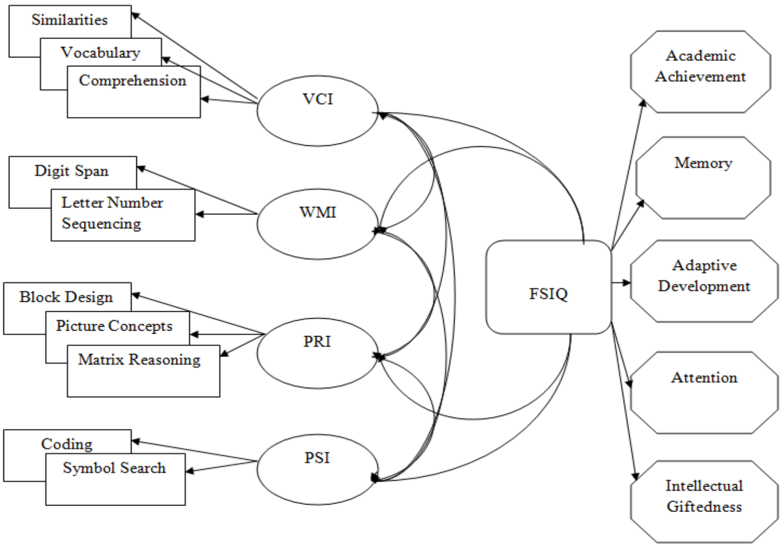


Figure 6.1: Hypothetical WISC-IV(US) nomological network adapted from WISC-IV(US) Validity Studies (Wechsler, 2003b)[11].

With regard to the antecedents, it is known that SES and education explain a significant portion of variance in WISC-IV (US) performance (e.g. Otrosky – Solis & Lozano, 2006; Weiss, Harris, Prifitera, Courville et al., 2006; Gregoire, Saklofske, Van de Vijver, Wierzbicki, et al., 2008). Additionally, parental education was shown to explain about 18% of the variance in FSIQ scores (Weiss, Harris, Prifitera, Courville et al., 2006). It is important to remember, however that since parental education is confounded by intelligence, these associations may reflect a gene-environment correlation. Another environmental variable, income explained a further 3% of the variance in FSIQ performance in the same sample (Weiss, Harris, Prifitera, Courville et al., 2006). In Chapter 2, the results of the pilot studies pointed to the possible influence of SES, US

---

[11] Note: This is a simplified interpretation. The five supplemental subtests as well as the specific associations between WISC-IV (US) factors and components of the outcome variables are not presented here.

exposure and quality of education on WISC-IV (US) model fit in T&T children. Additionally, in Chapter 5, the strong inter-correlations among the Letter Number Sequencing, Matrix Reasoning, Picture Concepts and the verbal comprehension subtests were hypothesized to reflect the contribution of informal and formal education. In light of the above findings, the relationship between WISC-IV (US) measured intelligence and the environmental variables of SES (parental education and income), quality of education and US exposure to will be further examined.

Analyses

To meet the aims of this study, correlational analyses were performed. Also, four multiple regression analyses were conducted using SPSS 21 to examine the relationship between the ability factors, VCI, PRI, WMI and FSIQ, and four hypothesized antecedent variables of School Type, School Performance, Parent Education, and US Exposure. One additional hierarchical analysis was performed in which the dependent variable was PSI. For this analysis, the Gender variable (Male = 1; Female = 2) was added to the list of independent environmental variables because of its moderate correlation with PSI. Gender was entered in the first step and the hypothesized antecedent variables were entered in the second step. Because homogeneity of variance assumptions for the US Exposure variable were not met, analyses using Welch ANOVA with Games-Howell were conducted. These post hoc analyses are recommended for use in the case of unequal variances and unequal group sizes (Field, 2009). Assumptions of univariate and multivariate normality as well as linearity were met.

Results

Inter-correlations

Table 6.1 shows the Pearson correlation coefficients, highlighting the associations that are of significance at the .01 and .05 probability levels.

Table 6.1.

Inter-correlations for Environmental Variables and WISC-IV Global and Specific IQ scores.

	Sch Type	SPer	Age	Gen	PED	USEx	LD	VCI	PRI	WMI	PSI
SPer	.433**										
Age	-.135	-.224**									
Gen	.101	.021	.040								
PED	.468**	.455**	-.024	.135							
USEx	.315*	.482**	-.018	.035	.481**						
LD	-.072	.018	.111	.019	-.031	-.003					
VCI	.242**	.378**	-.123	-.047	.418**	.369**	.016				
PRI	.237**	.374**	-.034	-.035	.300**	.344**	.071	.707**			
WMI	.115	.243**	-.053	-.075	.204**	.277**	-.011	.641**	.654**		
PSI	.168*	.232**	-.060	.144*	.290**	.322**	-.027	.491**	.449**	.417**	
FSIQ	.225**	.374**	-.081	-.027	.366**	.396**	.033	.890**	.875**	.788**	.649**

**. Correlation is significant at the 0.01 level (2-tailed)

*. Correlation is significant at the 0.05 level (2-tailed)

Note. SchType = Private vs Government funded schools, SPer = API score of School, Parent ED = Parental Education, USEx = US Exposure, LD = Learning Disability, PRI = Perceptual Reasoning Index, VCI = Verbal Comprehension Index, WMI = Working Memory Index, PSI = Processing Speed Index, FSIQ = Full Scale IQ.

With the exception of the relationship between WMI and School Type, results show statistically significant correlations among School Type, School Performance, Parental Education, US Exposure and the WISC-IV (US) composite scores. The Gender variable was found to be moderately correlated with processing speed, in that girls performed moderately better than boys on this test. There were no statistically significant differences in IQ scores related to age in this sample. Additionally, the correlations between learning disability and performance on all of the WISC-IV (US) composites were not statistically significant.

Multiple Regressions

Verbal Comprehension. The regression explained 23% of variance in performance on the verbal comprehension index (F (4, 197) = 15.91, $p < .001$). Parent Education, School Performance and US exposure contributed significantly to Verbal Comprehension. Verbal Comprehension was most significantly associated with Parental Education ($\beta = .29$), followed by School Performance ($\beta = .18$). Initial analyses revealed that the US exposure variable which was initially a 4 level categorical variable (1 = Very Little/No Exposure; 2 = Some Exposure; 3 = Moderate Exposure; 4 = Very Much Exposure) contributed significantly to performance on all measures of intellectual ability, however an examination of the post-hoc analyses revealed no statistically significant differences among children who had some exposure, moderate exposure or high exposure.

Analyses were subsequently redone with US exposure as a bivariate variable indicating whether participants had Exposure to US culture (N = 196) or No Exposure to US Culture (N = 26). For the analysis, Exposure to US Culture was assigned a code of 1 and No Exposure to US Culture was assigned a 0. Results revealed a regression which explained 25% of variance in performance on the verbal comprehension index (F (4, 197) = 17.94, p < .001; Table 6.2). In this case, US exposure explained most of the variance in VCI performance (β = .24) followed by Parental Education (β= .21).

Table 6.2.

Standardized β coefficients for predictors of Verbal Comprehension Index (VCI) score after multiple regression.

	VCI Score	
	β	*T*
School Type	.03	.35
School Performance	.17	2.20*
Parent Education	.21	2.57*
US Exposure	.24	3.19**
Regression Model	F (4,197) = 17.94**	
Adjusted R²	.25	

**. *p*< 0.01 level (2-tailed), *. *p*< 0.05 level (2-tailed)

Perceptual Reasoning Index. The regression explained 19% of variance in performance on the perceptual reasoning index (F (4, 197) = 12.64, p < .001; Table 6.3). School Performance and US exposure contributed significantly to Perceptual Reasoning ability. US Exposure (β = .27) demonstrated the strongest correlation with PRI followed by School Performance (β = .22).

Working Memory Index. The regression explained 12% of variance in performance on the working memory index (F (4, 197) = 6.55, p < .001; Table 6.4). Only US exposure contributed significantly to Working Memory ability.

Table 6.3.

Standardized β coefficients for predictors of Perceptual Reasoning Index (PRI) score after multiple regression.

	PRI Score	
	β	*T*
School Type	.11	1.41
School Performance	.22	2.82**
Parent Education	.03	.32
US Exposure	.28	3.50**
Regression Model	F (4,197) = 12.64**	
Adjusted R²	.19	

**. *p*< 0.01 level (2-tailed), *. *p*< 0.05 level (2-tailed)

Table 6.4.

Standardized β coefficients for predictors of Working Memory Index (WMI) score after multiple regression.

	WMI Score	
	β	*T*
School Type	.02	.22
School Performance	.13	1.56
Parent Education	-.02	-.23
US Exposure	.27	3.29**
Regression Model	F (4,197) = 6.55**	
Adjusted R²	.12	

**. *p*< 0.01 level (2-tailed), *. *p*< 0.05 level (2-tailed)

Processing Speed Index. The final regression explained 11% of variance in performance on the processing speed index (F (5, 196) = 5.93, p < .001; Table 6.5). Only US exposure contributed significantly to Processing Speed ability.

Table 6.5.

Standardized β coefficients for predictors of Processing Speed Index (PSI) score after multiple regression.

	PSI Score	
	β	*T*
Gender	.15	2.05*
Regression Model	F (1,199) = 4.79	
Adjusted R²	.02	
Gender	.11	.17
School Type	.04	.56
School Performance	.08	.93
Parent Education	.09	.98
US Exposure	.21	2.55*
Regression Model	F (5,196) = 5.93**	
Adjusted R²	.11	

**. *p*< 0.01 level (2-tailed), *. *p*< 0.05 level (2-tailed)

Full Scale IQ. The regression explained 23% of variance in achievement on FSIQ (F (4, 197) = 16.24, p < .001; Table 6.6). Both US exposure and School Performance contributed significantly to general intellectual ability.

Table 6.6.

Standardized β coefficients for predictors of Full Scale IQ (FSIQ) score after multiple regression.

	WMI Score	
	β	*T*
School Type	.05	.69
School Performance	.18	2.35*
Parent Education	.09	1.04
US Exposure	.31	4.03**
Regression Model	F (4,197) = 16.24**	
Adjusted R²	.23	

**. *p*< 0.01 level (2-tailed), *. *p*< 0.05 level (2-tailed)

Discussion

The aim of the study was to investigate the contribution of the environmental variables, Parental Education, Type of School, School Performance, and US Exposure to WISC-IV (US) performance (VCI, PRI, WMI, PSI and FSIQ) in a sample of 11 and 12-year old T&T children. In the case of PSI performance, the contribution of the demographic variable Gender was also examined. The rationale for exploring the relationship between Gender and PSI performance was provided by the statistically significant correlation found in this study between the two variables, and also by evidence from other research that girls tend to outperform boys on this measure (Keith et al, 2006, Keith, Reynolds, Patel & Ridley, 2008; Goldbeck, Daseking, Hellwig-Brida, Wallman & Petermann, 2010). This study also investigated the association between learning disability (LD) and WISC-IV (US) performance. In Chapter 4, it was suggested that the low processing speed scores of the T&T sample may be explained by the disproportionately large number of children diagnosed with learning disabilities. An examination of the relationship between LD and PSI did not support this hypothesis.

Hierarchical analyses identified a statistically significant effect for **Gender** on PSI performance, meaning that girls outperformed boys on the test of processing speed. The effect size for Gender was found to be very small (2%). Also, once the other environmental variables were added in the second step, the relationship between Gender and PSI performance was significantly weakened. As such, Gender is not considered to be a major antecedent of WISC-IV (US) performance in this sample.

The multiple regression analyses also revealed that **US exposure** explained a significant portion of variance in WISC-IV (US) performance; however, post-hoc analyses revealed that there was no difference in performance among children who had ‘some’, ‘moderate’ or ‘very much’ exposure to US culture. The results indicate that T&T

children who are very unfamiliar with American culture performed poorest compared with other children in the sample. This group of children comprise about 12% of the sample. The figure is surprising since US influence is pervasive in T&T. Also examinations of frequencies revealed that the parents of the Very Little or No Exposure group were less well educated than the parents of children who were exposed to US culture (Table 6.7). Furthermore, 23.1% of the Very Little or No Exposure group attended private school compared to 50% of the group with US exposure (See Table 6.7). These figures demonstrate that the US exposure variable may be confounded substantially by other indicators of SES, such as access to material resources and opportunities for environmental exposure.

Table 6.7.

Comparisons of No Exposure and Exposure group data by Type of School, Age and Parental Education

	No Exposure (N=26) %	Exposure (N=176) %
Type of School		
Public School	76.9	50.0
Private School	23.1	50.0
Age		
11 years old	53.8	50.6
12 years old	46.2	49.4
Parental Education		
0-7 years	53.8	2.3
8-14 years	46.2	34.7
15-19 years		47.2
19+ years		15.3
No information		.6

The US exposure questionnaire was constructed to ascertain the participant’s familiarity with US culture. The initial intention was to gauge US exposure with a Likert’ style scale consisting of 4 options: Very Little or No Exposure, Some Exposure, Moderate Exposure and Very Much Exposure. Admittedly the choice labels were vague on their own, but they were each accompanied by descriptive statements which are summarized in Figure 6.2. After re-examining the descriptors, it was determined that items may have overlapped two or more categories. To illustrate, it is not clear which category should be endorsed if the participant has a basic idea about American culture through television (Some Exposure) but also interacts regularly with Americans (Very Much Exposure). This lack of clarity may have weakened the validity of the scale as a measure of US exposure. Instead the scale's validity may have been strengthened by asking unambiguous questions such as: "How many times a week does the participant look at US news?"; "How many times a week does the participant watch US entertainment shows?" or "How many times has your child travelled to the US?".

How exposed is your child to American culture via television, internet and travel?

☐ Very little (No access to television or internet. My child has never travelled to the United States)

☐ Some (My child watches American television shows and has some idea about American culture music and events.)

☐ Moderate (My child has access to television and the internet, has a moderate amount of knowledge about the American culture and has travelled to the United States on one or two occasions)

☐ Very Much (My child has travelled to the United States on several occasions and is very knowledgeable about American culture through television, internet and also American family members and friends)

Figure 6.2. US Exposure Item of the Demographic Questionnaire.

It is possible that the questionnaire identified children on the basis of their access to television or internet – a possible indicator of poverty, rather than measured US exposure as such. Actual between group differences in the frequency values for School Type and Parental Education, which serve as proxy variables for SES in this study support the hypothesis. The assumption is also supported by the moderate and significant inter-correlations between US exposure and the variables of School Type and Parent Education. It is therefore proposed that the questionnaire can be used as a proxy for identifying children living in environmentally deprived or impoverished conditions.

Children growing up in impoverished circumstances generally lack the resources necessary for intellectual stimulation and therefore suffer from restricted personal growth and creativity as well as lessened opportunities for learning and exploration (Neiss & Rowe, 2000; Petrill, Pike, Price, & Plomin, 2004; Weiss, Harris, Prifitera, Courville et al., 2006). This lack of environmental enrichment has been shown to predict lower IQ scores in children (Weiss, Harris, Prifitera, Courville et al., 2006; Hart, Petrill, Deckard, & Thompson, 2007; Nisbett et al. 2012). The study therefore may have revealed an association between impoverishment and WISC-IV (US) performance in T&T children. Based on the aforementioned discussion, *the US Exposure variable will hereafter be referred to and included in the path model as the Environmental Deprivation variable.*

The results of this study demonstrated that **School Performance** was also related to WISC-IV (US) performance. T&T children who attended schools with higher API scores performed better on the VCI, PRI and FSIQ than children from low performing schools. The findings are consistent with studies that link verbal and reasoning ability with schooling (Ceci & Williams, 1997; Cahan & Cohen, 1989; Gottfredson & Saflofske, 2009; Lohman & Lakin, 2009; Rinderman, Flores-Mendoza & Mansur-Alves, 2010).

Implications of the findings must be considered in the unique context of the T&T education system which will be discussed in detail later in the chapter. The results suggest that students may be benefiting intellectually from attending institutions with higher performance rankings. Conceivably, the higher performing T&T schools, a large number of which are private or denominational, have greater access to the teaching and material resources necessary for optimal cognitive growth. Alternatively, many top performing T&T schools are known to actively recruit brighter students in order to maintain their high performance levels. In such a case, the higher average intellectual ability of the school population is expected to result in better overall school performance. In the following analyses, *School Performance will be included in the broader path model as a covariate of VCI and PRI.*

Parental Education demonstrated a significant positive relationship with VCI performance only. The correlation between VCI and parental education in the T&T sample was found to be moderate (r = .42) and similar to the correlation between parental education and FSIQ scores in a sample of US children (r = .43) (Weiss, Harris, Prifitera, Courville et al., 2006). The results suggested that children of highly educated parents demonstrated a greater fund of formal academic knowledge than children whose parents were not as well educated.

The higher crystallized ability scores may reflect strong parental attitudes about academic achievement (Davis-Kean, 2005; Dubow, Boxer & Huesmann, 2009). Also, in the extremely competitive T&T education system, parents with higher levels of education are more likely to provide their children with the best chances for success through thoughtful selection of schools and tutoring programmes, and by providing academic support and guidance at home. Highly educated parents are also expected to make more realistic assessments of their children’s academic potential and performance and are therefore likely to be better at optimizing academic skill development by modifying interventions to meet their children's specific needs (Alexander, Entwisle & Bedinger, 1994). *Parent education will therefore be included in the broad interpretive model as a predictor of VCI performance in the T&T sample.*

The study did not find any association between School Type and WISC-IV (US) Performance. As public or private school attendance was used in this study as a proxy variable for income, the results demonstrated that there was no difference in WISC-IV (US) performance between T&T children based on family income. These results are inconsistent with some of the findings of previous studies (eg. Mercy & Steelman 1982; Neiss & Rowe, 2000; Ganzach, 2014), but also find support in the literature (e.g. Lemos, Almeida & Colom, 2011). The absence of a significant effect for income on IQ scores seems to belie the strong correlations among School Type, Parent Education and School Performance.

Still, an explanation for the unexpected findings can be offered. The contribution of income to WISC-IV performance may have been weakened by the presence of the other SES variable (Parent Education) in the model and perhaps even the US Exposure (now *Environmental Deprivation*) variable. Chapter 7 of this thesis presents the results of the Structural Equation Modelling (SEM) that tests this assumption. Additionally, School Type may have been a weaker measure of income than first anticipated. In the past, private schools were affordable mainly to persons of the upper classes, however, in recent times, an increasing number of middle class parents have been investing in private school education for their children. Additionally, this study found that a little less than one quarter of the children whose parents reported little or no US exposure also attended private schools. This additional data may also weaken the School Type variable as a proxy for income.

Study 6

School Performance and Global Ability as Predictors of Academic Performance.

Beyond individual success, predicting academic performance can have social and economic implications for the wider society. Economically, countries benefit from having a highly educated population. Data from 23 countries show that reading ability and years of schooling are positively correlated with national earnings from employment (Hanushek & Zhang, 2009). Similarly, Hanushek and Woessman (2008) demonstrated that improvements in academic skills are associated with increases in income, income distribution, and economic growth. For that reason, countries invest substantially towards the monitoring and enhancement of their education systems (Spinath, Freudenthaler, & Neubauer, 2010). Where the focus is on strengthening national academic performance, policy-makers are interested in the environmental contributors to educational achievement.

So far, analyses of twin data show that genetic influences far surpass environmental contributions to academic achievement. Estimations are that genes contribute to between 50% and 80% of the variance in academic performance, and 20% to 30% is explained by the shared environment such as family, neighbourhood and school factors (Harlaar, Spinath, Dale & Plomin, 2005; Friend, DeFries & Olson, 2008; Friend, DeFries, Olson, Pennington, et al., 2009). Furthermore, environment serves a mediating function between hereditary factors and phenotypic expression; therefore, genetic potential is maximized when the environment provides optimal support for academic skills development (Olson, Keenan, Byrne & Samuelsson, 2014).

As with cognitive ability, estimations of genetic contributions to academic performance are contingent on the extent of environmental variation within the sampled population (Olson, Keenan, Byrne & Samuelsson, 2014). So that in populations that demonstrate less variability in terms of wealth distribution or access to education, heredity estimates are expected to be higher than in heterogeneous populations where the distribution of these resources are less even. Correspondingly, environmental contributions are expected to exceed genetic influence in heterogeneous populations because the shared variables necessary for maximum genetic expression are not equally distributed throughout the population. As an example, an international study of factors related to word recognition skills in children found that prior to school, literacy-promoting variables identified as ‘parent reading behavior’, years of parent education, and parent-initiated literacy based activities were more influential than hereditary factors in accounting for individual differences in reading competency (Samuelsson, Byrne, Wadsworth, Corley, DeFries, Willcutt, Hulslander & Olson, 2007). However, after one year of formal schooling, environmental input decreased substantially. Samuelson and colleagues argue that shared environment diminished as an explanatory factor once there was more equitable access to literacy instruction.

The structural model, described in this thesis, features the environment as both an interacting and intervening variable in the relationship between genes and academic performance. However, specific environmental contributors to this model must be discussed. Quasi experimental studies show that socioeconomic status (SES) variables such as parental education and family income are positively related to academic performance (White, 1982; Hanushek, 1986; Sirin, 2005). Furthermore, the relationship between SES and academic achievement is mediated by parental expectations, additional provision of academic resources and parental support towards building academic skills in the home (Davis-Kean, 2005; Sirin, 2005). Sirin (2005) also demonstrated that SES indirectly affects academic achievement through school choice and classroom environment. The implication is that beyond family variables, school and classroom factors are also associated with academic outcomes.

But what are these school related effects? A US study by Coleman (1966) challenged the then-prevailing view that variations in academic performance could be explained by such features as curriculum and school facilities. According to Coleman, these between school factors had relatively little impact on academic performance in US students. Rather, Coleman demonstrated that a substantially larger portion of educational achievement could be explained by teacher characteristics such as level of training and verbal skills. Coleman’s pioneering work was criticized for its survey methodology which reviewers argued did not account for possible contamination by covariates such as school selection and teacher assignment practices (Rivken, Hanushek and Kain, 2005); but in spite of its limitations, Coleman’s findings found empirical support in the Tennessee Class Size Experiment (Nye, Konstantopolous and Hedges, 2004). Using randomized assignment, Nye and colleagues found that teacher effects are at least twice as large as school effects. In fact, over and above genetics, cognitive ability and family factors, teacher effects account for approximately 7% of variance in student achievement (Kovas, Haworth, Dale & Plomin, 2007; Byrne, Coventry, Olson, Wadsworth, et al., 2010). Additionally, the results of education production function studies indicate that teacher characteristics of experience and education are positively correlated with academic performance (Rivkin, Hanushek & Kain, 2005).

If we view teacher recruitment as resource acquisition, one may have a strong rationale for hypothesizing that a significant portion of the observed differences in performance between individuals can be explained by school membership. Research shows that higher income schools generally outdo low income schools at attracting well experienced and trained teachers (Darling-Hammond, 1995; Lankford, Loeb & Wyckoff, 2002). The Coleman (1966) case studies found the influence of school resources on educational outcomes was stronger for minority students than for majority students. Uneven access to school resources (including effective teachers) within minority student populations seemed to explain this statistical interaction. A key finding from the Coleman (1966) report is that heterogeneity of access to resources seems to act as a moderator between school factors and academic achievement; and of course, by including both private and public schools in his later research, Coleman indeed found an effect for school membership (Coleman, Hoffer & Kilgore, 1982). Coleman and colleagues (1982) examined the difference in academic performance between children attending public and private schools in the United States. Results demonstrated that after controlling for background factors such as SES, race, family size and structure, academic resources in the home and parental aspirations, private school students demonstrated higher levels of academic achievement and academic growth than public school students. The results seem to strengthen the argument that when resources are not evenly distributed the effect for school membership increases.

The heterogeneity proposition in favour of school effects is further bolstered by cross cultural research which has demonstrated strong school effects in poor countries and weaker effects in richer countries (Gamoran and Long, 2006). Gamoran and Long (2006) noted that compared to 20% in US samples, 40% to 60% of variation in academic achievement can be attributed to between school factors. Other studies also have demonstrated a large effect for school resources even after controlling for family background factors (Willms & Somers, 2001; Buchmann, 2002). Montagnes (2001) explains that these findings reflect differences in educational resource variance between rich and poor countries. This assertion seems to have some merit. An examination of 2011 World Bank figures on government expenditure on education revealed that wealthier

European countries spent in excess of 15% of GDP per capita on each primary school. This is compared to Latin American and Caribbean where 2/3 of the countries listed invested less than 15%. There is some overlap (e.g. Jamaica, Brazil, St. Vincent, St. Lucia), however, it is likely that developed countries are better able to allocate learning resources more even-handedly thereby reducing variance in this area, while the same may not be true for poorer countries. Hypothetically, poor governments may focus larger portions of their budgetary allocations on certain subpopulations, in urban and heavily industrialized areas for example, leaving other populations wanting. Alternatively, poorer governments may choose to distribute limited resources evenly, however the result is an education system that as a whole is lamentably short of infrastructural benchmarks. In either case, with access to public education so severely limited, only the wealthy can opt for quality education through privately funded institutions. In either case school effects are likely to be large.

Trinidad and Tobago (T&T), one of the wealthier states within the Latin American and the Caribbean region, does not have problems allocating resources towards education. In fact, the largest allocation in its National Budget for 2015 (approximately 33% of total expenditure) went to the Ministry of Education (MOE) (Trinidad and Tobago Ministry of Finance and the Economy, 2014). In spite of these facts, there is a perception of unequal distribution of educational resources within the education system which might contribute to differences in academic performance. These inequities arguably are based in T&T’s history and culture which will be discussed later in the chapter. The following sections will provide a rationale for the hypothesis that school membership is related to academic success in T&T children.

Academic success in a developing nation: The Trinidad and Tobago education system. For developing states that are working at enhancing their position within a fiercely competitive global market, the development of an educated workforce is a matter of high priority. T&T continues to modify its developmental plans and policies in order to meet its stated goals. Topmost among its policies is the enhancement of a world class education system to produce a competent and productive workforce. A related goal is to also create a system wherein every child has the opportunity to realize his or her academic potential (Trinidad and Tobago Ministry of Education, 2011).

In the short term, the drive for an academic system that meets international standards has produced some favourable results. Data from the 2009-2010 report on global competitiveness placed the T&T primary education system at position 39 out of 134 countries, the general education system in 35th place, and math and science in 27th place (DeLisle, Seecharan, & Ayodike, 2010; & World Economic Forum, 2010). Currently T&T ranks at position 45 for general and primary education and a position of 36 out of 148 countries in math and science education (World Economic Forum, 2014). Beyond governmental initiatives, one can also argue that T&T owes its international success to a history of elitism and intense competition within its education system. This view is supported by international studies that have actually identified major problems of inequity across schools in T&T (World Bank, 1995; Mullis, Martin, Kennedy, & Foy, 2007)

Campbell (1996), in his book on the social history of education in T&T, wrote that following the abolition of slavery in 1834, the British colonial government of Trinidad and Tobago announced its intention to educate the ‘Coloureds’ (persons of mixed African and European ancestry) and newly emancipated Blacks on the island. Though the popular rhetoric at the time espoused building equity within the system, the strategy of opening education to all irrespective of race, creed, religion, or financial status was more likely an attempt to anglicize the population, the majority of whom were oriented to French and Spanish languages and cultures. About a decade later, immigrants from India began arriving to the island of Trinidad to work as indentured labourers, followed by the Chinese in 1853. By this time, Trinidad had become a truly heterogeneous society with a large diversity of languages, religions, ethnicities, and practices, thereby making it crucial for the British leadership to ensure that an education system was in place to produce an English speaking population, but also one with British values (Campbell, 1996). According to Williams (1964), this strategy had very little to do with changing an existing elitist system that was effectively restricted to children of the minority elite in the society: the French, Spanish, and English, who could either afford secondary education, or who were employed in privileged senior positions in the public service.

According to Williams (1964), between 1859 and 1869, there were only 3 secondary institutions in Trinidad: a public male secondary school, a female catholic secondary school, and a male catholic secondary school. Of the 206 pupils enrolled at the public school during that decade, less than a fifth of the student population were coloured, and none of the students were of either African descent or East Indian descent. A similar situation existed at the denominational schools that generally enrolled children of the upper and middle class French and Spanish members of the Roman Catholic Church.

Indeed, education, particularly at the secondary levels, was reserved only for the privileged class in the society, some of whom would then be offered places in tertiary institutions abroad (Williams, 1964). It was only in 1879, with the introduction of the College Exhibition (De Lisle, 2012), that secondary education opened up to the poor children of African or mixed descent in Trinidad and Tobago. The College Exhibition was an entrance examination offered to a select number of pupils at the primary levels who were deemed to have the potential to progress to secondary education. No more than 3 or 4 of these scholarships were available every year (Campbell, 1996). The limited availability of these awards must have been a source of great anxiety for many poor and middle class families for whom education was their only path towards upward social mobility, and who could not otherwise afford to send their children to school.

The following years saw a growth of denominational and public secondary schools, as well as an increase of both private and government funded scholarships for secondary education, and by the 1930s, the College Exhibition opened up to all primary school students increasing the competition for limited secondary school places (Campbell, 1996). This development, along with the expansion of primary schools across the country, and the increasing prosperity of the post-World War II population resulted in a greater demand for secondary school places in the 1950s (Williams, 1964).

With independence in 1962, policy makers aspiring to distribute educational resources more equitably at the secondary level, embarked on a programme of expansion of secondary education institutions which at the time accommodated less than 40% of children leaving primary school (De Lisle, Seecharan, & Ayodike, 2010). And so, from three government secondary schools in 1957, the number of public secondary schools increased to 21 in 1967 (Campbell, 1996). Campbell noted that this was in addition to the 23 denominational secondary schools that were in the system at that time.

In the late 1970s, the new government policy of free education increased access to secondary school (London, 1989). By then, the College Exhibition had been replaced with the Common Entrance Examination which determined if and where the child would be placed after primary school. Therefore, once the child passed this national exam, and there was space available, they would be placed in a secondary school. Still, the portion of students admitted to secondary school remained at an unsatisfactory level. In his study on equality in the 1980s education system, Baksh (1984) proposed that the education system in Trinidad and Tobago continued to restrict the mobility of the lower classes, as students from lower socio-economic backgrounds tended to be assigned to more vocational secondary institutions than traditional grammar schools.

In the early 2000s, a policy of universal secondary education was introduced to ensure that all primary school children accessed places at secondary school once they were of age and were deemed ready to transition to higher education (Trinidad and Tobago, Ministry of Education, 2014). Although the philosophy was one of egalitarianism in structure and delivery of education services, the postcolonial inequality and elitism remained within the system. It has been argued that major contributing factors to the continued inequity within the current system are (1) differentiation and diversity in the quality of education providers at the primary and secondary level, and (2) longstanding beliefs and preferences that perpetuate elitism within the system (De Lisle, Seecharan & Ayodike, 2010).

Based on figures in the Education Sector Strategic Plan 2011 – 2015 (Trinidad and Tobago Ministry of Education, 2011), the primary school system consists of 476 public schools, 71 private schools, and 12 specialized instruction schools. Of the public schools, 137 are government funded and managed and 339 are government-funded and religious board-managed. Private primary schools are considered to be elite schools as they have traditionally been the choice of the upper class in the society and have produced some of the best results at the national examinations (Anderson, George & Herbert, 2009). Traditionally private primary schools were reserved for the wealthy descendants of the Spanish, French, and English elite in the society, however in current times, they have become within reach of university educated parents and other upwardly mobile parents for whom the school fees are manageable. Differentiation also exists within the public school system, where the distinction is made between the government-managed public schools and the denominational public schools. Like private schools, denominational schools historically have been regarded as more desirable as they attracted the elite members of church. Beyond that, they have been admired for their perceived stricter models of discipline and higher standards of performance (Anderson, George & Herbert, 2009), as well as their focus on religious and moral instruction.

Differentiation among primary schools. The data seem to support the notion of class based differentiation among schools within the primary education system. Every year, the Ministry of Education uses its National Tests as a means of measuring a school's ability to meet national benchmarks of academic performance in Language Arts, Science, Social Studies and Mathematics, and also to make comparisons between schools (Anderson, George & Herbert, 2009).

Based on the average performance of its students on the national tests, a school is categorized under one of four performance indicators as follows:

- Excelling (an extremely high number of students exceed benchmarks)
- Mostly Effective (an adequate or high number of students meet benchmarks)
- Academic Watch (an inadequate number of students meet benchmarks)
- Academic Emergency (an inadequate number of students meet benchmarks and urgent intervention is required) (De Lisle, Smith, Lewis, Keller, McDavid et al., 2009)

Data from the 2014 academic performance indices compiled by the Trinidad and Tobago Ministry of Education revealed that out of the 525 schools assessed (57 Private and 468 Public schools) 38 exceeded benchmarks and 99 were on academic watch. Of the 38 excelling schools, 68.4% were private schools. Of those on academic watch, only 5%

were private schools. Of the public schools that met the criteria of excellence on the national tests, 66.7% were denominational schools.

Differentiation among secondary schools. The 2012 Strategic Plan report (Trinidad and Tobago Ministry of Education, 2011) listed 134 public and 29 private secondary schools. Of the public schools 91 are government run and 43 are denominational. The situation at the secondary level differs from the primary school system. At the secondary level, private schools are typically lowest in status and are by and large the option for students who are expected to struggle in the mainstream; however, the competition exists between the denominational secondary schools and the government secondary schools (Jackson, 2008). According to Jackson (2008), denominational secondary schools have been most favoured by parents. Although they make up less than one third of the secondary schools in T&T, denominational schools have traditionally produced the best results at higher levels of national assessment. Furthermore, the majority of secondary school graduates who are awarded government scholarships for study at local and international universities come from denominational secondary schools. Following the release of results of the Caribbean Advanced Proficiency Examinations (CAPE) students are offered scholarships for university education based on their performance. Open scholarships fully fund all levels of tertiary education at any university in the world, while Additional scholarships fund tertiary education within any of the campuses of the University of the West Indies (Ministry of Public Administration, 2014). Of the Open Scholarships awarded in 2014, 97.6% were won by students of denominational secondary schools; and for additional scholarships, 92.9% were awarded to denominational schools (Ministry of Public Administration, 2014, Trinidad and Tobago Guardian, 2014).

In spite of government policy that assures every child of a place at the secondary level, there remains stiff competition among primary school students for enrolment at the few high performing denominational secondary schools. Therefore, it is not surprising that the old ‘College Exhibition’ system of selection or ‘gate keeping’ persists.

The Secondary Entrance Assessment

The modern incarnation of the College Exhibition, the Secondary Entrance Assessment (SEA) is a high stakes, ‘one-shot’ examination that is administered to students preparing for transition from the primary level to the secondary level. Since Standard Five is equivalent to year 7 in a Trinidadian primary school and students enrol in the First Year class at 5 years old, then the minimum age for writing the exam is typically 11 years old. Because children move at different paces through the curriculum they can remain eligible to write the exam until they are 15 years old (Trinidad and Tobago Ministry of Education, 2014), failing which they transition seamlessly into a secondary vocational institution – one that is judged to be appropriate to care for their particular needs. In T&T, such schools usually provide interventions designed for children with intellectual and physical disabilities or follow curricula that focus on preparation for employment and independent living (Trinidad and Tobago Ministry of Education, 2014). Otherwise, students may complete a school leaving examination following which they are no longer required to attend school (Trinidad and Tobago Ministry of Education, 2014).

In summary, students eligible to write the SEA generally fall between the ages of 11 and 15 years based on the speed of their progress through primary school and once they have satisfactorily completed the standard Five curriculum (Trinidad and Tobago Ministry of Education, 2014). The scores of this exam help establish the student’s readiness for placement in secondary school, and also determine where the student is placed (Trinidad and Tobago, Ministry of Education, 2014). The examination traditionally measured performance on Language Arts, Mathematics and Creative Writing papers, with the score on each component of the exam contributing to a total composite score. The year 2014, however, saw the introduction of a Continual Assessment component which assesses student performance in Art, Music and Physical Education (Trinidad and Tobago, Ministry of Education, 2014). Placement is based on a total composite score (DeLisle, Smith, & Jules, 2005).

Prior to the exam, parents are asked to request six schools for possible placement. They are asked to rank their choices from one to six, with one being the most highly valued choice, and six being the least favoured choice. This process is complex and determined by many factors (De Lisle, Keller, Jules & Smith, 2009); as such it is difficult to conclusively state what would motivate a parent to choose one school over another. De Lisle and colleagues suggest that academic factors such as the school's reputation for good academic performance as well as non-academic criteria (e.g. discipline, extracurricular activities, religion instruction or gender composition) may influence decision-making. The evidence suggests however, that with very few exceptions, denominational secondary schools tend to be the first choice schools for most Trinidadians. Data from 2001 to 2005 show that the first choice for over 50% of male students and over 60% of female students were denominational schools (De Lisle et al., 2009). Additionally, the 5th and 6th choice schools were all government schools.

Where a child is eventually placed is based on at least one of the following criteria: the child's performance on the examination, the parent's choice of secondary school, the child's gender (if the choice is a single-sex school), where the child resides (Trinidad and Tobago Ministry of Education, 2014) and also the availability of spaces at the school of choice. Because there is more than one criterion for placement, there are no clear guidelines as what scores guarantee placement in one type of school or another. As a qualification to this predominantly merit based rule, 20% of the first year enrolment to denominational schools can be made at the discretion of the school's principal (Trinidad and Tobago Ministry of Education, 1960).

Normally a student has two chances to write this exam. If the student performs below expectations or is not placed in a secondary school of their choice, they have the option of repeating the exam the following year. If the student scores below 30%, they are not eligible for placement at secondary school and are expected to rewrite the exam the following year. Placement is an important event in the education of the T&T child and has major implications for his or her academic future and career. Along with a motivation for success is an avoidance of failure which can prove a major embarrassment for the child, the family and the school as placements are reported in the daily newspapers and on the Ministry of Education website. Conversely, excellent performance can serve as a source of pride for all involved as the top 200 performing students are also featured in the daily newspapers.

With such high stakes involved, the Secondary Entrance Assessment is regarded with much apprehension and anxiety by children, parents and teachers. For at least two years prior to the exam, there is intense preparation. This has traditionally been accomplished by highly specialized teaching methods and extra classes after school and on weekends. In the first year of preparation, many teachers focus on completing the 2-year syllabus, followed by the final year for practice and revision. Some argue that focus tends to be on rote memorization and repeated practice on problems that are likely to appear in the exam. It may well be that success in the SEA is a function of how many hours of drilling occur prior to this exam.

In 2014, 18,500 students completed the SEA and 200 students received recognition and a monetary award for achieving the top scores in the examination. Of those 200 students, 63.5% were from denominational primary schools, 25.5% were from private schools and 11% were from government schools. The data supports the long-held perception that in T&T, denominational and private primary schools and denominational secondary schools outperform government primary and secondary schools respectively.

These between-school differences may be partially explained by selection bias**.** Children who are better at academics will be placed at higher performing secondary schools, and these secondary schools are in turn more likely to demonstrate higher performance averages in future national exams. Additionally, based on the literature about the relationship between selective hiring of teachers and school effects (Darling-Hammond, 1995; Lankford, Loeb & Wyckoff, 2002), both denominational schools and private primary schools are likely to owe their stronger APIs to their recruitment of highly qualified and experienced teachers. With regard to student enrolment, the majority of the private school populations are likely to be from higher SES backgrounds. Otherwise, most schools are expected to have a fairly normal distribution of students based on intellectual ability. However, the truth may actually reveal significant negative skew in certain high performing private and denominational schools. This is because schools have been known to engage in a selection process that effectively rules out children who are less likely to master the curriculum. The mediating role of school performance in the relationship between IQ and academic performance will therefore be explored with the use of SEM in Chapter 7 of this thesis.

In light of the above the analyses of the present study will test the following hypotheses:

1. After controlling for parental education and environmental deprivation, IQ will explain a significant portion of variance in academic achievement.
2. After controlling for parental education and environmental deprivation, School Performance will explain a significant portion of variance in academic achievement.

Analyses

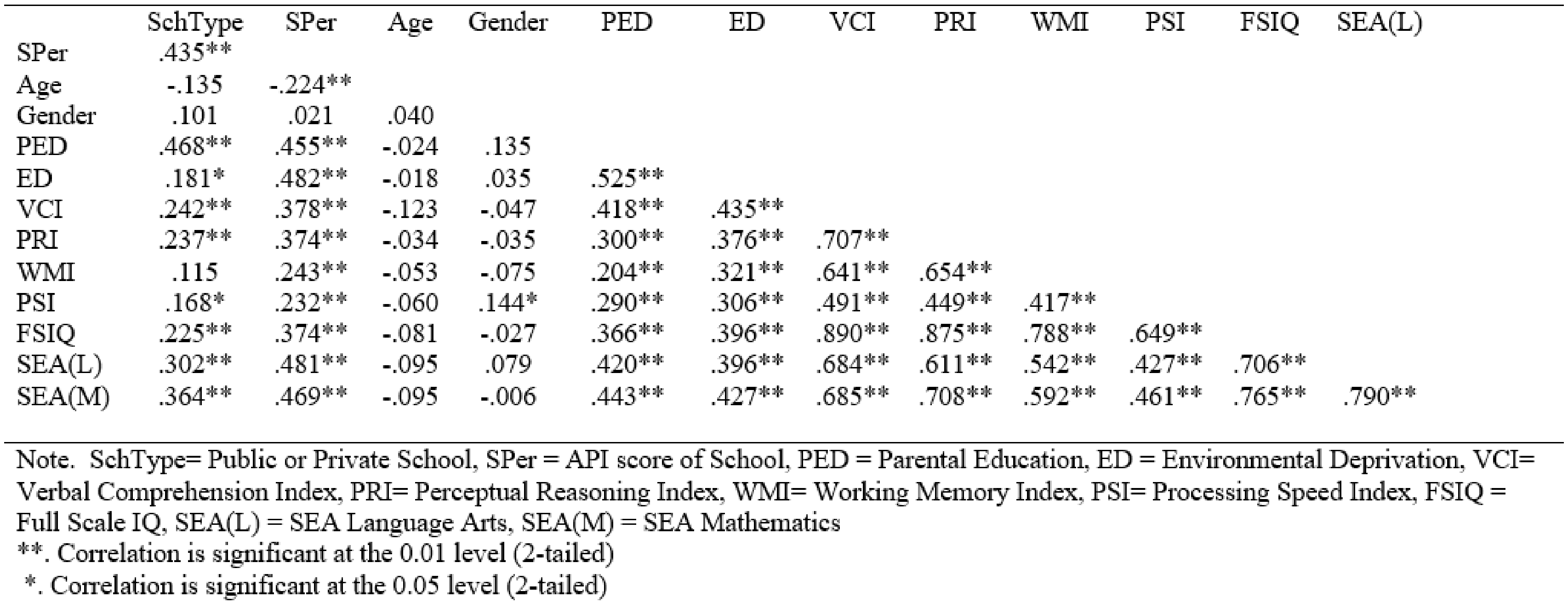
Two sequential multinomial logistic regression analyses were conducted using SPSS 21 to investigate the predictors of one of five levels of academic performance in SEA Language Arts and SEA Mathematics (0-30 marks, 31-50 marks, 51-70 marks, 71 - 85 marks, and 86-100 marks). To control for environmental factors shown in the previous study to be related to IQ, Environmental Deprivation (measured through US exposure) and Parental Education, are entered first, followed by global ability (Full Scale IQ) and then School Performance. For this study, the measure of global ability, FSIQ was included in the model as a predictor of academic achievement because global ability has been shown to account for more variance in academic achievement than specific ability factors (Kahana & Glutting, 2002; Watkins, Glutting & Lei, 2007; Watkins, 2010)

Results

Prior to the regression analyses, Pearson correlation coefficients were calculated for the WISC-IV (US) composites (VCI, PRI, WMI, PSI and FSIQ), the demographic variables Age and Gender, environmental variables of School Performance, School Type, Environmental Deprivation and Parental Education and the academic achievement variables of SEA Language Arts and SEA Mathematics. Coefficients are summarized in Table 6.8.

Table 6.8.

Inter-correlations for Environmental factors, IQ and Achievement.



	SchType	SPer	Age	Gender	PED	ED	VCI	PRI	WMI	PSI	FSIQ	SEA(L)
SPer	.435**											
Age	-.135	-.224**										
Gender	.101	.021	.040									
PED	.468**	.455**	-.024	.135								
ED	.181*	.482**	-.018	.035	.525**							
VCI	.242**	.378**	-.123	-.047	.418**	.435**						
PRI	.237**	.374**	-.034	-.035	.300**	.376**	.707**					
WMI	.115	.243**	-.053	-.075	.204**	.321**	.641**	.654**				
PSI	.168*	.232**	-.060	.144*	.290**	.306**	.491**	.449**	.417**			
FSIQ	.225**	.374**	-.081	-.027	.366**	.396**	.890**	.875**	.788**	.649**		
SEA(L)	.302**	.481**	-.095	.079	.420**	.396**	.684**	.611**	.542**	.427**	.706**	
SEA(M)	.364**	.469**	-.095	-.006	.443**	.427**	.685**	.708**	.592**	.461**	.765**	.790**

Note. SchType= Public or Private School, SPer = API score of School, PED = Parental Education, ED = Environmental Deprivation, VCI= Verbal Comprehension Index, PRI= Perceptual Reasoning Index, WMI= Working Memory Index, PSI= Processing Speed Index, FSIQ = Full Scale IQ, SEA(L) = SEA Language Arts, SEA(M) = SEA Mathematics

**. Correlation is significant at the 0.01 level (2-tailed)

*. Correlation is significant at the 0.05 level (2-tailed)

As can be seen from the table, gender does not correlate significantly with any of the measures. Age is significantly correlated with School Performance and no other variable. All other variables were significantly inter-correlated. FSIQ demonstrated higher correlations with the academic achievement variables than any of the other ability composites. It was therefore included in the following logistic regression analyses as the ability predictor of academic achievement.

SEA Language Arts achievement. Results indicated that the full model was statistically significant (χ^2 = 118.76, p < .001; *df* = 4). The predictors of the model as a group can be said to reliably predict Language Arts performance in this sample of T&T children.

Post hoc analyses produced complicated results in which school performance did not explain the difference in academic performance between the children in categories 2 (31-50 marks) and 5 (86 to 100 marks); and categories 4 (71 to 85 marks) and 5 (86 to 100) marks. To make the model less complicated, the 5 –levelled outcome variable was transformed by combining levels 1 and 2, and levels 3 and 4. The result was a 3-category variable (Level 1 = 0 to 50 marks; Level 2 = 51 to 85 marks and Level 3 = 86 to 100 marks). The results are presented in Table 6.9 and 6.10.

Table 6.9.

Logistic Regression Analysis of the relationship between School Quality and Global Ability and Language Arts Achievement, controlling for Parental Education and Exposure (Reference category is 0-50 marks).

Language Arts Achievement	*B*	SE	Wald	*df*	*p*	Exp (β)	95% CI for Exp (β)	
51-85 marks								
Intercept	-14.18	2.23	40.58	1	.000			
FSIQ	.11	.02	31.76	1	.000	1.12	1.08	1.16
School Per	.02	.01	13.88	1	.000	1.02	1.01	1.03
86-100								
Intercept	-34.57	5.74	36.30	1	.000			
FSIQ	.24	.04	29.84	1	.000	1.27	1.17	1.38
School Per	.03	.01	11.23	1	.001	1.03	1.01	1.05

R^2 = .45 (Cox & Snell), .60 (Nagelkerke)
Note: B = Coefficient for the constant, SE = Standard Error, Wald = Wald chi-square test of the null hypothesis, *df* = degrees of freedom, Exp (β) = odds ratio.

Table 6.10.

Logistic Regression Analysis of the relationship between School Quality and Global Ability and Language Arts Achievement, controlling for Parental Education and Exposure (Reference category is 51-85 marks)

Language Arts Achievement	*B*	SE	Wald	*df*	*p*	Exp (β)	95% CI for Exp (β)	
0-50 marks								
Intercept	14.18	2.23	40.58	1	.000			
FSIQ	-.11	.02	31.76	1	.000	.90	.86	.93
School Per	-.02	.01	13.88	1	.000	.98	.97	.99
86-100								
Intercept	-20.39	5.27	14.98	1	.000			
FSIQ	.13	.04	10.71	1	.001	1.14	1.05	1.23
School Per	.01	.01	2.50	1	.114	1.01	1.00	1.03

R^2 = .45 (Cox & Snell), .60 (Nagelkerke)
Note: B = Coefficient for the constant, SE = Standard Error, Wald = Wald chi-square test of the null hypothesis, *df* = degrees of freedom, Exp (β) = odds ratio.

Mathematics achievement. Results indicated that the full model was statistically significant (chi square = 139.07, p < .001 with df = 4). The predictors of the model as a group can be said to reliably predict Mathematics performance. As with Language Arts, Global Ability predicted Math performance across all categories. In this case School Performance was a predictor of Math performance, except when comparing students who scored between 51 and 85 marks and those who scored above 85 marks. Nagelkerke's R² of .57 and Cox and Snell's R² of .52 suggest that the relationship between the grouped variables and academic performance are moderately strong (See Tables 5.11 and 5.12). To simplify the model, the Mathematics Achievement variable was converted to a three category variable (1 = 0 to 50 marks, 2 = 51 to 85 marks and 3 = 85 to 100 marks).

Table 6.11.

Logistic Regression Analysis of the relationship between School Quality and Global Ability and Mathematics Achievement, controlling for Parental Education and Exposure (Reference category is 0-50 marks).

Mathematics Achievement	B	SE	Wald	Df	*P*	Exp (β)	95% CI for Exp (β)	
51-85 marks								
Intercept	-11.30	1.97	32.82	1	.000			
FSIQ	.10	.02	24.46	1	.000	1.11	1.06	1.15
School Per	.01	.01	6.21	1	.013	1.01	1.01	1.02
86-100								
Intercept	-24.20	3.30	53.78	1	.000			
FSIQ	.22	.03	53.95	1	.000	1.25	1.18	1.32
School Per	.01	.01	6.07	1	.014	1.01	1.00	1.02

R² = .50 (Cox & Snell), .57 (Nagelkerke)
Note: B = Coefficient for the constant, SE = Standard Error, Wald = Wald chi-square test of the null hypothesis, *df* = degrees of freedom, Exp (β) = odds ratio.

Table 6.12.

Logistic Regression Analysis of the relationship between School Quality and Global Ability and Mathematics Achievement, controlling for Parental Education and Exposure (Reference category is 51-85 marks).

Mathematics Achievement	B	SE	Wald	Df	*p*	Exp (β)	95% CI for Exp (β)	
0-50 marks								
Intercept	11.30	1.97	32.82	1	.000			
FSIQ	-.10	.02	24.46	1	.000	.90	.87	.94
School Type	-.01	.01	6.21	1	.013	.99	.98	1.00
86-100								
Intercept	-12.90	2.69	22.94	1	.000			
FSIQ	.12	.02	28.86	1	.000	1.13	1.08	1.18
School Type	.00	.01	.43	1	.514	1.00	.99	1.01

R^2 = .50 (Cox & Snell), .57 (Nagelkerke)
Note: B = Coefficient for the constant, SE = Standard Error, Wald = Wald chi-square test of the null hypothesis, *df* = degrees of freedom, Exp (β) = odds ratio.

Discussion

The aim of the study was to investigate the relationship between the dependent variable, academic performance, and the independent variables of school type, school performance and intellectual ability in a sample of T&T children. The FSIQ score was used as a measure of intellectual ability in this sample. School performance was measured by the Academic Performance Index (API), and Academic Achievement was measured by performance on the SEA Language Arts and Mathematics exams.

Prior to analyses, the Language Arts and Mathematics frequency distributions were found to demonstrate very significant negative skew. Since negative skew was also evidenced in the FSIQ data, one may argue that higher academic scores were produced by a 'more intelligent' sample. This explanation is consistent with the literature (e.g. Sattler, 2001, & Freberg, Vandiver, Watkins & Canivez, 2008), as well as the results of the inter-correlation analyses which reflect a strong statistically significant relationship between global IQ and measures of academic achievement. However, based on the results of the logistic regression it is clear that IQ only partially explained academic performance in the sample.

For Language Arts and Mathematics, the results pointed to an almost consistent effect for school API on academic performance. The findings are supported by studies that report significant school effects in systems where there is inequity in the distribution of educational resources (Coleman, 1966; Darling-Hammond, 1995; Lankford, Loeb & Wyckoff, 2002). Overall, the sequential logistic regression analyses identified that students at higher API schools outperformed students enrolled at low API schools on the Language Arts and Mathematics exams of the SEA, even after controlling for IQ. The results seem to suggest that, excepting selection bias, enrolling a child in a higher performing school places them at an advantage over a child of similar IQ who attends a lower performing school. An exception however was noted. Results demonstrated that school API did not explain the difference between qualifying for a 2nd choice school and qualifying for a 1st choice school based on Language Arts and Mathematics performance. The finding may suggest that after accounting for quality of instruction intelligence is the chief predictor of academic performance and the higher levels.

Results may also point to factors other than School Performance and IQ that can account for how well a child performs on these two exams. Indeed, numerous studies have provided evidence of a link between a number of variables (motivation, interests, personality, learning environment, educational level, gender, social/economic factors, and cultural/historical factors) and scholastic success (Sirin, 2005; Chamorro- Premuzic & Furnham, 2008; Winne & Nesbit, 2010). It is also proposed that this additional explanatory variable may be associated with private tutoring, a phenomenon that is prevalent in highly competitive education systems such as in T&T (Bray, 2009). Bray argued that private tutoring emerged as a way of bringing balance to historically inequitable systems, but instead served to perpetuate the inherent inequalities. In, T&T where many parents seek private tutoring, the competitive edge is gained by children whose parents can access the best quality instructors. Academic success may also be related to the amount of extra tutoring (in hours) that is received.

Conclusion

The first study reported in this Chapter (Study 5) examined the relationship between key antecedent variables and performance on the WISC-IV (US) composites. Results demonstrated that US exposure explained a significant portion of variance in WISC-IV (US) performance. The results however demonstrated that apart from those who had very limited access to US culture, either through television, travel, or contact with Americans, there was no difference in IQ performance among persons who had either had Some, Moderate or High levels of US exposure. The result suggested that the items measuring US exposure were instead measuring some other unidentified variable. An investigation of the correlates of the US exposure variable showed that the participants with no US exposure also attended predominantly public schools, and had parents with lower levels of education. The US exposure variable subsequently was argued to be a proxy measure or covariate of SES, and more specifically impoverishment and environmental deprivation. This variable was subsequently labelled the Environmental Deprivation variable and will be used in further analyses.

The results also demonstrated that school Academic Performance Index (API) predicted performance on the verbal comprehension and perceptual reasoning indices, thus providing additional support for the notion that schooling influences performance on both tests of crystallized ability and fluid reasoning. Finally, Parental Education predicted performance on the verbal comprehension index only, suggesting that highly educated parents may know how to increase academic knowledge in their children better than less educated parents. Gene-environment correlation processes may also be at play.

The second study reported in this chapter (Study 6) showed that global IQ consistently predicted academic achievement, measured by performance on SEA Mathematics and Language Arts. The results also demonstrated that students who attended higher API schools performed better on the SEA. Interestingly, school API did not predict whether the child obtained the 2nd highest or the highest category of scores on SEA Language Arts and Mathematics. This was an interesting finding, as it may have suggested that the difference between just doing well, and excelling in the SEA may not be explained by one's school. The difference for these children may be partially explained by the shadow system of education otherwise known as extra tutoring or 'lessons' (Bray, 2009). The phenomena of 'lessons' is a key part of national exam preparation in T&T. It is believed that parents, who can access as well as pay for the best quality tutoring, strongly enhance their children's chances of securing one of the very limited and coveted secondary school placements in the country.

The field requires a better understanding of the contribution of extra tutoring to academic achievement in T&T, particularly at the higher levels of performance. Unfortunately, such data were not available for this thesis, but variables associated with extra tutoring such as parent education, school type (Income), and Environmental Deprivation may provide interesting information. Based on the findings of the two studies it is hypothesized that the Interpretive Model of WISC-IV (US) Performance in T&T children should include the following variables: Parent Education, School Performance, School Type, Environmental Deprivation and Academic Achievement. The following chapter will assess the validity of the broad path model which will hopefully bring more clarity to the meaning of IQ scores in T&T children.

CHAPTER 7

INVESTIGATING THE CROSS-CULTURAL INTERPRETIVE MODEL

As discussed in Chapter 4 of this thesis, any investigation into the cross cultural validity of the WISC-IV (US) must challenge the broad definition of intelligence upon which the scale was developed. Any such exploration must take account of its internal factor structure, as well as the surrounding network of antecedent and outcome variables. To this end, this chapter will present two studies that assessed the validity of three alternative cross-cultural interpretive models of WISC-IV (US) measured intelligence in T&T children.

Study 7 was conducted to assess the validity of three alternative cross-cultural interpretive models of the WISC-IV (US). The content of the proposed models is guided by the results of studies that were presented in the previous chapters of this thesis and will be presented as path diagrams featuring two main components – the measurement model and the structural model. Study 8 used each alternative model to calculate global ability factor scores. The relationships between the respective derived factor scores and academic achievement were examined and compared for best fit. The following hypotheses were tested:

1. An interpretive framework featuring a direct hierarchical measurement model will provide the best fit to the data. In this model, the global ability factor is measured by school performance and environmental deprivation and the verbal comprehension factor is measured by parental education. Also global ability and school performance predicts academic achievement in this model
2. A factor score estimate derived from the direct hierarchical interpretive framework will be the best predictor of academic achievement in the sample.

Study 7

Investigating the Cross Cultural Interpretive Model of the WISC-IV (US)

Early in the history of the development of the Wechsler scales, racial differences in WISC and WISC-R IQ scores raised widespread concerns about the cross-cultural validity of the test. In an apparent response to these concerns, developers of subsequent editions of the WISC employed a series of rigorous statistical and non-statistical approaches to address problems of possible cultural bias within the test (Wechsler, 2003b). These strategies were useful at the item level but were not as successful at tackling problems in the way behaviours were measured (Weiss, Harris, Prifitera, Courville, et al., 2006). The concerns were related to expectations test developers may have about how children typically respond to questioning by an examiner as well as their familiarity with certain types of testing formats and stimuli (van de Vijver & Tanzer, 2004; He & van de Vijver, 2012). Some argue that the IQ testing format generally adopts a Western European perspective that may be at odds with the behaviours and practices of other cultural groups (Helms, 1992; Weiss, Harris, Prifitera & Courville, et al., 2006).

In spite of these concerns, the literature has consistently provided cross cultural evidence of the structural and predictive validity of the WISC tests (e.g. Ross-Reynolds & Reschly, 1983; Poteat, Wuensch, & Gregg, 1988; Weiss, Prifitera & Roid, 1993; Weiss & Prifitera, 1995; Georgas, Weiss, van de Vijver & Saklofske, 2003). However, as with means comparisons, evidence of structural and predictive validity does not completely rule out bias. In fact, although there may be agreement between two cultural samples about how the variables are organized, ambiguity about the nature of the underlying construct and its environmental covariates remain (see Flynn, 1987).

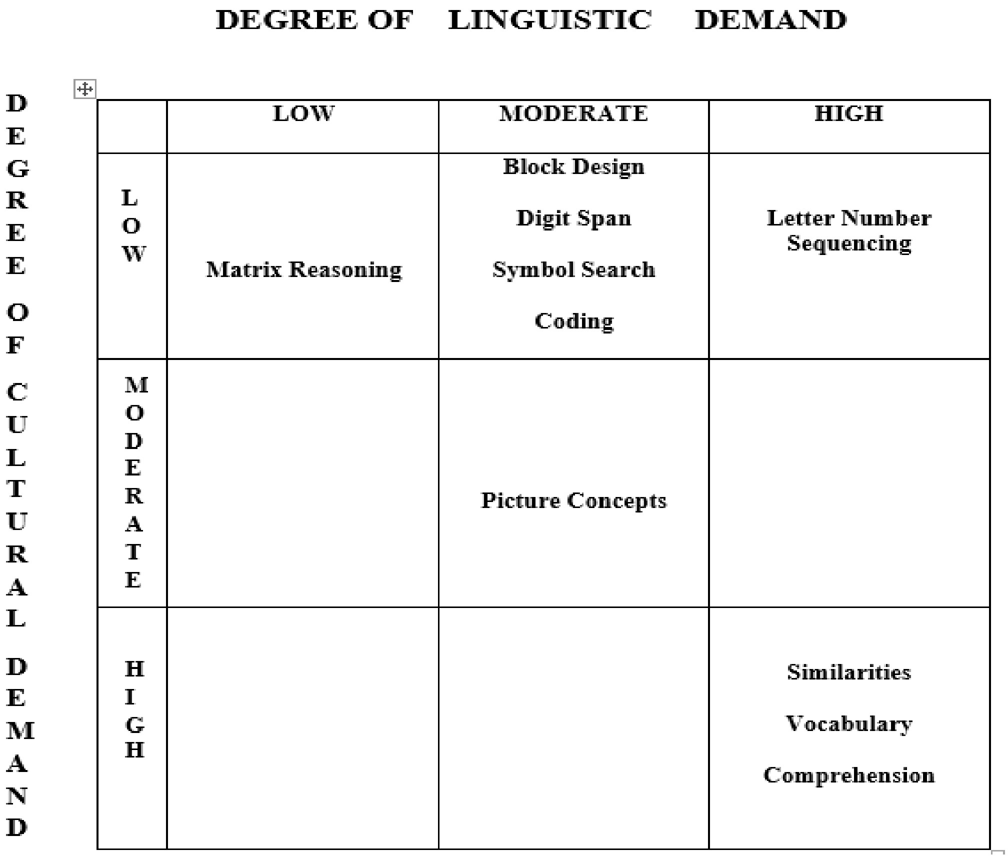
Consider once again the Kenyan study in which practical intelligence was shown to be inversely related to crystallized intelligence (Sternberg, Nokes, Geissler, Prince et al., 2001). Children who were not formally educated scored low on the measure of academic (crystallized) knowledge but high on measures of indigenous knowledge. On the other hand, children attending school scored higher on the measure of crystallized ability. Would a formally educated Kenyan child’s performance on a test of indigenous knowledge provide a fair estimate of their intellectual ability? Conversely, does an indigenous child’s low score on a measure of crystallized knowledge indicate low intellectual ability? Both practical and crystallized ability tests measure knowledge and skill acquisition - one in the natural environment and the other in a formal academic setting (Sternberg & Hedlund, 2002). Therefore, interpretations of the sample scores on either measure should not ignore the environmental context. In other words, the crystallized ability test should be described as a measure of both crystallized ability and exposure to formal education. Similarly, the practical intelligence test is tapping into crystallized ability and exposure to indigenous practices. The distinction is important as the measures cannot provide valid estimates of a particular type of knowledge acquisition, if there has been no access to that knowledge.

In the 12 nation study, Georgas and colleagues (2003) found that Lithuanian children demonstrated the weakest performance across national samples on the Digit Span test of working memory, while, the South Korean sample produced the highest scores. The authors hypothesized that, in spite of structural invariance, language related factors may have explained the performance differences between Lithuanian and South Korean children on the test of working memory. In this case, although the Georgas et al. factor analyses confirmed the grouping of the Digit Span and Letter Number Sequencing tests into a factor consistent with working memory; the results of the means comparisons may have implied the existence of a language covariate.

Many intelligence test manuals insist that cultural factors must be acknowledged in the assessment of intelligence, but it is rare to find an explicit interpretive framework to guide test score interpretations. Flanagan, Ortiz and Alfonso (2007) attempted to standardize US IQ test score interpretations with the Culture-Language Interpretive Matrix (C-LIM). This 9 squared matrix (see Figure 7.1) was developed to determine, based on the cultural loading or linguistic demands of test items, whether a particular person’s latent ability could be validly estimated by the test.

The C-LIM is designed along a vertical and horizontal axis. The vertical axis specifies the degree of cultural demand and the horizontal axis represents the degree of linguistic demand of an IQ test (Flanagan, Ortiz & Alfonso, 2007; Styck & Watkins, 2014). For this approach, the subtests of the test are assigned a cell within the matrix according to its degree of cultural and linguistic demand. For example, a US test with both low cultural and language loading would be at the upper left-hand corner while a test that taps highly into knowledge of the language and culture will be placed at the bottom right corner. The middle square represents the medium level of both cultural and linguistic demand. Figure 7.1 shows how the WISC-IV(US) subtests were assigned within the matrix (Flanagan, Ortiz & Alfonso, 2007). The organization of the subtests were guided by the results of studies which found that test scores declined linearly as a function of how much each subtest taps into an examinee's knowledge of the dominant language and culture (Flanagan, Ortiz & Alfonso, 2007).

To establish the validity of the test as a measure of an examinee's intelligence, the examinee's knowledge of English and their cultural assimilation is assessed. Expectations for performance are then determined. As an example, non-English speaking immigrants are expected to perform generally worse than the normative group, however the degree of attenuation is expected to depend on the level of acculturation and English language proficiency of the examinee. Newly arrived immigrants, for example, are expected to perform the worst on the WISC-IV, while their 3rd and 4th generation descendants are expected to perform more like the normative group (Flanagan, Ortiz & Alfonso, 2007). Additionally, examinees are expected to perform best on Matrix Reasoning and worst on Similarities, Vocabulary and Comprehension. If the pattern of scores meets expectations based on the examinee's English language skills and acculturation, the test may be considered an invalid measure of the examinee’s intelligence. Based on these guidelines, Styck and Watkins (2014) assigned the 10 WISC-IV core subtests to the C-LIM (see Figure 7.1) in order to investigate how well the matrix discriminated among children with different levels of English Language proficiency.



DEGREE OF LINGUISTIC DEMAND

DEGREE OF CULTURAL DEMAND

	LOW	MODERATE	HIGH
LOW	Matrix Reasoning	Block Design Digit Span Symbol Search Coding	Letter Number Sequencing
MODERATE		Picture Concepts	
HIGH			Similarities Vocabulary Comprehension

Figure 7.1. The Culture-Language Interpretive Matrix (C-LIM) for the WISC-IV (Styck& Watkins, 2014).

Neither this nor any previous study provided support for the C-LIM (Styck & Watkins, 2014). From a cross-cultural perspective, practitioners may benefit from the C-LIM as a guide for interpreting test scores. As a means of establishing validity, the C-LIM seems to be less promising. This is because the C-LIM's estimates of language and cultural loadings as well as its criteria for establishing test validity are not supported by data (Styck & Watkins, 2014). The authors' ideas about how the subtests are grouped in a cross-cultural context should have been tested using exploratory and confirmatory factor analyses. Understanding how the subtests are aggregated based on data from samples of varying levels of language proficiency and acculturation would have bolstered the validity of the C-LIM. Information on the actual loadings of subtests on culture or language factors may have certainly provided an empirically based rationale for subtest assignments within the matrix. In fact, any cross-cultural interpretive model should not be proposed without understanding how these key environmental factors fit statistically within the broad intelligence network.

The following section will describe a statistical cross-cultural interpretive model that was developed in 1978 to explain the performance of minority children in the US on the Wechsler scales. It was hypothesized that WISC performance was associated with key socio-cultural variables. A scoring template was then designed to adjust WISC scores in order to provide a more valid estimate of performance in minority children.

Adjusting IQ Scores Based on Socio-Cultural Variables – The SOMPA

As was discussed in Chapter 4, the comparisons of WISC-R performance among American ethnic groups revealed that ethnic minorities performed on average worse than members of the majority (Munford, 1978, Munford, Meyerowitz & Munford, 1980). As a consequence, larger numbers of students from minority groups within the US education system were being classified as ‘mentally retarded’ when they seemed to function normally outside of the school environment (Reschly, 1980). The System of Multicultural Pluralistic Assessment (SOMPA; Mercer & Lewis, 1978, cited in Oakland, 1977/1979; Vazquez-Nuttall, 1979; Reschly, 1980; Johnson & Danley, 1981 and Jirsa, 1983[12]) was developed as a response to this problem. The SOMPA was designed to adjust WISC-R scores based on socio-cultural variables. It was standardized on a sample of 2100 Californian public school students aged 5 to 11 years old. The standardization sample was stratified by ethnicity: 'Anglos' (N = 700); 'Blacks' (N = 700); and 'Hispanic – Americans' (N= 700) (Vazquez – Nuttall, 1979). The samples were administered the WISC-R and a battery of scales, among which were measures of the four sociocultural modalities - family size, family structure, SES and urban acculturation. Mercer and Lewis studied how the relationships between WISC-IV IQ scores and each of the four socio-cultural modalities differed among 3 US ethnic groups – ‘Anglo-American’, ‘Black American’ and ‘Hispanic American’. The correlation of WISC-R Full Scale scores with SES was highest for the ‘Anglo’ sample (β= .39). Also, WISC-R correlations with Urban Acculturation were highest for ‘Blacks’ (.30) and ‘Hispanics’ (.37) (Reschly, 1980). The regression formulae were then used to calculate adjusted scores, otherwise known as Estimated IQ for children from each population by summing each group’s specific raw scores and regression weights (Oakland, 1979, Reschly, 1980). Then the Estimated Learning Potential (ELP) was calculated using the following formula: (Actual IQ – Estimated IQ)/15) + 100. By controlling for the environmental modalities, ELPs were believed to provide a more accurate estimate of intellectual ability.

---

[12] The System of Multicultural Pluralistic Assessment was published by the Psychological Corporation in June of 1978 and no other version of the tool was published after that time. Additionally, NCS Pearson, Inc., which is a successor to the Psychological Corporation has no record of the book, neither does Amazon and other booksellers which lists the book as out of stock. Information about the SOMPA has been acquired from secondary sources.

Problems with the SOMPA were identified very early after its publication. First, the standardization samples were not considered to be representative of minority populations throughout the US (Vazquez – Nuttall, 1979; Reschly, 1980). The SOMPA was also criticized for its poor predictive validity. The correlation between ELP and academic ELP (*r* = .40) was found to be lower than the correlation between IQ and academic achievement (*r* = .64); therefore, ELP was considered to be no more useful than IQ at predicting academic outcomes in children (Oakland 1971/1979; Johnson and Danley, 1981). Additionally, Jirsa (1983) criticized the SOMPA's focus on estimated IQ. According to Jirsa, (1983), ability tests were concerned with current performance, and calculating ELPs did nothing to change the child's current functioning. Jirsa added that Mercer did not supplement ELPs with any useful strategies for improving current levels of test performance. Mercer (1979) however argued that test measures of concurrent validity did not provide a fair assessment of the test’s validity because the SOMPA focused on estimating learning potential rather than current achievement.

Jirsa's argument may be only partially valid. While it may be correct that IQ tests measure current performance, it is only in a literal sense. Current behaviours as measured by intelligence tests are seen as indicators or estimates of true ability. Testers are concerned with the underlying trait – the indicator of what the child is capable of learning - as much as, or perhaps even more than the actual behaviour. If this was not the case, ability tests would be no different from measures of achievement. Achievement which is a measure of acquired knowledge and skills is expected to change over time. The capacity to acquire the knowledge (ability) however, is expected to remain largely stable over time. Even in old age, when cognitive levels are expected to decline, correlations of current IQ scores with childhood IQ scores have been shown to range from .54 and .67 (Deary, 2014).

There is merit therefore in identifying ways by which tests can provide the best estimates of a child's true ability. Mercer's ELP's seemed to aspire to that purpose; however, as Jirsa (1983) suggested, ELPs lacked interpretive value and longitudinal predictive validity without knowledge of how interventions eventually could bring measured IQ levels closer to ELP levels. Indeed, a 10-year longitudinal study with 1184 children of the original standardization sample demonstrated that WISC-R IQ correlations with academic achievement (*r* = .42 for classroom GPA, .58 for Reading, and .68 for Math) exceeded that of ELP scores (*r* = .34 for classroom GPA, .47 for Reading, and .50 for Math)

Another problem with the SOMPA was that the design was based on the flawed premise that all members of any particular US racial group share similar socio-cultural characteristics. Furthermore, the use of three normative ethnic samples in test development raises the issue of generalizability and validity of the scoring framework. With regard to validity, ELP scores are calculated by imputing the weights of each socio-cultural modality into the regression formula, however, the accuracy of the modality weights may be in question. Regression weights for Black and Hispanic children were higher on Urban Acculturation and lower on SES. Are these differences in weights real or a function of within sample variability? Black and Hispanic minority samples are expected to show a narrower distribution in SES than an Anglo sample. In such a case, SES effects are expected to be greater for the Anglo sample. Similarly, the Anglo sample is expected to show less variability on the measure which taps into immersion into the dominant culture. As such, effects for acculturation are likely to be smaller for the Anglo sample compared with the minority groups. A broader sampling approach may have lessened these differences between groups. As it stands, the sampling methodology of the SOMPA study is believed to have negatively impacted the generalizability of the measure. Indeed, replication studies revealed that the regression formula weights could not be generalized to children in other states (Reschly, 1980). Based on the scoring guidelines, Cuban-American children in Florida or Puerto Rican Children in New York City would be assessed using the same norms as Californian Mexican children. Additionally, a Haitian-American child from a French speaking family would have to be compared with Black American norms.

The validity of the SOMPA may have been enhanced by stratified sampling of key populations in the US. This approach might have eliminated the need to assign persons to categories based on weak constructs such as ethnicity or race. Instead, each ELP estimate would have been made on the basis of sociocultural variables alone. With better sampling techniques, the weightings of each modality might have been more reliable. Also the measure may have been further strengthened by accounting for additional socio-cultural variables in the regression formula such as schooling and language.

The theoretical rationale for the SOMPA was strong, but because of serious methodological issues, as well as incremental improvements in WISC item review practices, the SOMPA waned significantly in popularity over the years. The test was eventually discontinued in 2003 (Dominguez de Ramirez, 2007). This is unfortunate because cross-cultural testing is in need of a clear and standardized approach for identifying and measuring key influential environmental variables and estimating how these variables will affect the actual IQ scores of real children. This thesis has underscored the importance of gathering detailed data on environmental variables as a way of adding meaning to WISC-IV (US) test performance. Understanding the contribution of environmental variables to the WISC-IV (US) model is important to making proper estimations of IQ.

Analyses

Measurement model. In Chapter 5, the CFA results supported the Wechsler (2003b) four 1st order oblique factor structure. Also the indirect hierarchical model featuring four 1storder oblique factors and a higher order general intelligence factor that is measured directly by the specific ability factors (Keith, 2005) demonstrated good fit to the data. Of the two models, the indirect model was shown to be the better of the two, and provided further empirical evidence of the validity of the Wechsler (2003b) scoring structure. However, the Watkins (2006) direct hierarchical model demonstrated the best fit of all the alternative WISC-IV (US) models. Based on the EFAs, a model which featured verbal and non-verbal crystallized ability, perceptual reasoning and working memory 1st order factors as well as a higher order general factor was also tested. This model was tentatively named the education/exposure model, the indirect and direct versions of which produced the 2nd and 3rd best fit respectively of all alternative models.

The differences among the model fit indices were observed to be very small. It also was not possible to determine whether the differences in fit were significantly different as comparisons were made between non-nested models. Of the five models demonstrating adequate fit, three were chosen for further analysis. They included the Keith (2005) indirect hierarchical model, the Watkins (2006) direct hierarchical model, and the direct education/exposure hierarchical model. They were included in alternative path models to determine best fit.

Structural model. The structural model features the environmental variables that correlated significantly with WISC-IV (US) composite scores. Results showed that the environmental deprivation variable correlated significantly with all composite areas of WISC-IV (US) performance. School Performance predicted performance on the VCI, PRI and FSIQ and Parental Education predicted performance on the VCI. Antecedent variables are therefore presented in the path diagrams as parental education, school performance, and environmental deprivation.

The outcome variable in the path model is academic achievement. In Study 6 of Chapter 6, results revealed an association between performance on the SEA subjects and both global IQ and School Performance. Notably, School Performance did not explain differences in performance between children scoring at the 2nd choice and those scoring at the 1st choice levels. This finding was argued to reflect the potential contribution of 'extra lessons' to the model. It was argued that the extent to which a child would benefit from 'extra lessons' depended upon parental knowledge about the importance of extra tutoring in the system, parental access to the best tutors, and parental ability to afford good tutoring. In that light, the SES variables, environmental deprivation, parental education and school type (income) were included in the model along with school performance as mediators in the relationship between global IQ and academic achievement.

Structural Equation Modelling

Structural Equation Modelling (SEM) is described as the step that follows CFA (Bollen, 1989). CFA assessed the fit of the internal latent factor structure of the WISC-IV (US), otherwise known as the measurement model. SEM will incorporate CFA and multiple regression to simultaneously assess the measurement model and the structural model (Tabachnick & Fidell, 2007). In other words, SEM explores the correlational matrix to assess the fit of a hypothesized factor structure; while simultaneously examining the linear relationship between independent variables. Independent and dependent variables may be either factors and or observed variables (Tabachnick & Fidell, 2007). As an illustration, the present thesis is concerned with an interpretive model of WISC-IV performance in T&T children. The model is concerned with elucidating the internal structure that best fits the sample data, but also examining how environmental factors such as SES, schooling and US exposure influence the structure of the test as well as the predictive validity of the factors that are measured by the WISC-IV (US). The model is specified by this author as a hypothesis of how the variables and factors are interrelated. It is hoped that through SEM, the models can be assessed for best fit and modified. Using SEM as a means of modifying a proposed model is not without controversy (Tabachnick & Fidell, 2007). Repeated tweaking of a model to provide best fit increases the risk of Type I error (Tabachnick & Fidell, 2007) and so precautions are necessary if this is the intent. Monitoring of significance levels as well as replication in future studies to validate the model is highly recommended (Kline, 2013).

To examine the validity of the WISC-IV (US) cross-cultural interpretive models, SEM was carried out using MPlus6. Two participants were missing data on demographic variables of Parent Education and Environmental Deprivation. These participants' entries were deleted by MPlus6 prior to analyses, leaving 201 cases. To estimate fit, normed chi square (χ^2/df) values were examined. Other goodness of fit indices were used including Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA). CFI and TLI values of greater than .95 and RMSEA values of less than .06 indicate adequacy of fit (Hu & Bentler, 1998, 1999). The hypotheses embedded in the three alternative cross-cultural interpretive models will be described as follows:

Model 1:

1. Verbal Comprehension Index (VCI) is measured by Similarities, Vocabulary, and Comprehension.
2. Perceptual Reasoning Index (PRI) is measured by Block Design, Picture Concepts and Matrix Reasoning.

3. Working Memory Index (WMI) is measured by Digit Span and Letter Number Sequencing.
4. Processing Speed Index (PSI) is measured by Coding and Symbol Search.
5. Global Ability (Full Scale IQ) is measured by VCI, PRI, WMI and PSI.
6. Global ability predicts Academic Performance in Mathematics and Language Arts.
7. VCI performance is predicted by Environmental Deprivation.
8. PRI performance is predicted by Environmental Deprivation.
9. WMI performance is predicted by Environmental Deprivation.
10. PSI performance is predicted by Environmental Deprivation.
11. VCI performance is predicted by School Performance.
12. PRI performance is predicted by School Performance.
13. VCI performance is predicted by Parental Education.
14. School Performance predicts Academic Performance in Mathematics and Language Arts.
15. School Performance acts as an intervening variable between global ability and academic performance in Mathematics and Language Arts.
16. Global Ability acts as an intervening variable between School Performance and Academic Achievement in Language Arts and Mathematics.
17. The SES variables, Parental Education, Environmental Deprivation and School Type act as intervening variables between global IQ and Academic Achievement in Language Arts and Mathematics.

Model 2:

1. Global Ability (Full Scale IQ) is measured by Similarities, Vocabulary, Comprehension, Block Design, Picture Concepts, Matrix Reasoning, Digit Span, Letter Number Sequencing, Coding and Symbol Search.
2. Verbal Comprehension Index (VCI) is measured by Similarities, Vocabulary, and Comprehension.
3. Perceptual Reasoning Index (PRI) is measured by Block Design, Picture Concepts and Matrix Reasoning.
4. Working Memory Index (WMI) is measured by Digit Span and Letter Number Sequencing.
5. Processing Speed Index (PSI) is measured by Coding and Symbol Search.
6. Global ability predicts Academic Performance in Mathematics and Language Arts.
7. Global Ability is predicted by School Performance and Environmental Deprivation.
8. VCI performance is predicted by Parental Education.
9. School Performance predicts Academic Performance in Mathematics and Language Arts.
10. School Performance acts as an intervening variable between global ability and academic performance in Mathematics and Language Arts.
11. Global Ability acts as an intervening variable between School Performance and Academic Achievement in Language Arts and Mathematics.
12. The SES variables, Parental Education, Environmental Deprivation and School Type act as intervening variables between global IQ and Academic Achievement in Language Arts and Mathematics.

Model 3:

1. Verbal Crystallized Index (VCI) is measured by Similarities, Vocabulary, and Comprehension.
2. Non-verbal Crystallized Index (NVCI) is measured by Picture Concepts, Digit Span and Letter Number Sequencing.
3. Perceptual Reasoning Index (PRI) is measured by Block Design and Matrix Reasoning.
4. Processing Speed Index (PSI) is measured by Coding and Symbol Search.
5. Global Ability (Full Scale IQ) is measured by VCI, NVCI, WMI and PSI.
6. Global ability predicts Academic Performance in Mathematics and Language Arts.
7. VCI performance is predicted by Environmental Deprivation.
8. NVCI performance is predicted by Environmental Deprivation.
9. PRI performance is predicted by Environmental Deprivation.
10. PSI performance is predicted by Environmental Deprivation.
11. VCI performance is predicted by School Performance.
12. NVCI performance is predicted by School Performance.
13. PRI performance is predicted by School Performance.
14. VCI performance is predicted by Parental Education.
15. School Performance predicts Academic Performance in Mathematics and Language Arts.
16. School Performance acts as an intervening variable between global ability and academic performance in Mathematics and Language Arts.
17. Global Ability acts as an intervening variable between School Performance and Academic Achievement in Language Arts and Mathematics.

18. The SES variables, Parental Education, Environmental Deprivation and School Type act as intervening variables between global IQ and Academic Achievement in Language Arts and Mathematics.

Results

Absolute fit of each model was evaluated using the normed chi square (χ^2/df) and model fit indices, CFI, TLI and RMSEA. The model fit indices for models 1 to 3 are presented in Table 7.1. The full unconstrained versions of Models 1, 2 and 3 all demonstrated a good fit to the data. The table shows that Model 2, based on the Watkins (2006) direct hierarchical measurement model demonstrated best fit to the data. The second best fitting model was Model 3 (based on education/exposure framework), and the third best model was the expansion on the Wechsler (2003b) scoring structure (Model 1).

Table 7.1.

Goodness of Fit Indexes for Evaluating Model Adequacy of Path Models (N= 203)

Model	χ^2	*df*	*p*	χ^2/df	CFI	TLI	RMSEA (CI90%)
1	87.96	82	.31	1.07	0.99	0.99	.02 (.00 to .05)
2	**84.41**	**85**	**.50**	**0.99**	**1.00**	**1.00**	**.00 (.00 to .04)**
3	85.03	81	.36	1.05	0.99	0.99	.02 (.00 to .04).

Note. χ^2 = chi square, *df* = degree of freedom, *p* = probability, χ^2/df = normed chi square, AIC = Akaike's information criterion, CFI = comparative of fit index; TLI = Tucker Lewis index, RMSEA= ¼ root mean square error of approximation.

Applying Model Constraints.

Model constraints were applied to all three models. To identify the most parsimonious and best fitting model, increasingly constrained nested models were assessed using normed chi square and chi-square difference testing. To determine their value to the model, parameters were removed and then chi-square difference values examined. Important parameters were retained in the model.

Applying constraints to Model 1. The following model constraints were imposed to Model 1 by constraining following parameters to zero in the model:

1. Constraint #1:
    a. The SES variables, Parental Education, Environmental Deprivation and School Type act as intervening variables between global IQ and Academic Achievement.
2. Constraint #2:
    a. School Performance predicts Academic Performance in Mathematics and Language Arts.
    b. School Performance acts as an intervening variable between global ability and academic achievement.
    c. Global Ability acts as an intervening variable between School Performance and Academic Achievement.
3. Constraint #3:
    a. VCI performance is predicted by Parent Education.
4. Constraint #4:
    a. VCI performance is predicted by School Performance.
    b. PRI performance is predicted by School Performance.
5. Constraint #5
    a. VCI performance is predicted by Environmental Deprivation.
    b. PRI performance is predicted by Environmental Deprivation.
    c. WMI performance is predicted by Environmental Deprivation.
    d. PSI performance is predicted by Environmental Deprivation.

Results

Applying Constraint 1 resulted in slightly reduced fit with non-significant chi-square difference values (See Table 7.2). Constraints 2, 3 and 4 resulted in significant changes to the model with good fit, however, constraint 5 resulted in a significantly worsened and poor fitting model ($\chi2$ (94) = 214.53, $p<.001$; CFI=.82; TLI= .80; RMSEA=.08 (90% CI: .06; .09); $\chi2$ diff (4) = 31.96; p <.001; Table 7.2).

Table 7.2.

Goodness of Fit Indices for Full and Constrained Models (N = 201)

Model	χ^2	*df*	*p*	χ^2/df	CFI	TLI	RMSEA (90% CI)
Full Model	87.96	82	.31	1.07	.99	.99	0.02 (.00 to .05)
Constraint #1	96.50	88	.25	1.10	.99	.98	0.02 (.00 to .05)
Constraint #2	98.60	90	.25	1.10	.99	.98	0.02 (.00 to .05)
Constraint #3	99.98	91	.24	1.10	.99	.98	0.02 (.00 to .05)
Constraint #4	102.69	93	.24	1.10	.99	.98	.02 (.00 to .05)
Constraint #5	214.53	97	.00	2.21	.82	.79	.08 (.06 to .09)

Note: χ^2 = chi square, *df* = degrees of freedom, *p* = probability, χ^2/df = normed chi square, CFI = comparative of fit index, TLI = Tucker Lewis index, RMSEA= ¼ root mean square error of approximation.

As a consequence, a model comprising the hierarchical scoring structure of the WISC-IV (US) (Keith, 2005), Parental Education predicting VCI, School Performance predicting VCI and PRI, Environmental Deprivation predicting VCI, PRI, WMI and PSI, FSIQ and School Performance predicting performance on SEA Language Arts and Mathematics, and School Performance mediating the relationship between global ability and academic achievement in both subject areas, and Global Ability mediating the School

Performance/Academic Achievement relationship was found to be the most parsimonious well-fitting model.

In the Structural Model, Parent Education explained 5% of the variance in VCI Performance. School Performance explained 3% and 5% of the variance respectively in VCI and PRI performance. Environmental Deprivation explained 6%, 11%, 11% and 7% respectively in VCI, PRI, WMI and PSI. Figure 7.2 presents the standardized coefficients for the full model. The unstandardized coefficients, standard error, standardized coefficients, and significance values for the full model are presented in Table 7.3.

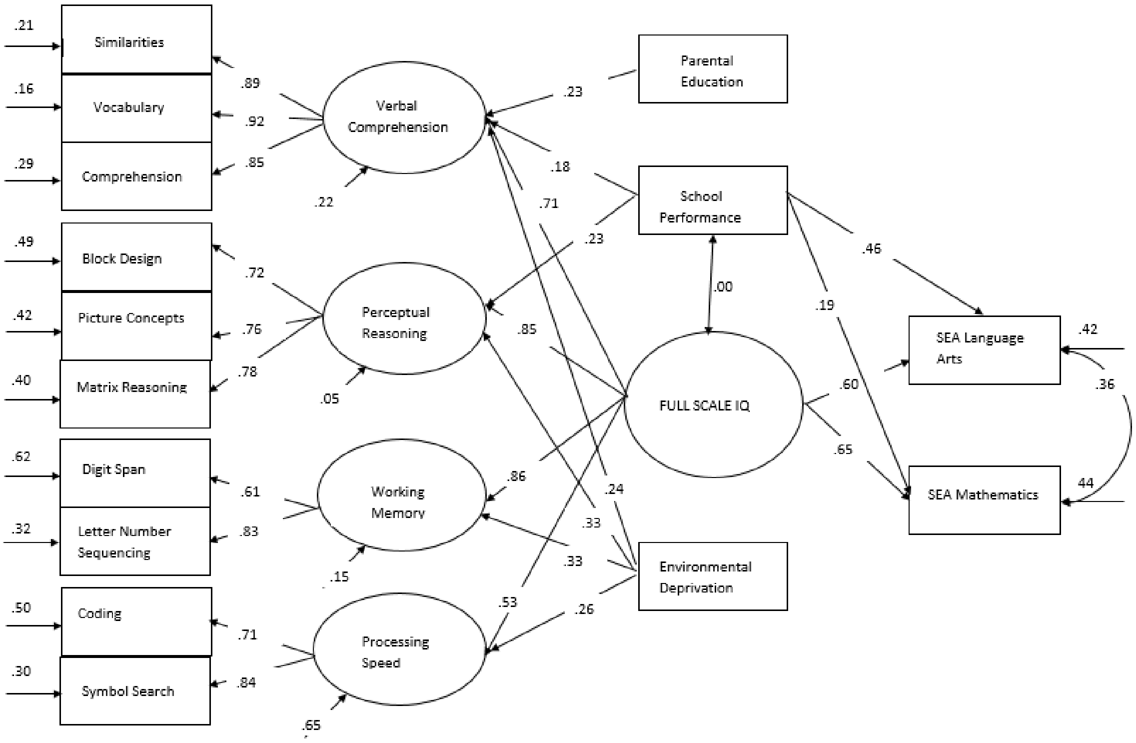


Figure 7.2. Standardized coefficients: Model 1

Table 7.3.

Unstandardized Coefficients, Standardized Coefficients, and Significance Levels for Unconstrained Model 1(Standard Errors in Parentheses; N = 201)

Parameter Estimate	*Unstandardized*	*Standardized*	*p*	*h²*
Measurement Model Estimates				
VCI by Similarities	1.00	.89	.00	.79
VCI by Vocabulary	.97 (.07)	.92	.00	.85
VCI by Comprehension	.80 (.07)	.85	.00	.74
PRI by Block Design	1.00	.72	Na	.56
PRI by Picture Concepts	1.19 (.13)	.76	.00	.58
PRI by Matrix Reasoning	1.16 (.11)	.78	.00	.61
WMI by Digit Span	1.00	.61	Na	.37
WMI by Letter Number Sequencing	1.25 (.13)	.83	.00	.69
PSI by Coding	1.00	.71	Na	.50
PSI by Symbol Search	1.44 (.21)	.84	.00	.71
FSIQ by VCI	1.00	.71	Na	.50
FSIQ by PRI	.83 (.09)	.85	.00	.72
FSIQ by WMI	.76 (.10)	.86	.00	.74
FSIQ by PSI	.46 (.09)	.53	.00	.28
Structural Model				
Language Arts on FSIQ	.35 (.04)	.60	.00	.36
Mathematics on FSIQ	.35 (.04)	.65	.00	.42
VCI on Environmental Deprivation	2.21 (.71)	.24	.00	.06
PRI on Environmental Deprivation	2.10 (.53)	.33	.00	.11
WMI on Environmental Deprivation	1.89 (.59)	.33	.00	.11
PSI on Environmental Deprivation	1.45 (.58)	.26	.01	.07
VCI on School Performance	.01 (.00)	.18	.02	.03
PRI on School Performance	.01 (.00)	.23	.00	.05
VCI on Parent Education	.17 (.07)	.23	.01	.05
Language Arts on School Performance	.01 (.00)	.46	.00	.21
Mathematics on School Performance	.00 (.00)	.19	.00	.04
FSIQ to Language Arts via School Performance	.00 (.00)	.00	.00	.00
FSIQ to Mathematics via School Performance	.00 (.00)	.00	.00	.00
Mathematics with Language Arts	.36	.36	.00	.13
School Performance to Language Arts via FSIQ	.00 (.00)	.00	.00	.00
School Performance to Mathematics via FSIQ	.00 (.00)	.00	.00	.00
Language Arts on School Type	.46 (.33)	.12	.16	.01
Language Arts on Parental Education	.20 (.26)	.08	.43	.01
Language Arts on Environmental Deprivation	.03 (.03)	.09	.39	.01

Mathematics on School Type	.61 (.33)	.18	.07	.03
Mathematics on Parental Education	.36 (.19)	.16	.06	,03
Mathematics on Environmental Deprivation	.04 (.03)	.16	.10	.03
FSIQ to Language Arts via School Type, Parental Education, Environmental Deprivation	.00 (.00)	.00	.00	.00
FSIQ to Mathematics via School Type, Parental Education, Environmental Deprivation	.00 (.00)	.00	.00	.00

Full Model 1: χ2 (82) =87.96, *p*=0.31; CFI=.99; TLI= .99; RMSEA=.02 (90% CI: .00; .05).
After 1st Constraint: χ2 (88) = 96.50, *p*=0.25; CFI=.99; TLI= .98; RMSEA=.02 (90% CI: .00; .05), χ2 diff (6) = 11.02, *p*=.09).
After 2ndConstraint: χ2 (90) = 98.60, *p*=0.25; CFI=.99; TLI= .98; RMSEA=.02 (90% CI: .00; .05), χ2 diff (8) = 22.17, *p*=<.001).
After 3rdConstraint: χ2 (91) =99.98, *p*=0.24; CFI=.99; TLI= .98; RMSEA=.02 (90% CI: .00; .05), χ2 diff (1) =6.60, *p*<.05).
After 4thConstraint: χ2 (93) =102.08, *p* =0.24; CFI=.99; TLI= .98; RMSEA=.02 (90% CI: .00; .05), χ2 diff (2) =12.10, *p*<.001).
After 5thConstraint: χ2 (97) =214.53, *p*<.001; CFI=.82; TLI= .80; RMSEA=.08 (90% CI: .06; .09), χ2 diff (4) =31.96, *p*<.001).
Note: Unstandardized coefficient = the change in Y latent or measured variable per change in X latent or measured variable (e.g. For 1point change in FSIQ, Language Arts score increases by .35 points).
Standardized coefficient = the number of SDs change Y per SD change in X.
h² = variance in Y explained by X. (e.g. For 1 SD change in VCI score, the Vocabulary score increases by .95 SD).
Na = Not available

Applying constraints to Model 2. The following model constraints were imposed on Model 2:

1. Constraint #1:
    a. The SES variables, Parental Education, Environmental Deprivation and School Type act as intervening variables between global IQ and Academic Achievement.
2. Constraint #2:
    a. School Performance predicts Academic Performance in Mathematics and Language Arts.

   b. School Performance acts as an intervening variable between global ability and academic achievement.
   c. Global Ability acts as an intervening variable between School Performance and Academic Achievement.
3. Constraint #3:
   a. VCI is predicted by Parental Education.
4. Constraint #4:
   a. FSIQ performance is predicted by School Performance and Environmental Deprivation.

Results

After the first constraint was applied to Model 2, the fit of the resulting model remained adequate though slightly reduced. The chi-square difference value after the first constraint was non-significant suggesting that the parameters removed in the first constraint added no significant value to the model. (See Table 7.4).

Table 7.4.

Goodness of Fit Indices for Full Model 2 and Constrained Models (N = 201)

Model	χ^2	*df*	*p*	χ^2/df	CFI	TLI	RMSEA (90% CI)
Full Model	84.41	85	.50	0.99	1.00	1.00	0.00 (.00 to .04)
Constraint #1	90.96	91	.48	1.00	1.00	1.00	0.00 (.00 to .04)
Constraint #2	92.97	93	.48	1.00	1.00	1.00	0.00 (.00 to .04)
Constraint #3	93.99	94	.48	1.00	1.00	1.00	0.00 (.00 to .04)
Constraint #4	211.68	96	.000	2.21	.83	.79	0.08 (.06 to .09)

After constraints 2 and 3, model fit reduced slightly but still remained adequate. The chi-square difference statistic was significant suggesting that School Performance as a predictor of academic achievement and an intervening variable between FSIQ and academic achievement adds significantly to the model. The inclusion of FSIQ as a mediator between School Performance and academic achievement was an important parameter. Parent Education accounted for 1% of the variance in VCI Performance (p = .29) and was therefore not included in the model. Constraint 4 resulted in a poorly fitting model ($\chi 2$ (96) =211.68, $p<.001$; CFI =.83; TLI = .79; RMSEA=.08 (90% CI: .06; .09); $\chi 2$ diff (3) =28.05; $p <.001$).

The final model consists of the direct hierarchical model (Watkins, 2006), with School Performance and Environmental Deprivation predicting FSIQ. Also FSIQ and predicts SEA Language Arts and Mathematics achievement, with School Performance as a mediating variable between FSIQ and academic achievement. School Performance also is a predictor of academic achievement in the model, with FSIQ as a mediator.

In the Structural Model, School Performance accounted for 4% of the variance in FSIQ performance and Environmental Deprivation explained 11% of the variance in FSIQ performance. Figure 7.3 provides a summary of the standardized coefficients for the full model. Unstandardized coefficients, standard error, standardized coefficients, and significance values for the Model 2 are presented in Table 7.5.

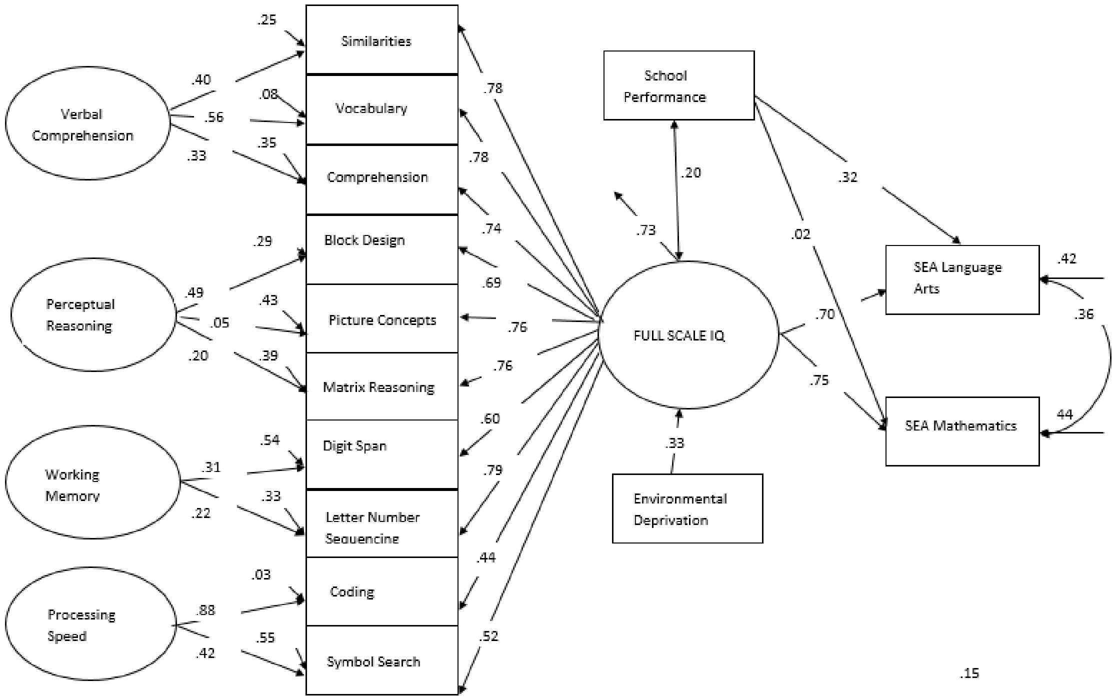


Figure 7.3. Standardized coefficients: Model 2

Table 7.5.

Unstandardized Coefficients, Standardized Coefficients, and Significance Levels for Unconstrained Model 2 (Standard Errors in Parentheses; N = 201)

Parameter Estimate	*Unstandardized*	*Standardized*	*p*	*h*²
Measurement Model Estimates				
FSIQ by Similarities	1.23 (.14)	.78	.00	.61
FSIQ by Vocabulary	1.18(.13)	.78	.00	.61
FSIQ by Comprehension	.99 (.12)	.74	.00	.55
FSIQ by Block Design	1.00	.69	.00	.48
FSIQ by Picture Concepts	1.24 (.14)	.76	.00	.58
FSIQ by Matrix Reasoning	1.18 (.11)	.76	.00	.58
FSIQ by Digit Span	.94 (.13)	.60	.00	.36
FSIQ by Letter Number Sequencing	1.17 (.13)	.79	.00	.62
FSIQ by Coding	.57 (.10)	.44	.00	.20
FSIQ by Symbol Search	.82 (.12)	.52	.00	.27
VCI by Similarities	.20 (.00)	.40	Na	.16
VCI by Vocabulary	.26 (.00)	.56	Na	.31
VCI by Comprehension	.14 (.00)	.33	Na	.11
PRI by Block Design	1.00 (.00)	.49	Na	.24

PRI by Picture Concepts	.12 (.16)	.05	.47	.00
PRI by Matrix Reasoning	.44 (.00)	.20	Na	.04
WMI by Digit Span	.09 (.00)	.31	Na	.10
WMI by Letter Number Sequencing	.06 (.00)	.22	Na	.05
PSI by Coding	.50 (.00)	.88	Na	.77
PSI by Symbol Search	.29 (.00)	.42	Na	.17
Structural Model				
Language Arts on FSIQ	.43 (.06)	.70	.00	.49
Mathematics on FSIQ	.43 (.05)	.75	.00	.56
FSIQ on Environmental Deprivation	2.00 (.53)	.33	.00	.11
FSIQ on School Performance	.01 (.00)	.20	.01	.04
VCI on Parental Education	.05 (.04)	.10	.26	.01
Language Arts on School Performance	.01 (.00)	.32	.00	.10
Mathematics on School Performance	.00 (.00)	.02	.55	.00
FSIQ to Language Arts via School Performance	.00 (.00)	.00	.00	.00
FSIQ to Mathematics via School Performance	.00 (.00)	.00	.00	.00
Mathematics with Language Arts	.36 (.08)	.36	.00	.13
School Performance to Language Arts via FSIQ	.00 (.00)	.14	.01	.02
School Performance to Mathematics via FSIQ	.00 (.00)	.15	.01	.02
Language Arts on School Type	.20 (.26)	.08	.43	.00
Language Arts on Parental Education	.01 (.03)	.02	.80	.00
Language Arts on Environmental Deprivation	.41 (.30)	.11	.18	.01
Mathematics on School Type	.36 (.19)	.16	.06	.03
Mathematics on Parental Education	.02 (.02)	.08	.30	.00
Mathematics on Environmental Deprivation	.24 (.30)	.07	.42	.00
FSIQ to Language Arts via School Type, Parental Education, Environmental Deprivation	.00 (.00)	.00	.00	.00
FSIQ to Mathematics via School Type, Parental Education, Environmental Deprivation	.00 (.00)	.00	.00	.00

Note: Full Model 1: χ2 (85) =84.41, *p*=0.50; CFI=1.00; TLI= 1.00; RMSEA=.00 (90% CI: .00; .04).

After 1st Constraint: χ2 (91) =90.96, *p*=0.48; CFI=1.00; TLI= 1.00; RMSEA=.00 (90% CI: .00; .04), χ2 diff (6) =7.95, *p*=.24).

After 2nd Constraint: χ2 (93) =92.97, *p*=0.48; CFI=1.00; TLI= 1.00; RMSEA=.00 (90% CI: .00; .04), χ2 diff (2) =14.54, *p*=<.001).

After 3rd Constraint: χ2 (94) = 93.99, *p*=0.48; CFI=1.00; TLI=1.00; RMSEA=.00 (90% CI: .00; .04), χ2 diff (1) = 1.31, *p*=0.25).

After 4thConstraint: χ2 (96) =211.68, *p*<.001; CFI=.83; TLI= .79; RMSEA=.08

(90% CI: .06; .09), $\chi 2$ diff (2) =28.05, *p*<.001).

Note: Unstandardized coefficient = the change in Y latent or measured variable per change in X latent or measured variable.
Standardized coefficient = the number of SDs change Y per SD change in X.
h^2 = variance in Y explained by X.
Na = Not available

Applying constraints to Model 3. The following constraints were imposed to Model 3:

1. Constraint #1:
    a. The SES variables, Parental Education, Environmental Deprivation and School Type act as intervening variables between global IQ and Academic Achievement.
2. Constraint #2:
    a. School Performance predicts Academic Performance in Mathematics and Language Arts.
    b. School Performance acts as an intervening variable between global ability and academic achievement.
    c. Global Ability acts as an intervening variable between School Performance and Academic Achievement
3. Constraint #3:
    a. VCI performance is measured by Parent Education
4. Constraint #4:
    a. VCI performance is predicted by School Performance
    b. NVCI performance is predicted by School Performance
    c. PRI performance is predicted by School Performance.
5. Constraint #5
    a. VCI performance is predicted by Environmental Deprivation.

b. NVCI performance is predicted by Environmental Deprivation.

c. PRI performance is predicted by Environmental Deprivation

d. PSI performance is predicted by Environmental Deprivation

Results

Model fit was slightly reduced after the first constraint was applied; however, the chi-square difference value was non-significant. Additionally, the model maintained good fit after the first constraint. Constraint 2 resulted in a significant reduction in fit which suggests that these parameters added important value to the model. Adding constraints 3 and 4 also significantly reduced the fit of the model; however, good fit was maintained until adding constraint 5, which resulted in a poorly fitting model ($\chi2$ (97) =213.90, *p*<.001; CFI=.83; TLI= .79; RMSEA=.08 (90% CI: .06; .09); $\chi2$ diff (4) = 32.08; *p* <.001; See Table 7.6).

Table 7.6.

Goodness of Fit Indices for Full Model 2 and Constrained Models (N = 201)

Model	χ^2	*df*	*p*	χ^2/df	CFI	TLI	RMSEA (90% CI)
Full Model	85.03	81	.36	1.05	.99	.99	0.02 (.00 to .04)
Constraint #1	93.50	87	.30	1.07	.99	.99	0.02 (.00 to .04)
Constraint #2	95.57	89	.30	1.07	.99	.99	0.02 (.00 to .04)
Constraint #3	96.94	90	.29	1.07	.99	.99	0.02 (.00 to .04)
Constraint #4	100.06	93	.29	1.08	.99	.99	0.02 (.00 to .04)
Constraint #5	213.90	97	.000	2.21	.83	.79	0.08 (.06 to .09)

The final model consists of the education /exposure measurement model. The structural model consists of Parental Education, School Performance and Environmental Deprivation predicting VCI, School Performance and Environmental Deprivation predicting NVCI and PRI, Environmental Deprivation predicting PSI, FSIQ and School Performance predicting SEA Language Arts and Mathematics achievement, and both School Performance and FSIQ acting as mediators between academic achievement and the predictors, global ability and school performance respectively. In the Structural Model, Parent Education explained 5% of the variance in VCI. School Performance accounted for 3%, 3% and 6% respectively of the variance in VCI, NVCI and PRI performance. Environmental Deprivation explained 6%, 12% 10% and 7% respectively of the variance in VCI, NVCI, PRI and PSI. The standardized coefficients for the full Model 3 are presented in Figure 7.4. Unstandardized coefficients, standard error, standardized coefficients, and significance values for the Model 2 are presented in Table 7.7.

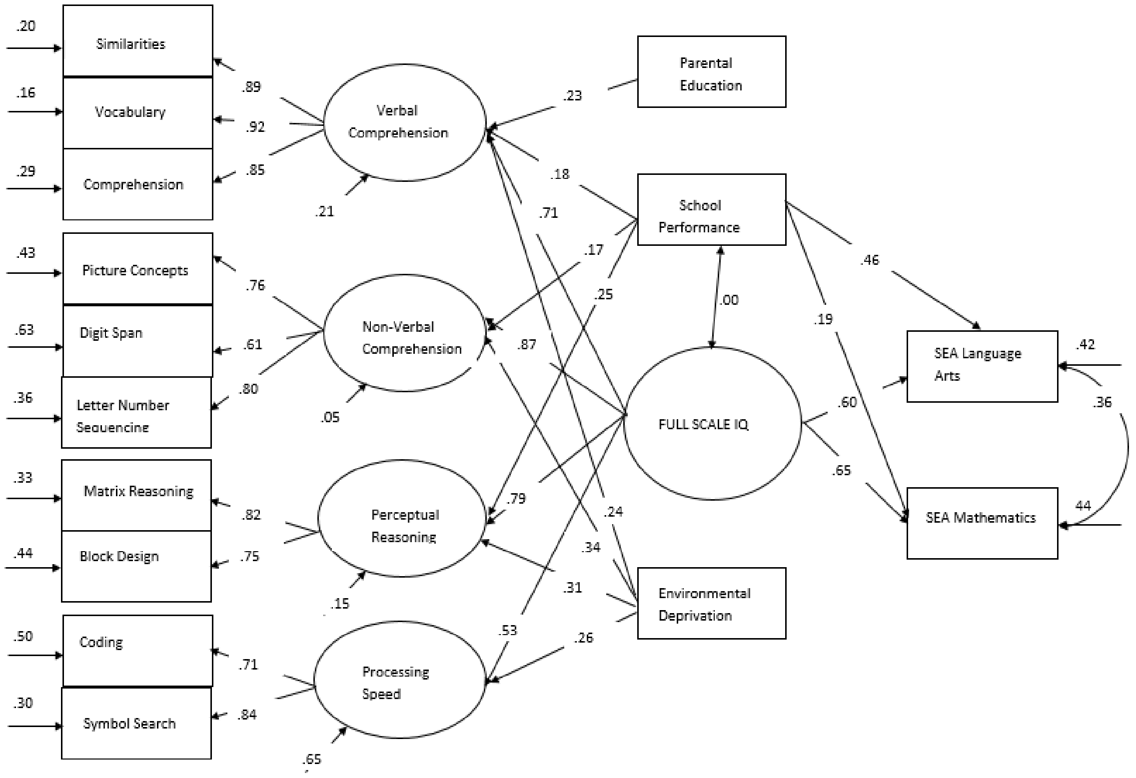


Figure 7.4. Standardized coefficients: Model 3

Table 7.7.

Unstandardized Coefficients, Standardized Coefficients, and Significance Levels for Unconstrained Model 3 (Standard Errors in Parentheses; N = 201)

Parameter Estimate	*Unstandardized*	*Standardized*	*p*	*h²*
Measurement Model Estimates				
VCI by Similarities	1.00	.89	Na	.79
VCI by Vocabulary	.96 (.07)	.92	.00	.85
VCI by Comprehension	.78 (.07)	.85	.00	.72
NVCI by Digit Span	1.00	.61	Na	.37
NVCI by Picture Concepts	1.28 (.16)	.76	.00	.58
NVCI by Letter Number Sequencing	1.22 (.13)	.80	.00	.64
PRI by Matrix Reasoning	1.00	.82	Na	.67
PRI by Block Design	.87 (.09)	.75	.00	.56
PSI by Coding	1.00	.71	Na	.50
PSI by Symbol Search	1.50 (.24)	.84	.00	.71
FSIQ by VCI	1.00	.71	Na	.50
FSIQ by NVCI	.76 (.10)	.87	.00	.71
FSIQ by PRI	.95 (.10)	.79	.00	.62
FSIQ by PSI	.44 (.09)	.53	.00	.28
Structural Model				
Language Arts on FSIQ	.35 (.04)	.60	.00	.36
Mathematics on FSIQ	.35 (.04)	.65	.00	.42
VCI on Environmental Deprivation	2.21 (.71)	.24	.00	.06
NVCI on Environmental Deprivation	1.94 (.53)	.34	.00	.12
PRI on Environmental Deprivation	2.39 (.64)	.31	.00	.10
PSI on Environmental Deprivation	1.45 (.60)	.26	.01	.07
VCI on School Performance	.01 (.00)	.18	.02	.03
NVCI on School Performance	.01 (.00)	.17	.05	.03
PRI on School Performance	.01 (.00)	.25	.00	.06
VCI on Parent Education	.16 (.07)	.23	.01	.05
Language Arts on School Performance	.01 (.00)	.46	.00	.21
Mathematics on School Performance	.00 (.00)	.19	.00	.04
FSIQ to Language Arts via School Performance	.00 (.00)	.00	.00	.00
FSIQ to Mathematics via School Performance	.00 (.00)	.00	.00	.00
Mathematics with Language Arts	.36 (.07)	.36	.00	.13
School Performance to Language Arts via FSIQ	.00 (.00)	.00	.00	.00
School Performance to Mathematics via FSIQ	.00 (.00)	.00	.00	.00
Language Arts on School Type	.20 (.26)	.08	.43	.01
Language Arts on Parental Education	.03 (.26)	.09	.39	.01

Language Arts on Environmental Deprivation	.46 (.33)	.12	.16	.01
Mathematics on School Type	.36 (.19)	.16	.06	.03
Mathematics on Parental Education	.04 (.03)	.16	.10	,03
Mathematics on Environmental Deprivation	.61 (.33)	.18	.07	.03
FSIQ to Language Arts via School Type, Parental Education, Environmental Deprivation	.00 (.00)	.00	.00	.00
FSIQ to Mathematics via School Type, Parental Education, Environmental Deprivation	.00 (.00)	.00	.00	.00

Note: Full Model 1: χ2 (82) =87.96, *p*=0.31; CFI=.99; TLI= .99; RMSEA=.02 (90% CI: .00; .05).

After 1st Constraint: χ2 (88) = 96.50, *p*=0.25; CFI=.99; TLI= .98; RMSEA=.02 (90% CI: .00; .05), χ2 diff (6) = 11.02, *p*=.09).

After 2nd Constraint: χ2 (90) = 98.60, *p*=0.25; CFI=.99; TLI= .98; RMSEA=.02 (90% CI: .00; .05), χ2 diff (8) = 22.17, *p*=<.001).

After 3rd Constraint: χ2 (91) = 99.98, *p*=0.24; CFI=.99; TLI= .98; RMSEA=.02 (90% CI: .00; .05), χ2 diff (1) = 6.60, *p*<.05).

After 4thConstraint: χ2 (93) =102.08, *p* =0.24; CFI=.99; TLI= .98; RMSEA=.02 (90% CI: .00; .05), χ2 diff (2) =12.10, *p*<.001).

After 5th Constraint: χ2 (97) = 214.53, *p*<.001; CFI=.82; TLI= .80; RMSEA=.08 (90% CI: .06; .09), χ2 diff (4) = 31.96, *p*<.001).

Discussion

The present study examined how observed scores on the WISC-IV (US) subtests are organized into factors, the extent to which environmental variables, parental education, school performance and environmental deprivation relate to these WISC-IV (US) factors and how these environmental variables and WISC-IV (US) factors predict academic achievement in a sample of T&T children. The results of the current study showed that the WISC-IV (US) data can be organized in three distinct ways. One framework organized the 10 core subtests into four oblique 1st order factors (VCI, PRI, WMI and PSI), and a higher order general factor (FSIQ) that is measured by the four 1st order factors. Also the model demonstrated that a T&T child's performance on the WISC-IV (US) can be linked to their parent's education levels, the overall academic performance of their school and also to whether or not they live in impoverished circumstances.

Another framework featured a higher order global ability factor (FSIQ) that was measured directly by the 10 core subtests and four orthogonal 1st order factors (VCI, PRI, WMI, PSI). In this model, performance on the FSIQ can be explained by their school's API and also their level of environmental deprivation. Variance in academic achievement was accounted for by both global ability and school performance with both variables acting as mediators in the predictive model.

In the third model, WISC-IV (US) observed scores were organized into 4 oblique 1st order factors measuring verbal crystallized ability, non-verbal crystallized ability, perceptual reasoning and processing speed, and a higher order global ability factor measured by the 1st order factors. WISC-IV (US) performance was demonstrated to be linked to parental education, school API and impoverishment.

Overall, the results highlighted the robustness of FSIQ as a predictor of academic achievement. The findings also support the hypothesis that global ability is linked to School Performance and Environmental Deprivation; and to a lesser extent, Parental Education. In the area of academic achievement, children with higher global ability scores performed better on the SEA Language Arts and Mathematics exams. The hypotheses that the relationship between School Performance and FSIQ is bidirectional, and that School Performance, Parental Education, School Type and Environmental Deprivation act as intervening variables between FSIQ and academic achievement partially found support in the data. The contributions of Parental Education, School Type and Environmental Deprivation to the predictive model were not found to be statistically significant. School Performance's contribution to the predictive model was statistically significant. Consistent with previous findings, global ability explained between 36% and

56% of the variance in academic achievement in both subject areas. Parental Education, School Type and Environmental Deprivation did not account for individual differences in academic achievement over and above global ability. As predicted, if global ability is controlled for, academic performance can be predicted by the academic performance rating of the school that the child attends.

Implications of the Findings

On the whole, the results of the present study suggest that interpretations of the WISC-IV (US) scores of the T&T sample must consider the potential influence of explanatory environmental variables. Indeed, consistent with the findings of the pilot studies that demonstrated poor model fit for children of lower SES, the environmental deprivation variable demonstrated the strongest association with all areas of WISC-IV (US) performance. Factors associated with impoverishment, such as poor health care and nutrition, lack of environmental stimulation, poor maternal care, low income and social isolation can severely delay intellectual development in children (; Rowe, Jacobson & van den Oord, 1999; Hart, Petrill, Deckard & Thompson, 2007), but can also negatively impact test taking skills and motivation (Kieffer & Goh, 1981). If replicated, these findings may help build a case against assessing children in very impoverished circumstances with the WISC-IV (US). Results also indicated that children who attended high performance schools performed better on measures of both verbal and non-verbal crystallized ability and perceptual reasoning ability. Crystallized and fluid reasoning abilities have been associated with formal education (Gottfredson & Saklofske, 2009; Lohman & Lakin, 2009); but it is not clear from the data which of the two plausible explanations - schooling improves intelligence or higher performing schools recruit ‘brighter’ students - best explains the data. The association between parental education and intelligence was evident only in the measure of verbal crystallized ability (VCI). The link between parental education and the fostering of academic knowledge in children was discussed in Chapter 6 as reflection of parental academic sophistication, strict attitudes towards education, and provision of a high standard of academic support at home (Alexander, Entwisle & Bedinger, 1994; Davis-Kean, 2005; Dubow, Boxer & Huesmann, 2009).

The implications of the findings are that in spite of evidence of structural equivalence, questions remain about construct equivalence between national samples. Although, the WISC-IV (US) purports to measure intelligence in US children, it may not measure, or may only partially measure intellectual ability in the T&T sample. Some of the variance in the WISC-IV (US) of T&T children may be accounted for by their familiarity with intelligence tests. Problems with construct equivalence generally mean that the T&T and US samples scores cannot be meaningfully compared.

The impact of construct in-equivalence on test interpretation is given considerable attention in test manuals and other literature (e.g. Weiss, Harris, Prifitera, Courville et al., 2006; Weiss, Prifitera, Holdnack, Saklofske et al., 2006). Practitioners are encouraged to use their clinical judgement to discuss how aspects of a child’s medical, family, social and educational history contribute to observed test scores (Prifitera, Saklofske & Weiss, 2005; Weiss, Harris, Prifitera, Courville et al., 2006; Gregoire, Georgas, Saklofske, van de Vijver et al., 2008). In T&T for example, it is common practice for psychologists to indicate that standard scores which are derived from comparisons with a US national sample, may actually underestimate the client’s true latent ability. These interpretations however are largely speculative, because as of yet, the field has not agreed on a way of quantifying the impact of cross-cultural variables on actual scores.

Study 8 will attempt to assess the validity of the three cross-cultural interpretive models by investigating the relationship between model-derived estimates of ability and academic achievement in the T&T sample.

Study 8

Using Factor Score Estimates to Assess the Predictive Validity of the Cross Cultural Interpretive Model

As was discussed earlier in the chapter, previous interpretive frameworks that have been offered for the WISC tests have been criticized for methodological problems, a lack of generalizability and poor predictive validity. The present study sought use global ability factor score estimates to investigate the relationship between WISC-IV (US) global ability and academic achievement. Analyses compared the fit of regression models using global ability factor score estimates that were generated from the aforementioned cross cultural interpretive models; as well as from the original WISC-IV (US) indirect hierarchical model (Keith, 2005). The validity of the models will be assessed on two indicators, absolute fit and relative fit.

Specifically, the study assessed the hypotheses that:

1. The Model 1 factor scores VCI, PRI, WMI and PSI measure global ability; and that the global ability factor score predicts performance in SEA Language Arts and Mathematics.
2. The Model 2 observed scores (Similarities, Vocabulary, Comprehension, Block Design, Picture Concepts, Matrix Reasoning, Digit Span, Letter Number Sequencing, Coding and Symbol Search measure global ability (FSIQ), and that the global ability factor score predicts performance in SEA, Language Arts and Mathematics.

3. The Model 3 factor scores VCI, NVCI, PRI and PSI measure global ability; and that the global ability factor score predicts performance in SEA Language Arts and Mathematics.
4. Wechsler (2003a) factors scores VCI, PRI, WMI and PSI measure global ability; and that the global ability factor score predicts performance in SEA Language Arts and Mathematics.

Analyses

Using CFA in MPlus6, the four alternative models were analysed and global ability factor scores were saved using the MPlus6 Regression Method (Muthén, & Muthén, (1998-2010). These factor scores were then entered as predictors of SEA Language Arts and SEA Mathematics in a regression model which was assessed using logistic regression in SPSS 21. Absolute fit was assessed using model fit indices of normed chi-square, AIC and BIC. Because the models are not nested models, they were compared on the basis of fit values.

Results

Tables 7.8 and 7.9 present the fit indices for the 4 alternative models tested. All with the exception of the Wechsler (2003b) based predictive model showed evidence of good fit to the data ($p<.001$). The Wechsler (2003b) based model failed to meet statistical significance on separate analyses ($p =.16$; $p =.05$). The predictive model that was derived from the Watkins (2006) based path model demonstrated better AIC and BIC values, but worse normed chi-square values than the other models. Overall, the factor scores derived from the alternative models were shown to predict SEA Language Arts and Mathematics performance in this sample of T&T children.

Table 7.8.

Goodness of Fit Indices for Evaluating Model Adequacy for Using Factor Score Estimates (Language Arts)

	χ^2	*df*	χ^2/df	*p*	AIC	BIC
Model 1	71.08	4	11.77	.000	482.67	509.10
Model 2	98.41	4	24.60	.000	448.41	474.83
Model 3	66.30	4	16.58	.000	487.45	513.88
Model 4	6.53	4	1.63	.16	538.09	564.52

Table 7.9.

Goodness of Fit Indices for Evaluating Model Adequacy for Using Factor Score Estimates (Mathematics)

	χ^2	*df*	χ^2/df	*p*	AIC	BIC
Model 1	93.67	4	23.42	.000	527.95	554.38
Model 2	129.08	4	32.27	.000	489.77	516.19
Model 3	88.21	4	22.05	.000	537.57	563.99
Model 4	9.44	4	2.36	.05	604.44	630.66

Discussion

Revisiting Factor Score Estimations in Test Score Interpretation

Factor score estimations can be a useful way of providing an objective summary of the impact of environmental variables on intellectual ability. In clinical practice, the adjusted score may provide the psychologist with a more accurate view of the child’s intellectual and academic potential. This approach may improve diagnoses as well as interventions. For example, a child who scores below 70 on a measure of IQ may be diagnosed with either Borderline Intellectual Functioning or an Intellectual Disability. Either diagnosis predicts difficulty with mainstream academic instruction, and in such cases, interventions may focus on teaching survival reading skills, writing and math, and fostering the development of daily living and vocational skills. If, however, factor score adjustments result in an increase in IQ score, this may affect diagnostic decision-making and educational planning. Intervention strategies are likely to focus on enhancing the child’s living conditions in order to promote academic success. Student support teams may also supplement academic interventions with social, financial and psychological support for the family, parental training, and provision of academic support at home.

The current study assessed whether factor score estimates derived from the hypothesized models adequately predicted academic achievement in a sample of T&T children. The study also determined if the proposed cross-cultural models were better than the Wechsler (2003a) model at predicting academic outcomes in this sample. To meet the aims of the study, global ability factor scores were estimated from each competing model. The relationship between the respective global ability scores and performance on SEA Language Arts and Mathematics were then examined. The results showed that alternative models provided some evidence of fit but the results were not very convincing. Interestingly however, the factor summary scores derived from the cross-cultural models predicted academic achievement better than the factor estimates of the original model.

Conceptually, the idea of a cross cultural model has demonstrated some promise. It reinforced the notion that in non-US populations, the Wechsler scoring structure is substantially improved by including important explanatory variables. While, both CFA and SEM analyses have supported the hypotheses imbedded in the model; the failure to demonstrate robustness in the predictive validity model, may reflect on the inability of the studies described thus far to identify other important environmental contributors to the nomological network, such as extra tutoring. Also, future research is needed to clarify the relationship between WISC-IV (US) performance and other possible explanatory variables such as monthly income, parental occupation, and use of Standard English in the home. Another hypothesized correlate of WISC-IV performance, US Exposure should also be re-examined. The failure of the US Exposure hypothesis to find support in this study was perhaps related to methodological rather than conceptual limitations. However, the variables hypothesized to covary with US exposure, parent education and income, were demonstrated to add significantly to the overall model.

Alternatively, the use of scores from a standardized academic achievement test may result in better fit than tests like the SEA by minimizing the impact of confounding variables (test taking speed and rote memorization) which may weaken the relationship between global ability and academic achievement.

The results of the current study supported the findings of the studies described earlier in this thesis. The WISC-IV (US) indirect hierarchical model was not sufficient to explain the performance of the sample of T&T children. Instead models featuring Parental Education, School Performance and Environmental Deprivation as predictors of WISC-IV (US) performance; and both global ability and School Performance as predictors of academic achievement provided adequate fit to the data. Although the present study produced evidence of the validity of a cross-cultural interpretive model of the WISC-IV (US) in a T&T sample, it also demonstrated that there are problems within the models that limit their generalizability and clinical applicability. For example, calculating factor scores and imputing them into the model did not result in convincing model fit. The idea of a cross-cultural interpretive paradigm that features environmental variables as part of its extended network holds considerable potential, but the ability of the model to influence clinical practice cannot be determined without more research.

CHAPTER 8

CONCLUSION

American Educational Research Association (AERA; 1999) guidelines stipulate that imported intelligence tests must be assessed for validity before they are adopted for use. This is because the skills, knowledge and abilities that test developers measure to predict success within one context, may not be relevant in another. This recommendation is consistent with a relativistic view of intelligence as a culturally specific construct (see Sternberg, 2004; Sternberg & Grigorenko, 2004). Such perspectives acknowledge that while the mechanisms underlying intelligence may be universal, the way intelligence is defined cross-culturally reflects the unique demands of each particular cultural environment (Sternberg, 2004). Investigations of a test's cross-cultural validity must therefore determine if the instrument measures abilities equally; and also, how well the abilities predict important life outcomes across cultures.

The purpose of this thesis was to assess the validity of the US version of the Wechsler Intelligence Scale for Children 4th Edition WISC-IV (US) (Wechsler, 2003a) for use in Trinidad and Tobago (T&T). The WISC-IV (US) has been used by T&T psychologists since 2003 to estimate the intelligence of 6 to 16-year-old children. The literature has provided extensive evidence that the WISC-IV is valid for use both in the US and in other countries. Indeed, the Wechsler (2003a) four 1st order factor scoring structure, consisting of verbal comprehension, perceptual reasoning, working memory and processing speed ability has found support in its standardization study as well as in external research (e.g. Wechsler, 2003a; Flanagan & Kaufman, 2004; Keith 2005; Keith, Fine, Taub, Reynolds & Kranzler, 2006; Watkins, Wilson, Kotz, Carbone & Babula,2006). Evidence consistent with a model featuring the four factor structure and an overarching higher order general intelligence factor was also provided in additional studies (e.g. Keith, 2005; Watkins, 2010). Furthermore, studies provided empirical evidence of the validity of the WISC-IV (US) as a predictor of academic achievement (Wechsler, 2003b; Freberg, Vandiver, Watkins & Canivez, 2008). By exploring the data of 11 to 12-year-old T&T children, the thesis attempted to determine whether current results would corroborate prior evidence of internal and external validity. The information was expected to contribute to the development of an interpretive model of the WISC-IV (US) for T&T children.

Chapters 1 and 2 presented the background and rationale of the present thesis. Chapter 1 provided a review of the literature on intelligence – its historical foundations and current trends in intelligence theory and measurement. The chapter explored the contributions of twin, adoption, intervention and correlational studies to our understanding of the causes of variability in intelligence. Specifically, the chapter discussed how genes and environmental factors such as parental education, income, environmental enrichment, home stability and formal education fit into the broad network of intelligence. The chapter also explored the life outcomes of intelligence such as health, criminality and occupational success, but paid specific attention to academic achievement which was the main outcome variable in the interpretive model. The chapter discussed how traditional and contemporary conceptualizations of intelligence have influenced the design of modern day tests. The chapter demonstrated that in spite of empirical support for current models, the actual definition of intelligence itself remains a source of controversy. Questions about the nature of intelligence persist as does the debate over what intelligence tests really measure. Issues about the cultural specificity of intelligence and the inability of test developers to construct a universal measure of intelligence were discussed. The chapter argued that because intelligence is such a complex construct, validity studies should broaden their scope from the internal structure, to the vast array of antecedent and outcome variables that form part of the extended intelligence network.

Chapter 2 described the WISC-IV (US), its uses, development, content and psychometric properties. Some of the potential consequences of using the test without first assessing its validity in T&T were discussed, and an argument was put forward that the test may be more suitable for T&T children who have access to US culture. The chapter also presented the findings of pilot studies that were conducted by the author of this thesis, and that used CFAs to examine the structural validity of the WISC-IV (US) in a referred sample of T&T students. The results demonstrated fairly good fit for the WISC-IV (US) four 1st order factor model. Multi-sample analyses however demonstrated good fit for a high income sub-sample and poor fit for a low income sub-sample. Furthermore, the difference in fit between the two groups was attributed to socioeconomic status, income, and also quality of education which were argued to impact access to US culture through television, internet and travel. In light of the findings of the pilot studies, further exploration was considered necessary to clarify the relationship between WISC-IV (US) structure and environmental variables such as US exposure, parental education, income and quality of schooling.

Investigating the Performance of the T&T Sample

Study 1 that was described in Chapter 3 was conducted to describe the performance of the T&T sample on the WISC-IV (US), by examining frequency distributions for normalcy, comparing mean scores with the US standardization sample, and comparing within-sample subtest and index scores. Both 11 and 12-year-old T&T children performed consistently lower than their US counterparts on all WISC-IV (US) subtests. The differences between samples on the Coding subtest either equalled or exceeded 1standard deviation. In the absence of raw US standardization data, between group differences could not be tested for statistical significance. This fact, along with an absence of evidence of construct equivalence, meant that no further comparisons between groups could be made (van de Vijver & Tanzer, 2004).

The within sample comparisons produced interesting results, the most striking of which was the sample's very low performance on a measure of processing speed. Results were thought to reflect cultural differences in the way time is perceived. The notion is supported by studies that demonstrate slower work and walking pace in hotter, less developed, and less populated countries (Bartlett, 1984; Levine & Norenzayan, 1999). The low processing speed scores otherwise may have been explained by the disproportionately high number of children with learning disabilities (LD) in the sample (25.1%). Such instances of low processing speed relative to other subtests are not uncommon in studies with referred samples (i.e. children with ADHD, Autism or TBI) (e.g. Mayes & Calhoun, 2005 & 2006, Solanto et al., 2008; Goldstein et al., 2008; Allen, Thaler, Donohue & Mayfield, 2010). A follow up analysis was conducted to determine, if after removing LD cases from the dataset, whether the significant difference between processing speed scores will be obtained. Repeated Measures ANOVA demonstrated no major changes in mean score differences between the sample's score on the processing speed measure and the other subtests. The results did not support the learning disability hypothesis.

The sample's low performance in Comprehension was also noted. The findings were argued to highlight cultural differences between the US and T&T, as well as, T&T society's norms about how children communicate with adults. A surprising result was the sample's relative strength on the Picture Concepts subtest. This test was predicted to be one of the more challenging tests for T&T children because of its US content. However, the sample’s performance on this task exceeded that of the other subtests. This result was attributed to the larger number of high SES participants in the study. It was argued that high SES children are more likely to have access to US culture and therefore produce more correct responses on the test. Comparisons between groups with varying degrees of US exposure provided some support for this hypothesis, but there was evidence of a possible non-linear relationship between US exposure and Picture Concepts performance, in that children with ‘very much’ exposure to US culture performed no better on this test than children with some or moderate levels of exposure.

Investigating Internal Validity

Study 2 of Chapter 3 examined the inter-correlation matrix to determine how subtests were organized. The results supported a grouping together of the subtests into three and also four 1st order factors. Furthermore, the subtests, Picture Concepts, Letter Number Sequencing, Matrix Reasoning and the three verbal comprehension subtests (Similarities, Vocabulary, Comprehension) correlated very highly with each other. The clustering together of these six subtests were argued to reflect an underlying education/exposure cluster.

This hypothesis found partial support in Study 3 through the 1st order and higher order Exploratory Factor Analyses (EFAs). Oblique rotation and orthogonal rotation EFAs techniques were used to investigate the 1st order structure and the higher order structure of the test. The first order EFA demonstrated a cross-loading for Picture Concepts on factors associated with perceptual reasoning and working memory. The highest loading was on the factor related to working memory. Similar findings were obtained for the higher order EFA which extracted a general intelligence factor that was directly measured by the 10 core subtests and a similar grouping of subtests into uncorrelated 1st order factors. The factors were subsequently conceptualized as follows: Verbal crystallized (Similarities, Vocabulary and Comprehension); Non-verbal crystallized (Picture Concepts, Letter Number Sequencing and Digit Span); Perceptual Reasoning (Matrix Reasoning and Block Design) and Processing Speed (Coding and Symbol Search). Implied in the terms 'verbal and non-verbal crystallized' is the hypothesis that education and exposure explains a significant portion of variance on both factors.

This hypothesis was tested in Study 4, reported in Chapter3, along with alternative models including the 1st order scoring structure and both the Keith (2005) indirect and Watkins (2006) direct hierarchical versions of the Wechsler scoring structure. CFAs supported both the 1st order oblique factor model and the direct and indirect hierarchical models (Keith, 2005; Watkins, 2006). The Watkins (2006) direct hierarchical model demonstrated the best fit of all the tested models, but an interesting finding was that the indirect and direct versions of the education/exposure model demonstrated the 2nd and 3rd best fit respectively. Fit indices of the five well-fitting models were very similar, which complicated selection of a measurement model for the larger interpretive framework, especially in the absence of statistical significance estimates for non-nested model comparisons (Hooper, Coughlan & Mullen, 2008). Eventually, the 3 alternative models with the highest model fit indices were selected for further analysis. They included the indirect hierarchical scoring structure model (Keith, 2005); the direct hierarchical model proposed by Watkins et al. (2006); and the indirect education/exposure hierarchical model (developed as part of this thesis).

Environmental Variables and WISC-IV (US) Performance

Study 5 of Chapter 6 investigated the relationship between the environmental variables (parental education, school performance, school type and US exposure) and the five WISC-IV (US) ability composites (VCI, PRI, WMI, PSI and FSIQ). The study further explored the hypothesis that the 25.1% presence of children with learning disabilities may explain the sample's overall weak performance on the processing speed subtest. This was done by examining the inter-correlation matrix to determine if Learning Disability (LD) and Processing Speed Index (PSI) performance were related. The inter-correlation matrices revealed low and nonsignificant correlations between LD and all PSI measures. A significant correlation between Gender and PSI was observed. Hierarchical multiple regression analyses uncovered a significant effect for Gender on PSI performance. In this sample, girls outperformed boys on the measure of processing speed, but the effect was negligible (*d* = .02), and disappeared once other environmental variables were added to the predictive model for PSI. The resulting model highlighted US exposure as the sole predictor of PSI performance in the sample. US exposure also explained the largest percentage of variance in the other WISC-IV (US) composites. Post-hoc analyses demonstrated that on all of the ability composites, there was no difference in WISC-IV (US) performance among children who had 'some', 'moderate' or 'very much' exposure to US culture. Instead, the results demonstrated that, compared with the rest of the sample, children who had very little or no US exposure performed poorest on the WISC-IV (US). This subsample also had less well educated parents and additionally, fewer children in this group attended private school. Based on the questionnaire, these children may also lack access to television or internet. In this light, the US exposure questionnaire was thought to be an indicator of poverty and was renamed the Environmental Deprivation variable.

Study 5 demonstrated that children of highly educated parents performed relatively better on the verbal comprehension test. Children who attended schools with high scores on the Academic Performance Index (API) also performed better on the verbal comprehension, perceptual reasoning and general intelligence composites. These results were consistent with studies that found statistically significant associations between intellectual ability and parental education or schooling (Cahan & Cohen, 1989; Weiss, Harris, Prifitera, Courville et al., 2006; Gottfredson & Saklofske, 2009; Lohman & Lakin, 2009; Rinderman, Flores-Mendoza & Mansur-Alves, 2010). Based on the findings, School Performance, Parental Education and Environmental Deprivation were included in the cross-cultural interpretive model as antecedents of WISC-IV (US) performance.

Environmental Variables, Global Ability and Academic Performance

Study 6 of Chapter 6 investigated the relationship between academic achievement measured by performance on the Language Arts and Mathematics, SEA national exams, and global ability. The study also examined the extent to which school performance, school type, parental education and US exposure contributed to the predictive model. Logistic regression analyses investigated the relationship between academic performance and the predictors, global ability, school performance and school type, after controlling for parental education and environmental deprivation. Results of the first regression showed that children with higher global IQ scores performed better on the Mathematics and Language Arts national exams. After controlling for IQ, students attending higher API schools performed better on both national exams. Notably, school performance could not explain why some children attained the highest scores in Mathematics over and above other children of similar IQ. It was argued that this advantage may have been gained through extra lessons, a variable that should be investigated in future studies.

Assessing the Cross-Cultural Interpretive Model

Study 7 of Chapter 7 was conducted to assess the validity of three alternative cross-cultural interpretive models of WISC-IV (US) measured intelligence in T&T children. The interpretive hypotheses were presented in the form of a path model with a measurement model surrounded by a network of antecedent and outcome variables. The interpretive framework to be assessed consisted of the Wechsler (2003a) hierarchical scoring structure as the measurement model. The second model featured the Watkins (2006) direct model as the central construct. Model 3 used the education/exposure model. All measurement models were regressed on parental education, environmental deprivation and school performance. The global ability factor and school performance were specified as predictors of SEA Mathematics and SEA Language Arts. The environmental variables, school type, parental education, environmental deprivation and school performance were also included in the model as mediators between global ability and academic achievement. Global ability was also featured as a mediator between school performance and academic achievement.

The results of the study supported the notion that adding antecedent variables to the WISC-IV (US) intelligence network substantially enhances model fit. All alternative models demonstrated good fit to the data, with the Watkins (2006) based model demonstrating best fit. Overall, the findings implied that children whose parents were highly educated were more likely to perform better on the measures of verbal crystallized ability. Also children who attended high performance schools performed better on verbal crystallized ability, nonverbal crystallized ability, perceptual reasoning ability, and global ability. Environmental deprivation was shown to explain the largest portion of variance in WISC-IV (US) performance.

The results also demonstrated that global ability predicted academic achievement in the sample. Furthermore, school performance was found to explain a significant portion of variance in academic achievement. Additionally, school performance was demonstrated to act as a mediator between global ability and academic achievement. Global ability was also found to be an intervening variable in the school performance/academic achievement relationship. The role of school performance as a mediator can be explained in one of two ways. Either the difference in performance between two children of similar IQ can be explained by the quality of school they attend, or high performance schools actively recruit children of either higher IQ or higher prior achievement to enhance their performance ratings. If the 2nd explanation is true, then the API scores are not true indicators of school ‘quality’ because, in this case, differences in academic achievement are not the result of ‘school added value’.

Factor score estimation. As a first step towards testing the validity of the interpretive models, Study 8 of Chapter 7 estimated factor scores by using the original indirect model, and the three alternative path models. The factor score estimates were then imputed into four separate logistic regression models in which global ability was hypothesized to predict SEA Language Arts and SEA Mathematics. The fit of four alternative regression models were assessed and compared. Results of Study 8 showed that the regression model that was derived from the Wechsler (2003b) original scoring structure did not fit well with the data. The three alternative models demonstrated some evidence of good fit; however, it was difficult to determine, based on the model fit indices which of the alternative models fit best.

Implications of the Findings

Enhancing WISC-IV (US) score interpretations. WISC-IV(US) test score interpretation generally starts with a description of the client's general intelligence (FSIQ), followed by an analysis of performance on the specific cognitive ability domains; and at times, an examination of individual subtests (Wechsler, 2003a; Weiss, Prifitera, Holdnack, Saklofske et al., 2006). The FSIQ, which is calculated from the four ability indices (Wechsler, 2003b), is often considered less useful than the specific abilities (Weiss, Prifitera, Holdnack, Saklofske et al., 2006), especially in educational and clinical settings where detailed profile analysis is important for decision making and educational planning. As such, a clear understanding of the abilities, knowledge and skills associated with the various WISC-IV (US) domains and subtests (see Chapter 2) is essential for a thorough and comprehensive interpretation of scores. As an example, a low score on the verbal comprehension index may indicate overall poor verbal skills, but may also indicate a weakness in verbal fluid reasoning ability, expressive language or formal education. A high score on the Block Design task may reflect good visual spatial reasoning ability, as well as strong visual acuity, visual processing and visual-motor coordination. The aforementioned interpretations adhere to the basic guidelines detailed in the Technical and Interpretive Manual (Wechsler, 2003b); however, this thesis may have uncovered potentially useful alternative frameworks to the traditional indirect scoring structure that are worth further exploration.

Watkins (2006) direct hierarchical measurement model. The most obvious is the Watkins (2006) measurement model which was replicated in Study 3 of Chapter 5. This model gives the global ability factor priority over the specific abilities. A problem for practitioners is likely to emerge when attempting to quantify this direct higher order factor. This is because there is no established scoring algorithm in the WISC-IV (US) for a model where the global factor is measured directly from the core subtests. In fact, in spite of exposing the conceptual limitations of the original scoring structure, Watkins (2010) recommends using the FSIQ score as an estimate of global ability.

Perhaps the main argument behind the model is that after accounting for a general factor, the specific abilities explain very little of the common and total variance and thus provide a less meaningful measure of ability. In such a case, the FSIQ should be used as the sole estimate of intellectual ability. As a corroboration of this interpretive approach, the inter-correlation analysis of Chapter 6 revealed stronger correlations between academic achievement (the SEA subjects) and FSIQ compared with the other ability composites (VCI, PRI, WMI, and PSI). Similar findings were presented by Weiss and colleagues (2006) who found correlations of .87 to .89 between FSIQ and academic achievement regardless of the size of the discrepancies among the individual specific ability composites. This information provides a clear rationale for using only the FSIQ in the achievement/ability discrepancy analyses that are used in T&T for diagnosing Learning Disabilities.

The utility of this approach can be seen in actual practice where large achievement/ability differences can result in the diagnosis of a specific learning disability (Prifitera, Saklofske, Weiss & Rolfhus, 2005). The use of specific ability composite scores such as GAI (which is a combination of the VCI and PRI) or VCI scores as a substitute for FSIQ may result in smaller achievement/ability difference values (Prifitera, Saklofske, Weiss & Rolfhus, 2005; O’Donnell, 2009). This is especially true in cases where children with learning disabilities also demonstrate lower scores in verbal comprehension or perceptual reasoning ability.

Education/exposure measurement model. Perhaps the closest to a cross-cultural measurement model was revealed by the CFAs in Study 4 of Chapter 5. This model is identified by the presence of a verbal crystallized and a non-verbal crystallized factor among the four 1st order abilities. In the T&T sample, a relative weakness in 'non-verbal crystallized ability' may reflect a lack of environmental or academic exposure. However, such an explanation can be plausible only if accompanied by relative weaknesses on the verbal measures of crystallized ability as well as actual evidence of under-stimulation or poor education. Replication studies along with studies which examine the correlations between the verbal crystallized and non-crystallized factors in other similar samples will provide further light in this area.

Cross-cultural interpretive model. The most important models to emerge from this thesis are the cross-cultural interpretive path models. Some of the broad conclusions of these models are not new. For example, the results of Chapter 6 support previous findings that schooling, parental education and impoverishment are linked to IQ test performance (Cahan & Cohen, 1989; Weiss, Harris, Prifitera, Courville et al., 2006; Gottfredson & Saklofske, 2009; Lohman & Lakin, 2009; Rinderman, Flores-Mendoza & Mansur-Alves, 2010; Nisbett et al. 2012). In the US standardization population, these antecedents may not be an explicit part of the original model because the standardization process minimizes the contribution of culture by stratifying the sample on important socio-cultural demographics such as SES, ethnicity and gender (Weiss, Harris, Prifitera, Courville, et al., 2006; Gregoire, Georgas, Sakloffske, van de Vijver, et al., 2008). In this thesis, regressing the ability factors on to the observed environmental variables provided useful information about their statistical contribution to the broad model. This means that the interpretations can potentially ascertain in numerical terms the impact of external variables on the model.

Factor score estimation and cross-cultural means comparisons. Means comparisons are an important aspect of cross-cultural research in intelligence. It allows the field to identify how samples differ on key predictors of life success, thereby providing a greater understanding of cross-cultural differences in academic achievement, economic growth, health behaviour, crime and other major life outcomes. Additionally, between-group differences in means have prompted researchers to make observations and test hypotheses about the cross cultural variables that may explain differences in performance. Therefore, means comparisons are important to the development of a comprehensive cross-cultural model of intelligence. In this regard, the estimation of factor scores from cross cultural interpretive path models can potentially have important implications for such research.

If a construct is determined to be variant across two cultures, means comparisons may become invalid (van de Vijver & Tanzer, 2004). It follows that a test that does not measure the same construct in two nations or two ethnic groups will produce scores that cannot be meaningfully compared. Interestingly, the popular approach of test adaptation which has been used to address this problem, may improve construct validity but may also decrease cross-cultural equivalence (see van de Vijver & Tanzer, 2004). The approach of adjusting scores to account for environmental variables has been used, but has also been criticized for methodological flaws. Rather than abandoning this strategy, work should focus on improving it. Study 7 showed that factor scores derived from the regression weights of the cross-cultural models predicted academic achievement better than those derived from the original model. In spite of not convincingly achieving good fit, factor score adjustment appears to be a promising way of controlling for environmental variables when making means comparisons. Essentially, corrections can potentially result in a better and more reliable estimate of IQ and allow for more meaningful cross-cultural comparisons of means. The usefulness of this approach can be assessed through replication studies.

Factor score estimation and adjusted IQ scores. Often, guidelines for the diagnosis of various intellectual, learning, neuropsychological and language disorders specify a cut-off IQ score. In Chapter 2, it was suggested that invalid estimations of IQ can negatively impact clinical decision making in such cases. If replications support the findings that test validity is influenced by environmental variables, then score adjustments may improve diagnostic accuracy. Adjusted scores perhaps can be used as a loose indicator of how specified environmental variables may have influenced IQ test performance.

It is not suggested that adjusted scores replace actual IQ scores; rather, adjusted scores can be used as additional information and as an advanced method of case conceptualization. Psychological reports can communicate with other professionals how adjusted scores were obtained, the purpose of their use, and what the scores imply for intervention planning. So if by using a cut off score of 70, a referred student with an IQ of 69 meets the diagnostic criteria for Borderline Intellectual functioning, then if the IQ score adjustment results in a score higher than the cut off (e.g. 78), then it can be argued that, considering that the child’s current environmental conditions and academic instruction are less than optimal, the child’s academic learning potential is likely to be closer to the Below Average level. As a consequence, interventions can be tailored to stimulate intellectual development and thereby build the foundations for academic learning.

Strengths and Limitations of the Thesis

This thesis has made an important contribution to the literature on cross-validity in intelligence testing, by gathering evidence to support the development of a cross cultural interpretive model of WISC-IV measured intelligence in T&T children. This is a novel endeavour, since there are few studies that have examined the cross-cultural validity of intelligence tests in the Caribbean, and to my knowledge none have applied path model analysis to assess validity of intelligence tests in this population. Another strength of this thesis is the use of the structural equation modelling to assess validity. This is contrasted with traditional EFA and CFA approaches which are relevant only to measurement models. The use of structural equation modelling is an approach that can have encouraging implications for future research and clinical practice in the Caribbean.

Lamentably the findings of the thesis are not generalizable to the broader population. This was an unavoidable consequence of recruiting participants in a society that is unfamiliar with psychological research and wary of divulging personal data. Recruitment of participants was a slow and painstaking process that resulted in a convenience sample that did not represent the T&T population in ethnicity, SES and perhaps in the prevalence of learning disabilities. This undoubtedly made the study susceptible to the effects of group related nuisance variables. Acknowledging the problems associated with test bias, caution is taken before making sweeping statements about the findings in wider populations when they may be relevant to this group only. Although the sample exceeded the required N for SEM, longer term data collection may have ensured a more representative sample. Unfortunately, this was not possible as a result of time and financial constraints and difficulty with recruitment.

Other limitations may have impacted negatively on the generalizability of the findings. Use of more than one administrator was necessary but could have negatively impacted reliability. Inter-rater reliability estimates were high, which is a good indicator that scoring procedures were adhered to; however, it was not possible to determine without recording or observing sessions, whether standards of administration were consistently followed.

Another limitation of the thesis was its inability to collect data about income due to problems respondents seem to have about disclosing such information. The majority of participants left this category open and so a proxy variable for income, School Type, was used. It was not possible to determine how well the School Type variable correlated with income, although traditionally, only the wealthy in T&T society could afford to send their children to private schools. In recent times, however, increasing numbers of children of middle class families have been enrolling at expensive private schools. In this light, it is reasonable to argue that the correlation between school type and income has diminished over the years. There was no evidence of extreme collinearity between School Type and School Performance, however, the moderate correlation between the two variables (r = .43) may have been a problem. Apart from denominational public schools, private schools produce some of the best academic scores in T&T. It is therefore possible that the presence of school performance in the model muted the contribution of school type. A solution to this problem would be to seek other methods of collecting information on income. If disclosure is a real problem, a more reliable proxy such as occupation may be useful. Salary ranges can be used to delineate these income categories. In spite of the inability of the study to identify an effect for school type or income, they should not be removed from the model. A suitable approach may be to group highly correlated environmental variables into factors.

Future Directions

Results suggests that the WISC-IV (US) performance of T&T children who are similar to those of the sample that participated in this study is explained by their parents' education levels, the quality of school they attend, and perhaps the amount of stimulation their environment provides. Results also showed that performance on the SEA is predicted by intellectual ability and the performance ranking of their schools. Sampling limitations imply that these findings cannot be generalized to the broader population of T&T children, however, it is not unreasonable to predict that similar findings will be obtained if the study were redone. To test this hypothesis, replication studies should be carried out to determine the reliability of these findings and also the generalizability of the model to other populations. Future studies should employ stratified sampling techniques for a more representative T&T sample. Also, studies in other Caribbean countries as well as populations like the UK and the US where there is a large multi-ethnic population may be very valuable for assessing the effects of cross-cultural variables. Specifically, studies to investigate the validity of the WISC-IV (Wechsler, 2003) and soon the WISC-V (Wechsler, 2014) should address the following aims:

1. To examine the factor structure of the 15 WISC-IV and 16 WISC-V core and supplemental tests in a representative sample of children.
2. To make comparisons between the current WISC-IV and WISC-V model and alternative models to determine best fit.
3. To determine the impact of carefully selected environmental variables on WISC-IV and WISC-V performance.
4. To use differential item functioning analyses for item-level examinations to identify which test items, if any, negatively impact model fit.
5. To determine if revisions to questionable items will improve model fit, by administering alternative versions of the test to two similar samples. Comparisons of the model fit indices of the two samples will determine which of the alternate versions result in better fit.
6. To determine if removal or modification of questionable items significantly improve model fit over and above the original version.

7. To investigate how WISC-IV and WISC-V performance and performance on classroom and national exams are related.
8. To use longitudinal studies to investigate the relationship between WISC-IV and WISC-V performance and academic performance from primary to tertiary levels.

Another approach to replication may be to test the cross-cultural model using other measures of intelligence. Further research also is needed to strengthen the paradigm by clarifying the contribution of additional explanatory variables. Other variables that remained unexplored at the end of these studies: the US exposure variable, extra tutoring, parental attitudes towards education and pace of life should also be included in future research.

If replicated, the cross-cultural interpretive model might underscore the need to develop quantitative measures of key background variables for use in clinical practice. These may include established and validated measures of SES, school performance ratings and even US acculturation. Beyond academic goal setting, the model may also influence implementation of social interventions. Support from visiting specialized instructors who can provide advice, guidance and training for teachers at low API schools may help improve academic performance standards. Training, counselling, guidance and financial assistance can also be offered to parents to help improve the provision of academic support at home.

Final Remarks

In spite of its limitations, this thesis has added evidence in support of the notion that the interpretation of performance of T&T children on WISC-IV (US) cannot be understood without considering the potential impact of culturally relevant environmental variables. One can argue that these results offer a strong rationale for the adaptation of the WISC-IV into a version that is compatible with the T&T context. This is certainly a reasonable option and one that has been adopted by a variety of countries (e.g. WISC-IV (Canadian) (Wechsler, 2004); WISC-IV (Spanish) (Wechsler, 2004); WISC-IV (UK) (Wechsler, 2004); WISC-IV (French) (Wechsler, 2005); HAWIK-IV (German) (Petermann & Petermann, 2010); and WISC-IV (Vietnam) (Dang et al., 2011)).

But do adaptations solve the problem of cross-cultural in-equivalence? How do test adaptations account for the vast number of sub-cultures that may co-exist within a dominant culture? It is difficult to conceive of an adaptation of WISC-IV items that will thoroughly address the problems of in-equivalence even within an apparently homogenous country. As an illustration one can consider the difference between a child who grows up in an isolated rural Greek community with little access to television, internet or the wider society and another Greek child who lives in an urban cosmopolitan environment and has access to all the modern conveniences. These children will gain from their own experiences specific practices, communication styles, dialects, and understandings of how the world works that underpin survival within their specific environmental context. A test developer therefore may find it very difficult to measure each child’s abilities without either constructing a cumbersome, item-heavy instrument, or conversely one that is unfairly weighted in one direction or another.

Once culture specific knowledge is measured outside of its cultural context, environmental variables are introduced that may diminish the test’s capacity to gauge what is of primary interest, the capability to “reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience” (Gottfredson, 1997, p.13).

The best solution may be to develop a non-context specific definition of intelligence; but in light of all that was said prior, this solution may also be the most impractical. The prospect that there are multiple intelligences to be separately defined and measured is one that seems to defy the laws of parsimony, however in reality, relying on a global model of intelligence may have also be a limiting factor for the field. Perhaps the historical controversies that have accompanied intelligence testing should be blamed on relying on too narrow definition of intelligence.

It is believed that a resolution to this dilemma will come from diligent continued attempts to study how key environmental variables, language, pace of life, SES, education systems, or societal norms determine how intellectual skills are manifested across cultures. For example, how language orthography is related to visual spatial reasoning and long term verbal memory; or how topography of the natural environment impact visual processing ability (see Roberson, Davidoff, Davies & Shapiro, 2004). Such investigations into how cultural variables account for differences in performance between groups will likely make an invaluable contribution to intelligence research and it is believed that the findings will provide the field with a broader understanding of the nature of intelligence.

References

American Educational Research Association (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Alexander, K. L., Entwisle, D. R., & Bedinger, S. D. (1994). When expectations work: race and socioeconomic differences in school performance. *Social Psychology Quarterly, 57*, 283 – 299.

Allen, D.N., Thaler, N.S. & Donohue, B. (2010). WISC–IV profiles in children with traumatic brain injury: similarities to and differences from the WISC–III. *Psychological Assessment, 22*(1), 57 – 64.

Alloway, T.P. & Alloway, R.G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20 – 29.

American Psychological Association (2000). *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (TR)*. Arlington, VA, American Psychological Association

American Psychological Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Arlington, VA, American Psychiatric Association.

Anderson, J.O., George, J. & Herbert, S. (200). Factors impacting on student learning: a preliminary look at the National Test of Trinidad and Tobago. *Caribbean Curriculum, 16*(2), 99 – 126.

Andersson, U. & Lyxell, B. (2008). Working memory deficit in children with mathematical difficulties: a general or specific deficit? *Journal of Experimental Child Psychology, 96*(3), 197 – 228.

Arinoldo, C.G. (1981). Black–White differences in the General Cognitive Index of the McCarthy Scales and in the Full Scale IQs of Wechsler's Scales. *Journal of Clinical Psychology, 37*(3), 630 – 638.

Asbury, K., Wachs, T. D. & Plomin, R. (2005). Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. *Intelligence, 33*(6), 643 – 661.

Baddely, A. (2002). Working memory and language processing. *Advances in Cognitive Science, 4*(2), 5 – 14.

Baksh, I. J. (1984). Factors influencing occupational expectations of secondary school students in Trinidad and Tobago. *Social and Economic Studies, 33*(3), 1 – 29.

Barber, N. (2005). Educational and ecological correlates of IQ: a cross-national investigation, *Intelligence, 33*(3), 273 – 284.

Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and Individual Differences, 42*, 815 – 824.

Beins, B.C. (2010). Teaching the history of psychology. Teaching measurement through historical sources. *History of Psychology, 13* (1), 89 – 94.

Benson, E. (2003). Intelligent intelligence testing. *Monitor, 34* (2), 48.

Bentler, P. M., & Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods, 15* (2), 111 – 123.

Binet, A. & Simon, T. (1905/1916). New methods for the diagnosis of the intellectual level of subnormals. In H.H. Goddard (Ed.), *Development of Intelligence in Children (the Binet – Simon Scale)* (pp. 199 – 244). Baltimore, Williams & Wilkins.

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24* (3), 383 – 405. Retrieved from http://dx.doi.org/10.1076/jcen.24.3.383.981

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology, 15,* 417-424.

Bollen, K.A., 1989. *Structural Equations with Latent Variables*. Wiley, New York.

Bracken, B. & McCallum, S. (1998). *Universal Nonverbal Intelligence Test.* Itasca, IL: Riverside.

Bradley, R.H. & Casey, P.H. (1992). Family environment and behavioral development in low birthweight children: annotation. *Developmental Medicine and Child Neurology, 34,* 822-826.

Bray, M. (2009). *Confronting the Shadow Education System: What Government Policies for What Private Tutoring?* Paris: UNESCO International Institute for Educational Planning (IIEP). Retrieved from http://www.unesco.org/iiep/eng/publications /recent/abstracts /2009/Bray_Shadoweducation.htm

Brooks-Gunn, J., McCarton, C.M., Casey, P.H., McCormick, M.C., Bauer, C.R., Bernbaum, J.C., Tyson, J., Swanson, M., Bennett, F.C., Scott, D.T., et al. (1994). Early intervention in low-birth-weight premature infants. results through age 5 years from the infant health and development program. *JAMA, 272*(16), 1257 – 1262.

Buchmann, C. (2002). Measuring family background in international studies of education: conceptual issues and methodological challenges. In A.C. Porter & A. Gamoran (Eds), *Methodological Advances in Cross-National Surveys of Educational Achievement* (pp. 150 – 197). Washington DC, National Academy Press.

Byrne B. 2004. Testing for multigroup invariance using AMOS Graphics, a road less traveled. *Structural Equation Modeling, 11,* 272 – 300.

Byrne, B., Coventry, W. L., Olson, R. K., Wadsworth, S. J., Samuelsson, S., Petrill, S. A., Wilcutt, E.G. & Corley, R. (2010). "Teacher effects" in early literacy development: evidence from a study of twins. *Journal of Educational Psychology, 10*2, 32 – 42.

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development, 60* (5), 1239 – 1249.

Calhoun, S. L., & Mayes, S. D. (2005). Processing speed in children with clinical disorders. *Psychology in the Schools, 42*, 333 – 343.

Campbell, C.C. (1996). *The Young Colonials: A Social History of Education in Trinidad and Tobago, 1834-1939*. Press University of the West Indies.

Campbell, E A. & Ramey, C. T. (1994). Effects of early intervention on/ intellectual and academic achievement: a follow-up study of children from low-income families. *Child Development, 65*, 684 – 698.

Canivez, G. L. (2014). Construct validity of the WISC–IV with a referred sample: direct versus indirect hierarchical structures. *School Psychology Quarterly, 29,* 38–51. doi:10.1037/ spq0000032.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Cattell, J.M. (1890). Mental tests and measurements. *Mind, 15,* 373 – 381. Retrieved from http://psychclassics.yorku.ca/Cattell/mental.htm

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153 – 193.

Cattell, R. B. (1987/1971). *Intelligence: Its structure, growth and action.* Amsterdam: Elsevier.

Ceci, S.J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*(5), 703 – 722.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence and income. *American Psychologist, 52*, 1051 – 1058.

Chabris, C.F., Hebert, B.M., Benjamin, D.J., Beauchamp, J. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science, 23*(11), 1314 - 1323. doi: 10.1177/0956797611435528.

Chamorro-Premuzic, T; & Furnham, A., (2005). *Personality and Intellectual Competence.* London: Lawrence Erlbaum Associates.

Chen, H., Keith, T. Z., Chen, Y. -H., & Chang, B. -S. (2009). What does the WISC– IV measure? Validation of the scoring and CHC-based interpretive approaches. *Journal of Research in Education Sciences, 54,* 85 – 108.

Chen, H., Keith, T., Weiss, L., Zhu, J., & Li, Y. (2010). Testing for multigroup invariance of second-order WISC– IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences, 49,* 677 – 682. doi:10.1016/j.paid.2011 .10.006

Church, A. T., & Katigbak, M. S. (1987). Ecocultural bias in culture-specific intelligence tests in an ecologically diverse culture: The Philippines. *Intelligence, 11*, 371-389.

Cocodia, E.A. (2014). Cultural perceptions of human intelligence. *Journal of Intelligence, 2,* 180 – 196. doi:10.3390/jintelligence2040180

Coleman, J. (1966). *Equality and Educational Opportunity.* U.S. Department of Health, Education, and Welfare.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). Cognitive outcomes in public and private schools. *Sociology of Education, 55*, 65 – 76.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M. & Kyllonen, P.C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence, 32*, 277 – 296.

Cores, E. V., Vanotti, S., Eizaguirre, B., Fiorentini, L., Garcea, O., Benedict, R.H.B. & Cáceres, F. (2015). The effect of culture on two information-processing speed tests. Applied *Neuropsychology: Adult, 22*(4), 241 – 245. doi:10.1080/23279095.2014.910214

Cornoldi, C., Giofre, D., Orsini, A. & Pezuti, L. (2014). Differences in the intellectual profile of children with intellectual vs. learning disability. *Research in Developmental Disabilities, 35,* 2224 – 2230.

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281 – 302.

Dang, H., Weiss, B. Pollack, A. & Nguyen, M.C. (2011). Adaptation of the Wechsler Intelligence Scale for Children-IV (WISC-IV) for Vietnam. *Psychological Studies, 56*(4), 387 – 392.

Darling-Hammond, L. (1995). Changing conceptions of teaching and teacher development. *Teacher Education Quarterly, Fall*, 9 – 26.

Daseking, M., Petermann, U. & Petermann, F. (2007). Assessment of intelligence with the HAWIK-IV. *Kindheit und Entwicklung, 16,* 250 – 259.

Davis-Kean, P. E. (2005). The Influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of Family Psychology, 19*(2), 294 – 304.

Deary. I.J. (2014). The stability of intelligence from childhood to old age. *Current Directions in Psychological Science, 23*(4), 239 -245. doi: 10.1177/0963721414536905

Deary, I.J. & Johnson, W. (2010). Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology, 39(*5), 1362 – 1369. doi:10.1093/ije/dyq072

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience, 11*, 201 – 211.

Deary, I. J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13 – 21.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology, 86*(1), 130 – 147.

DeLisle, J. (2012). Secondary School Entrance Examinations in the Caribbean: Legacy, policy and evidence within an era of seamless education. *Caribbean Curriculum, 19,* 103 – 143.

DeLisle, J., Keller, C., Jules, V. & Smith, P. (2009). When choosing might mean losing: the construction of secondary school choice in the Republic of Trinidad and Tobago. *Caribbean Curriculum, 16* (1), 131 – 176.

DeLisle, J., Seecharan,H. & Ayodike, A.T. (2010). *Is the Trinidad and Tobago education system structured to facilitate optimum human capital development? New findings on the relationship between education structures and outcomes from national and international assessments.* Paper presented at Turmoil and Turbulence in Small Developing States: Going beyond Survival. Sir Arthur Lewis Institute of Social & Economic Studies (SALISES) 11th Annual Conference, Trinidad and Tobago. Retrieved from http://sta.uwi. edu/conferences/09/salises/documents/J%20De%20Lisle.pdf

De Lisle, J., Smith, P., & Jules, V. (2005). Which males or females are most at risk and on what? An analysis of gender differentials within the primary school system of Trinidad and Tobago. *Educational Studies, 31*(4), 393-418.

De Lisle, J., Smith, J., Lewis, Y., Keller, C., Mc David, P., Jules, V., Lochan, S., Hackett, R., Pierre, P. & Seunarinesingh, K. (2009). In the context of Trinidad and Tobago, how do we identify schools that are succeeding or failing amidst exceptionally challenging circumstances? *Curriculum, 16*(1), 547 – 562.

Devena, S. E., Gay, C., & Watkins, M. W. (2013). Confirmatory factor analysis of the WISC-IV in a hospital sample. *Journal of Psychoeducational Assessment, 31,* 591 – 599.

DiStefano, C., Zhu, M. & Mindrila, D. (2009). Understanding and using factor scores: considerations for the applied researcher. *Practical Assessment, Research and Evaluation, 14*(20). Retrieved from http://pareonline.net/getvn.asp?v=14&n=20

Dominguez de Ramírez, R. (2007). System of multicultural pluralistic assessment. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics.* (pp. 989-990). Thousand Oaks, CA: SAGE Publications, Inc. doi: http://dx.doi.org/10.4135/9781412952644.n449

Dubow, E.F., Boxer, P. & Huesmann, L.R. (2009). Long-term effects of parents' education on children's educational and occupational success: mediation by family interactions, child aggression, and teenage aspirations. *Merrill Palmer Q (Wayne State University Press), 55*(3), 224 – 249. doi:10.1353/mpq.0.0030

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working-memory, short-term memory and general fluid intelligence: a latent variable approach. *Journal of Experimental Psychology, General, 128,* 309 – 331

Fancher, R.E. (1985). *The intelligence men: Makers of the IQ controversy*. New York: W.W. Norton & Company.

Fancher, R.E. (2009). Scientific cousins: The relationship between Charles Darwin & Francis Galton. *American Psychologist, 64* (2), 84 – 92. (Special Issue: Charles Darwin & Psychology, 1809-2009)

Fenollar-Cortés, J., Navarro-Soria, I., González-Gómez, C. &García-Sevilla, J. (2015). Cognitive profile for children with ADHD by using WISC-IV: subtype differences? *Revista de Psicodidáctica, 20*(1), 157 – 176. doi: 10.1387/ RevPsicodidact.12531

Field, A. P. (2009). *Discovering statistics using SPSS (3rd edition).* London: Sage Publications.

Fina, A.D., Sanchez-Escobedo, P., Hollingworth, L. (2010). *Adaptation of the WISC-IV for use in Mexico –a validity Study.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Denver, CO.

Fiorello, C.A., & Primerano, D. (2005). Research into practice: Cattell-Horn-Carroll cognitive assessment in practice: eligibility and program development issues. *Psychology in the Schools, 42*(5), 525 – 536.

Fischer, R. & Milfont, T. L. (2010). Standardization in psychological research. *International Journal of Psychological Research, 2*, 88 – 96.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV Assessment.* Hoboken, NJ: John Wiley & Sons.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). Use of the cross-battery approach in the assessment of diverse individuals. In A. S. Kaufman & N. L. Kaufman (Eds.), *Essentials of Cross-Battery Assessment - Second Edition* (2nd ed., pp. 146 –205). Hoboken, NJ: Wiley.

Floyd, R.G., Evan, J.J. & McGrew, K.S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*(2), 155–171.

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The relative contributions of the Cattell-Horn-Carroll cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools, 45*(2), 132 – 144.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286.

Flynn, J.R. (1984). The mean IQ of Americans: massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29 – 51.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*(2), 171 – 191.

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC-III full scale IQ in predicting later academic achievement. *Applied Neuropsychology, 15,* 131 – 139.

Friend, A., DeFries, J.C. & Olson, R. (2008). Parental education moderates genetic influences on reading disability. *Psychological Science, 19*(11), 1124 – 1130. doi:10.1111/ j.1467-9280.2008.02213.x

Friend, A., DeFries, J.C., Olson, R.K., Pennington, B., Harlaar, N., Byrne, B., Samuelsson, S., Wilcutt, E.G., Wadsworth, S.J., Corley, R. & Keenan, J.M. (2009). Heritability of high reading ability and its interaction with parental education. *Behavioural Genetics, 39*(4), 427 – 436. doi: 10.1007/s10519-009-9263-2.

Fry, A. F., & Hale, S. (1996). Processing speed, working-memory, and fluid intelligence: evidence for a developmental cascade. *Psychological Science, 7,* 237 – 241.

. Gagné, F. & St Père, F. (2002). When IQ is controlled, does motivation still predict achievement? *Intelligence, 30*(1), 71 – 100. Retrieved from http://dx.doi.org /10.1016/S0160-2896(01)00068-X

Galton, F. (1869). *Hereditary Genius*. London: Macmillan.

Galton, F. (1904). Eugenics: its definition, scope and aims. *The American Journal of Sociology, X*(1), 1 – 16.

Gamoran, A. & Long, D.A. (2007). Equality of educational opportunity: a 40 Year retrospective. *International Studies in Educational Inequality, Theory and Policy*, 23 – 47. Retrieved from http://link.springer.com/chapter/10.1007%2F978-1-4020-5916-2_2

Ganzach, Y. (2014). Adolescents' intelligence is related to family income. *Personality and Individual Differences, 59*, 112 – 115.

Gardner, H. (2000). *Intelligence Reframed: Multiple Intelligences for the 21st Century.* New York: Basic Books Inc.

Garkov, V. (2002). Is the American approach to science education the best in the world? Market forces—driving science education in the United States. *Journal of College Science Teaching, 3*(6), 399 – 401.

Georgas, J., Weiss, L., van de Vijver, F., Saklofske, D.H. (2003). *Culture and Children's Intelligence: Cross Cultural Analysis of the WISC-III.* London: Academic Press

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education, 40*, 103 – 114.

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2012). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. *Psychological Assessment. Advance online publication*. doi:10.1037/a0030676

Goldbeck L., Daseking, M., Hellwig-Brida, S., Waldman, H.C. & Petermann, F. (2010). Sex differences on the German Wechsler Intelligence Test for Children (WISC-IV). *Journal of Individual Differences, 31*(1), 22 – 28.

Goldstein, G., Allen, D., Minshew, N., Williams, D., Volkmar, F., Klin, A., & Schultz, R. T. (2008). The structure of intelligence in children and adults with high functioning autism. *Neuropsychology, 22*, 301–312.

Goleman, Daniel (1996). *Emotional Intelligence: Why It Can Matter More Than IQ.* Bantam Books.

Gottfredson, L. S. (1997a). Mainstream science on intelligence: an editorial with 52 signatories, history, and bibliography. *Intelligence 24*, 13 – 23. doi: 10.1016/S0160-2896(97)90011-8

Gottfredson, L.S. (1997b). Why g matters: the complexity of everyday life. *Intelligence, 24*, 79 – 132.

Gottfredson, L. S. & Deary, I.J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13*, 1 – 4.

Gottfredson, L.S. & Saklofske, D.H. (2009). Intelligence: foundations and issues in assessment. *Canadian Psychology, 50* (3), 183 – 195.

Gottfried, A., & Gottfried, A. (1984). Home environment and cognitive development in young children of middle-socioeconomic-status families. In A. Gottfried (Ed.), *Home environment and early cognitive development* (pp. 57 – 112). Orlando, FL: Academic Press.

Gregoire, J., Georgas, J., Saklofske, D. H., Van de Vijver, F., Weirzbicki, C., Weiss, L. G., & Zhu, J. (2008). *Cultural issues in clinical use of the WISC-IV*. In A. Prifitera, D. Saklofske, & L. Weiss (Eds.), WISC-IV Clinical Assessment and Intervention (2nd ed) (pp. 517 – 544). San Diego, CA: Elsevier.

Grigorenko, E.L., Geissler, P.W., Prince, R., Okatcha, F., Nokes, C., Kenny, D.A., Bundy, D.A. & Sternberg, R.A. (2001). The organization of Luo conceptions of intelligence: a study of implicit theories in a Kenyan village. International Journal of *Behavioural Studies, 25*(4), 367 – 378. doi:10.1080/01650250042000348

Guilford, J.P. (1966). Intelligence: 1965 model. *American Psychologist, 21*(1), 20 – 26. Retrieved from http://dx.doi.org/10.1037/h0023296

Guthke, J., Beckmann, J.F. & Dobat, H. (1997). Dynamic testing — problems, uses, trends and evidence of validity. *Educational and Child Psychology, 14*(4) 17-32.

Hale, J.B., Fiorello, C.A., Kavanagh, J.A., Hoeppner, J.B. & Gaither, R.A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: are global and factor scores comparable? *School Psychology Quarterly, 16*(1), 31 – 55. doi:10.1521/scpq.16.1.31.19158

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 3847.

Hanscombe, K.B., Trzaskowski, M., Haworth, C.M.A., Davis, O.S.P., Dale, P.S. & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): in a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS ONE 7*(2), e30320. doi:10.1371/journal.pone.0030320

Hanushek, E.A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature, 24* (3), 1141 – 1177.

Hanushek, E.A.& Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46* (3), 607–668.

Hanushek, E.A., Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital, 3* (2), 107–143.

Harden, K.P., Turkheimer, E., & Loehlin, J.C. (2007). Genotype by environment interaction in adolescents' cognitive aptitude. *Behavior Genetics, 37,* 273–283

Harlaar, N., Spinath, F.M., Dale, P.S. & Plomin, R. (2005). Genetic influences on early word recognition abilities and disabilities: a study of 7-year-old twins. Journal of *Child Psychology and Psychiatry, 46*(4), 373 – 84.

Hart, S.A., Petrill, S.A., Deckard, K.D. & Thompson, L.A. (2008). SES and CHAOS as environmental mediators of cognitive ability: a longitudinal genetic analysis. *Intelligence, 35*(3), 233 – 242. doi:10.1016/ j.intell.2006.08.004.

Hart, B. & Risley, T. R. (1992). American parenting of language-learning children: persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology, 28,* 1096 – 1105.

Haworth, C.M.A, Wright, M.J., Luciano, M., Martin, N.G., de Geus, E.J.C., van Beijsterveldt, C.E.M., Bartels, M., Posthuma, D., Boomsma, P.I., Davis, O.S.P., Kovas, Y., Corley, R.P., et al. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15*(11), 1112 -1120. doi: 10.1038/mp.2009.55

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, 2*(2). Retrieved from http://dx.doi.org/ 10.9707/2307-0919.1111

Healy, W. & Fernald, G.M. (1911) *Tests for Practical Mental Classification.* Baltimore: The Review.

Helms, J.E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47*(9), 1083 – 1101.

Henry, L. & MacLean, M. (2003). Relationship between working memory, expressive vocabulary and arithmetical reasoning in children with and without intellectual disabilities. *Educational and Child Psychology, 20*(3), 51 – 63.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Journal of Business Research Methods, 6*, 53 – 60.

Horn, J.L. (1968). Organization of abilities and development of intelligence. *Psychological Review, 75,* 242 – 259.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53 – 91). New York: Guilford.

Hu, L., & Bentler, P. M. (1998). Fit indexes in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424 − 453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1 − 55.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340–362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin, 96*, 72–98.

Jackson, C. K. (2008). *Ability-grouping and academic inequality: Evidence from rule-based student assignments.* Retrieved from Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/workingpapers/79/

Jackson, D.L., Gillaspy, J.A. & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological Methods, 14*(1), 6 – 23. doi: 10.1037/a0014694.

Jensen, A.R. (1980). *Bias in Mental Testing*. Free Press: NY

Jensen, A.R. (1997). The puzzle of nongenetic variance. In R.J. Sternberg & E. Grigorenko (Eds.) *Intelligence Heredity and environment.* Cambridge University Press. Retrieved from http://arthurjensen.net/wp-content/uploads/2014/06/The-Puzzle-of-Nongenetic-Variance-1996-by-Arthur-Robert-Jensen.pdf

Jensen, A. R. (1998). *The g factor*. Westport, CT: Preager.

Jirsa, J.E. (1983). The SOMPA. A brief examination of the technical considerations, philosophical rationale and implications for practice. *Journal of School Psychology, 21*, 13 – 21.

Johnson, D. L. & Danley, W. (1981). Validity: comparison of the WISC-R and SOMPA estimated learning potential scores. *Psychological Reports, 49*, 123 – 131.

Johnson, M.R., Shkura, K., Langley, S.R., Delahaye – Duriez, A. et al. (2015). Systems genetics identifies a convergent gene network for cognition and neurodevelopmental disease. *Nature Neuroscience (Advance Online Publication),* 1 – 10. doi: 10.1038/nn.4205

Kahana, S.Y & Glutting, J.J. Factor and subtest discrepancies on the DAS. examining the prevalence and validity in predicting academic achievement. *Assessment, 9(*1), 82 – 93.

Kan, K.-J., Kievit, R. A., Dolan, C. V., & van der Maas, H. L. J. (2011). On the interpretation of the CHC factor Gc. *Intelligence, 39*(5), 292 – 302.

Kan, K.J, Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. J. (2013). On the nature and nurture of intelligence and specific cognitive abilities: the more heritable, the more culture-dependent. *Psychological Science, 24*(10), 2420 – 2428.

Kaufman, A.S., & Kaufman, N.L. (2004). *Kaufman Assessment Battery for Children Second Edition.* Circle Pines, MN: American Guidance Service.

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: similarities and differences across ethnic groups. *School Psychology Quarterly, 14,* 239 – 262.

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues (2nd ed)* (pp. 581 – 614). New York, NY: Guilford Press.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multipsample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: what does it measure? *School Psychology Review, 35,* 108 – 127.

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence, 36*, 502–525.

Kieffer, D.A. & Goh, D.S. (1981). The effect of individually contracted incentives on intelligence test performance of middle- and low SES children. *Journal of Clinical Psychology, 37*(1), 175 – 179. doi:10.1002/1097-4679(198101)37:1<175::AID-JCLP2270370135>3.0.CO;2-1

Kim, K.H. (2005). Learning from each other: creativity in East Asian and American education. *Creativity Research Journal, 17*(4), 337 – 347.

Kim, H.Y. (2013). Statistical notes for clinical researchers: assessing normal distribution using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38*(1): 52 – 54. doi:10.5395/rde.2013.38.1.52

Kline, P. (2000). *Handbook of Psychological Testing.* Routledge, New York & London.

Kline, P. (2006). Intelligence. *The Psychometric View*. Routledge: NY.

Kline, R. B. (2013). Assessing statistical aspects of test fairness in structural equation modeling. *Educational Research and Evaluation, 19*, 204-222.

Kosmidis, M.H., Zafiri, M. & Politimou, N. (2011). Literacy versus formal schooling: influence on working memory. *Archives of Clinical Neuropsychology, 26*(7), 575 – 582. doi:10.1093/arclin/acr063

Kovas, Y., Haworth, C.M., Dale, P.S. & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development, 72*(3), 1 – 144.

Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S.B., Dale, P.S., Plomin, R. (2013). Literacy and numeracy are more heritable than intelligence in primary school. *Psychological Science, 24*(10), 2048 – 2056. doi: 10.1177/095679613486982.

Krapohl, E., Rimfeld, K., Shakeshaft, N.G., Trzaskowski, M., McMilan, A., Pingault, J., Asbury, K., Harlaard, N., Kovas, Y., Dale, P.S. & Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Psychological and Cognitive Sciences, 111*(42), 15273 – 15278. doi:10.1073/pnas.1408777111

Kubiszyn, T.W., Meyer, G.J., Finn, S.E., Eyde, L.D, Kay, G.G., Moreland, K.L., Dies, R.R. & Eisman, E.J. (2000). Empirical support for psychological assessment in clinical health care settings. *Professional Psychology: Research and Practice, 31*(2), 119 – 130.

Kudiaki, C., Aslan, A. (2008). Executive functions in a Turkish sample: associations with demographic variables and normative data. *Applied Neuropsychology, 15*(3), 194 – 204. doi: 10.1080/09084280802324416.

Kyllonen, P. C. (1994). Aptitude testing inspired by information processing: a test of the four-sources model. *Journal of General Psychology, 120,* 375 – 405.

Laidra, K., Pullman, H., Allik, J. (2007). Personality and intelligence as predictors of academic achievement: a cross-sectional study from elementary to secondary school. *Personality and Individual Differences, 42,* 441 – 451.

Lankford, H., Loeb, S. & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: a descriptive analysis. *Educational Evaluation and Policy Analysis, 24*(1), 37 – 62.

Lecerf, T., Rossier, J., Favez, N., Reverte, I, & Coleaux, L. (2010). The four- vs. alternative six-factor structure of the French WISC-IV: comparison using confirmatory factor analyses. *Swiss Journal of Psychology, 69*(4), 221 – 232.

Lee, H., & Barratt, M. S. (1993). Cognitive development of premature low birth weight children at 5 to 8 years old. *Development and Behavioral Pediatrics, 14,* 242 – 249.

Lemos, G. C., Almeida, L. S., & Colom, R. (2011). Intelligence of adolescents is related to their parents' educational level but not to family income. *Personality and Individual Differences, 50*(7), 1062 – 1067. Retrieved from http://www.sciencedirect.com/ science/article/pii/S0191886911000468

Levine, R., & Bartlett K. (1984). Pace of life, punctuality and coronary heart disease in six countries. *Journal of Cross-Cultural Psychology, 15*, 233 – 255.

Levine, R.V. & Norenayan, A. (1999). The pace of life in 31 countries. Journal of *Cross-Cultural Psychology, 30*(2), 178 – 205.

Liaw, R, & Brooks-Gunn, J. (1993). Patterns of low-birth- weight children's cognitive development. *Developmental Psychology, 29*, 1024 – 1035.

Lissitz, R.W. (2009). *The Concept of Validity. Revisions, New Directions, and Applications*. Charlotte, NC: Information Age Publishing, Inc.

Locurto, C. (1990). The malleability of IQ as judged from adoption studies. *Intelligence, 15*, 295 – 312.

Lohman, D.F., & Lakin, J.M. (2009). Reasoning and intelligence. In R. J. Sternberg and S. B. Kaufman (Eds.), *Handbook of Intelligence.* New York: Cambridge University Press.

London, N.A. (1989). Selecting students for secondary education in a developing society: The case of Trinidad and Tobago. *Journal of Education, 24*(3), 281 – 292.

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problem*s. Hillsdale, Erlbaum Associates.

Lorenzo-Seva, U. & ten Berge, J.M.F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57 – 64.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "General Intelligence" objectively determined and measured. *Journal of Personality and Social Psychology, 86*(1), 96 – 111.

Mayes. S.D., & Calhoun, S.L. (2006). Frequency of reading, math and writing disabilities in children with clinical disorders. *Learning and Individual Differences, 16*, 145 – 157.

McCarton, C.M., Brooks-Gunn, J., Wallace, I.F., Bauer, C.R., Bennett, F.C., Bernbaum, J.C., Broyles, R.S., Casey, P.H., McCormick, M.C., Scott, D.T., Tyson, J., Tonascia, J. & Meinert, C.L. (1997). Results at age 8 years of early intervention for low-birth-weight premature infants. The infant health and development program. *JAMA, 277*(2), 126 – 132.

McClain, A. L. (1996). Hierarchical analytic methods that yield different perspectives on dynamics: aids to interpretation. *Advances in Social Science Methodology, 4*, 229 – 240.

McGill, R.J. & Canivez, G.L. (2016). Orthogonal higher order structure of the WISC-IV Spanish using hierarchical exploratory factor analytic procedures. Journal of

Psychoeducational Assessment, 1 -7. Advance Online Publication. doi:10.1177/0734282915624293

McGraw, K.O. & Wong, S. P. (1996). Forming inferences about some intra-class correlation coefficients. *Psychological Methods, 1*(1), 30-46. Retrieved from http://dx.doi.org/ 10.1037/1082-989X.1.1.30

McGrew, K. (2009). Editorial: CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research, *Intelligence, 37,* 1 – 10.

McGrew, K., Flanagan, D., Keith, T. & Vanderwood, M. (1997). Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26*(2), 189 – 210.

Mc Grew, K.S., & Wendling, B.J. (2010). Catell-Horn- Carroll cognitive-achievement relations: what have we learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651 – 675.

Mejias-Contreras, D.M. & Albarracin-Rodriguez, A.P. (2013). Preliminary study of the psychometric properties of the WISC-IV in school children in Bucaramanga. *Informes Psicologicos, 13*(2), 13 – 25.

Mercy, J.A., & Steelman, L.C. (1982). Familial influence on the intellectual attainment of children. *American Sociological Review, 47*, 532 – 542.

Mercer, J.R. (1979). In defense of racially and culturally non-discriminatory assessment. *School Psychology Digest, 8*, 89 – 115.

Moffitt, T. E., Gabrielli, W. F., Mednick, S. A., & Schulsinger, F. (1981). Socioeconomic status, IQ, and delinquency. *Journal of Abnormal Psychology, 90,* 152–156.

Montagnes, I. (2001). Textbooks and learning materials 1990-99. *Education for All 2000 Assessment, Thematic Studies:* UNESCO, ADEA, p27

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report. IEA's Progress in International Reading Literacy Study in primary schools in 40 countries.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Munford, P.R. (1978). A comparison of the WISC and WISC-R on black child psychiatric outpatients. *Journal of Clinical Psychology, 34*(4), 938 – 943.

Munford, P.R., Meyerowitz, B.E. & Munford, A.M. (1980). A comparison of black and white children's WISC/WISC-R differences. *Journal of Clinical Psychology, 36*(2), 471 – 475.

Munford, P.R. & Munoz, A. (1980). A comparison of the WISC and WISC-R on Hispanic children. *Journal of Clinical Psychology, 36*(2), 452 – 458. doi:10.1002/jclp. 6120360211

Muthén, L.K. and Muthén, B.O. (1998-2010). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.

Nakano, S. (2011). *Factor structure of the Wechsler Intelligence Scales for Children–Fourth Edition among referred Native American students.* Doctoral Dissertation, Arizona State University. Retrieved from http://repository.asu.edu/attachments /93096/content/tmp/ package-LDreXT/Nakano_asu_0010E_11015.pdf

National Center for Learning Disabilities. (2014). *The State of Learning Disabilities – 3rd Edition.* Retrieved from: https://www.ncld.org/wp-content/uploads/2014/11/2014-State-of-LD.pdf

Neiss, M. & Rowe, D.C. (2000). Parental education and child's verbal IQ in adoptive and biological families in the National Longitudinal Study of Adolescent Health. *Behavior Genetics, 30*(6):487 – 495.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J. & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist, 51*, 77 – 101. doi:10.1037/0003-066X.51.2.77

Nelson, C.A., Zeanah, C.H., Fox, N.A., Marshall, P.J., Smyke, A.T. & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: the Bucharest Early Intervention Project. *Science, 318*(5858), 1937 – 1940.

Nettle, D. (2003). Intelligence and class mobility in the British population. British *Journal of Psychology, 94*, 551 – 561.

Neubauer, A.C., & Benischke, C. (2002). A cross-cultural comparison of the relationship between intelligence and speed of information processing in Austria vs. Guatemala. *Psychologische Beitraege, 44*, 521 – 534.

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: current research in Cattell-Horn-Carroll-based assessment. *Psychology in the Schools, 47*(7), 621 – 634.

Nisbett, R.E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D.F. & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American Psychologist, 67*(2), 130 –159. doi: 10.1037/a0026699

Nye, B., Konstantopoulous, S. & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237 – 257.

Oakland, T. (1977). *Pluralistic norms and estimated learning potential.* Paper presented at the Annual Convention of the American Psychological Association. San Francisco, California.

Oakland, T. (1979). Research on the ABIC and ELP: A revisit to an old topic. *School Psychology Digest, 8*, 209-213.

OECD (2011). *History of DAC Lists of Aid Recipient Countries.* Retrieved from: http://www.oecd.org/dac/stats/historyofdaclistsofaidrecipientcountries.htm

Oliveras-Rentas, R.E., Kenworthy, L.E., Roberson, III, R.B., Martin, A. & Wallace, G.L. (2012). WISC-IV profile in high-functioning autism spectrum disorders: Impaired processing speed is associated with increased autism communication symptoms and decreased adaptive communication abilities. *Journal of Autism and Developmental Disorders, 42*(5), 655 – 664.doi: 10.1007/s10803-011-1289-7

Olson, R.K., Keenan, J.M., Byrne, B. & Samuelsson, S. (2014). Why do children differ in their development of reading and related skills? *Scientific Studies of Reading, 18*(1), 38 – 54. doi: 10.1080/10888438.2013.800521

Ostrosky-Solı´s, F. & Lozano, A. (2006). Digit span: effect of education and culture. *International Journal of Psychology, 41*(5), 333 – 341.

Panicker, A.S., Hirisave, U, & Subbakrishna, D.K (2006). . WISC-III UK: comparison of Indian and UK norms. *Journal of the Indian Association of Child and Adolescent. Mental Health, 2*(4), 108 – 111.

Parkin, J.R. & Beaujean, A.A. (2012). The effects of Wechsler Intelligence Scale for Children – Fourth Edition cognitive abilities on math achievement. *Journal of School Psychology, 50*, 113 – 128.

Passolunghi, M.C. & Cornoldi, C. (2008). Working memory failures in children with arithmetical difficulties. *Child Neuropsychology, 14*(5), 387 – 400. doi: 10.1080/09297040701566662

Payton, A. (2009). The impact of genetic research on our understanding of normal cognitive ageing: 1995 to 2009. *Neuropsychology Review, 19*(4), 451 – 477.

Petrill S.A., Pike, A., Price, T. & Plomin, R. (2004). Chaos in the home and socioeconomic status are associated with cognitive development in early childhood. Environmental mediators identified in a genetic design. *Intelligence, 32*, 445 – 460.

Plomin, R. & Deary, I.J. (2014). Genetics and intelligence differences: five special findings. *Molecular Psychiatry*, 1 – 11.

Plomin, R., Owen, M.J. & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science, 264*(5166), 1733 – 1739.

Plomin, R, & Rutter, M. (1998). Molecular genetics, and what to do with genes once they are found. *Child Development, 69*(4), 1223-1242.

Plomin, R. & Spinath, F.M. (2002). Genetics and general cognitive ability (g). *Trends in Cognitive Sciences, 6*(4), 169 – 176.

Poteat, G. M., Wuensch, K. L., & Gregg, N. B. (1988). An investigation of differential prediction with the WISC-R. *Journal of School Psychology, 26*, 59 – 68.

Prifitera, A., Weiss, L.G. Saklofske, D.H. & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In Prifitera, Saklofske & Weiss eds. *WISC-IV Clinical Use and Interpretation: Scientist-Practitioner Perspectives* (pp. 3 – 28). London, Elsevier, Inc.

Public Health England. (2013). *People with Learning Disabilities in England* 2013. Retrieved from http://www.improvinghealthandlives.org.uk/securefiles/ 1510251825/People%20with%20learning%20disabilities%20in%20England%202013.pdf

Raiford, S.E., Weiss, L.G., Rolfhus, E., & Coalson, D. (2005). *WISC-IV General Ability Index (WISC-IV Technical Report No. 4).* Retrieved from http://harcourtassessment. com/hai/Images/pdf/wisciv/WISCIVTechReport4.pdf

Rambhajan, B. (2007. Sea of tears. *Trinidad & Tobago Express.* Retrieved from www.trinidadexpress.com

Raven, J., Raven, J.C., & Court, J.H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford: Oxford Psychologists Press.

Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology, 118*(3), 219 – 235.

Reeve, C.L., & Bonaccio, S. (2011). The nature and structure of "intelligence. In T. Chamorro-Premuzic, S. von-Stumm & A. Furnham (Eds.), *The Wiley Blackwell Handbook of Individual Differences,* 1st Edition, Blackwell Publishing Ltd.

Reschly, D.J. (1980) Nonbiased Assessment. Iowa Department of Public Instruction. Iowa State University. Retrieved from http://files.eric.ed.gov/ fulltext/ED209810.pdf

Reschly, D.J. (1980). School psychologists and assessment in the future. *Professional Psychology, 11*(6), 841 – 848. Retrieved from http://dx.doi.org/10.1037 /07357028.11.6.841

Reschly, D.J. (1981). Psychological testing in educational classification and placement. *American Psychologist, 36*(10), 1094 – 1102. Retrieved from http://dx.doi.org/ 10.1037/0003-066X.36.10.1094

Reverte, I., Golay, P., Favez N., Rossier, J. & Lecerf, T. (2014). Structural validity of the Wechsler Intelligence Scale for Children (WISC-IV) in a French-speaking Swiss sample. *Learning and Individual Differences, 29,* 114 – 119.

Reverte I., Golay P., Favez N., Rossier J. & Lecerf, T. (2015) Testing for multigroup invariance of the WISC-IV structure across France and Switzerland: Standard and CHC models. *Learning and Individual Differences, 40*, 127 – 133.

Reynolds, C.R., Keith, T.Z., Fine, J.G., Fisher, M.E., & Low, J.A. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children-Second Edition: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly, 22*(4), 511 – 539.

Reynolds, C. R., Lowe, P. A., & Saenz, A. (1999). The problem of bias in psychological assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The Handbook of School Psychology (3rd ed)* (pp. 549 – 595). New York, NY: Wiley.

Reynolds, C.R., Sanchez, S. & Willson, V.L. (2006). Normative tables for calculating the WISC-III performance and full scale IQs when symbol search is substituted for coding. *Psychological Assessment, 8,* 378 – 382.

Rimfeld, K., Kovas, Y., Dale, P.S., Plomin, R. (2015). Pleiotropy across academic subjects at the end of compulsory education. *Scientific Reports, 5,* 11713. doi: 10.1038/srep11713.

Rietveld C.A., Medland S.E., Derringer J., Yang J., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340*(6139), 1467-1471. doi: 10.1126/science.1235488.

Rindermann, H., Flores-Mendoza, C., & Mansur-Alves, M. (2010). Reciprocal effects between fluid and crystallized intelligence and their dependence on parents' socioeconomic status and education. *Learning and Individual Differences, 20,* 544 – 548.

Rivkin, S.G., Hanushek, E.A. & Kain, J.F. (2005). Teachers, schools and academic achievement. *Econometrica, 73*(2), 417 – 458.

Roberson, D., Davidoff, J. Davies, L.R. & Shapiro, L.R. (2005). Colour categories in Himba: evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50,* 378 – 411.

Roberts, R.D., Schulze, B, & MacCann, C. (2008). The measurement of emotional intelligence: a decade of progress? In G. Boyle, D. Matthews & D. Saflofske (Eds), *The SAGE handbook of personality theory and assessment. Personality measurement and testing* (pp. 461 – 482). Thousand Oaks: Sage.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*(1), 83-92.

Ross-Reynolds, J. & Reschly, D. J. (1983). An investigation of item bias on the WISC-R with four sociocultural groups. *Journal of Consulting and Clinical Psychology, 51*(1), 144 – 146.

Rowe, D.C., Jacobson, K.C. & Van den Oord, E.J. (1999). Genetic and environmental influences on vocabulary IQ: parental education level as moderator. *Child Development, 70*(5), 1151 – 1162.

Rowe, E. W., Kingsley, J. M., & Thompson, D. F. (2010). Predictive ability of the General Ability Index (GAI) versus the Full Scale IQ among gifted referrals. *School Psychology Quarterly, 25*, 119 – 128. doi:10.1037/a0020148

Rust. J. & Golombok S. (1992). *Modern Psychometrics: The Science of Psychological Assessment.* London: Routledge.

Ryan, J. J., Glass, L. A., & Bartels, J. M. (2010). Stability of the WISC-IV in a sample of elementary and middle school children. *Applied Neuropsychology, 17,* 68 –72. doi:10.1080/09084280903297933.

Saklofske, D. H. & Janzen, H.L. (1990). School-based assessment research in Canada. *McGill Journal of Education. 25*(1), 5 – 23. Retrieved from http://mje.mcgill.ca/article/ viewFile/7896/5825

Saklofske, D. H., Prifitera, A., Weiss, L. G., Rolfhus, E., & Zhu, J. (2005). Clinical interpretation of the WISC–IV FSIQ and GAI. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC–IV Clinical Use and Interpretation: Scientist-Practitioner Perspectives* (pp. 33–65). New York: Academic Press.

Saklofske, D.H., Zhu, J., Coalson, D.L., Raiford, S.E., Weiss, L.G. (2008). Cognitive proficiency index for the Canadian edition of the Wechsler Intelligence Scale for Children-fourth edition. *Canadian Journal of School Psychology, 25*(3), 277 – 286. doi: 10.1177/0829573510380539

Salthouse, T. A. (1992). Working-memory mediation of adult age differences in integrative reasoning. *Memory and Cognition, 20,* 413 – 423.

Samuelsson, S., Byrne, B., Wadsworth, S., Corley, R., DeFries, J.C., Willcutt, E., Hulslander, J. & Olson, R.K. (2007). Genetic and environmental influences on prereading skills and early reading and spelling development: a comparison between United States, Australia, and Scandinavia. *Reading and Writing: An Interdisciplinary Journal, 20,* 51 – 75. doi:10.1007/s11145-006-9018-x

Sandoval, J. (1979). The WISC-R and internal evidence of test bias with minority groups. *Journal of Consulting and Clinical Psychology, 47,* 919 – 927.

Sattler, J.M. (2001). *Assessment of children (4th ed.).* San Diego, CA: Jerome Sattler, Publishers, Inc.

Scarr, S., & McCartney, K. (1983). How people make their own environments: a theory of genotype - environment effects. *Child Development, 54,* 424 – 435.

Schmid, J., & Leiman, J. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53 – 61.

Schmidt, F.L. & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162 – 173. doi: 10.1037/0022-3514.86.1.162

Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417 – 453.

Solanto, M., Gilbert, S., Raj, A., Zhu, J., Pope-Boyd, S., Stepak, B., Vail, L. & Newcorn, J.H. (2007). Neurocognitive functioning in AD/HD predominantly inattentive and combined subtypes. *Journal of Abnormal Child Psychology, 35*, 729–744.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201 – 292.

Spearman, C. (1939). Thurstone's work re-worked. *Journal of Educational Psychology, 30*(1), 1 – 16. Retrieved from http://dx.doi.org/10.1037/h0061267

Spence, J. (2007), Inadequate schooling. *Trinidad & Tobago Express.* Retrieved from www.trinidadexpress.com.

Spinath, B., Freudenthaler, H.H. & Neubauer, A.C. (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. Personality and Individual Differences, 48, 481 – 486.

Statistics Canada (2012). Learning Disabilities Among Canadians Aged 15 Years and Older. Retrieved from http://www.statcan.gc.ca/pub/89-654-x/89-654-x2014003-eng.htm

Steinbach, M. (2012). Obstacles to change in teacher education in Trinidad and Tobago. *The International Education Journal: Comparative Perspectives, 11*(1), 1443 – 1475.

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Intelligence.* Cambridge: Cambridge University Press.

Sternberg, R. J. (2004). *Culture and intelligence. American Psychologist, 59*, 325 – 338.

Sternberg, R. J., & Detterman, D. K. (1986). *What is Intelligence?* Norwood, USA: Ablex.

Sternberg, R. J., & Grigorenko, E. L. (2000). *Teaching for Successful Intelligence.* Arlington Heights, IL: Skylight.

Sternberg, R. J., & Hedlund, J. (2002). Practical intelligence, g, and work psychology. *Human Performance, 15*(1/2), 143 – 160.

Sternberg, R.J., & Kaufman, J.C. (1996). Innovation and intelligence testing: The curious case of the dog that didn't bark. *European Journal of Psychological Assessment, 12* (3), 175 – 182.

Sternberg, R.J., Nokes, C., Geissler, W., Prince, R., Okatchae, F., Bundy, D.A., Grigorenko, E.L. (2001). The relationship between academic and practical intelligence: a case study in Kenya. *Intelligence, 29*(5), 401 – 418.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*, 401 – 426.

Styck, K. & Watkins, M. (2013). Diagnostic utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children—Fourth Edition among referred students. *School Psychology Review, 42*(4), 367 – 382.

Swanson, H. L. & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research, 76,* 249 – 274.

Swanson, L. & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence, 35,* 151 – 168. doi:10.1016/j.intell.2006.07.001

Süß, H.M., Oberauer, K., Wittmann, W.W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability – and a little bit more. *Intelligence, 30*(3), 261-288.

Tabachnick, B.G. & Fidell, L.S. (2013). *Using Multivariate Statistics, 6th Ed.* Boston: Pearson.

Taub, G.E., Floyd, R.G., Keith, T.Z., & McGrew, K.S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23* (2), 187 – 198.

Taub, G.E., & McGrew, K.S. (2004). A Confirmatory factor analysis of Cattell-Horn-Carroll theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly, 19*(1), 72 – 87.

Tempest, P. (1998). Local Navajo norms for the Wechsler intelligence Scale for Children-Third Edition. *Journal of American Indian Education, 37*(3), 18 – 30.

Terman, L.M. (1917). Binet scale and the diagnosis of feeble-minded. Journal of *Criminal Law and Criminology, 7*(4), 530 – 543. Retrieved from http://scholarlycommons. law.northwestern.edu/cgi/viewcontent. cgi?article=1474& context=jclc

Terman, L.M., Lyman, G., Ordahl, G., Ordahl, L.E., Galbreath, N & Talbert, W. (1917). *The Stanford revision and extension of the Binet-Simon scale for measuring intelligence.* Baltimore, Warwick & York, Inc. Retrieved from https://archive.org/details/ stanfordrevision00term

Thaler, N.S., Bello, D.T. & Etcoff, L.M. (2013). WISC-IV profiles are associated with differences in symptomatology and outcome in children with ADHD. *Journal of Attention Disorders, 17*(4), 291 – 301. doi: 10.1177/1087054711428806.

Thorndike, R.M., Cunningham, G.K., Thorndike, R.L. & Hagen, E.P. (1992). *Measurement and Evaluation in Psychology and Education.* New York: Macmillan Publishing Company.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Tong, S., Baghurst, P., Vimpani, G. & McMichael, A. (2007). Socioeconomic position, maternal IQ, home environment, and cognitive development. *Journal of Pediatrics, 151*, 284 – 288. doi:10.1016/j.jpeds.2007.03.020.

Trinidad and Tobago Central Statistical Office (2011). *Trinidad and Tobago 2011 Population and Housing Census: Demographic Report.* Port of Spain: The Central Statistical Office.

Trinidad and Tobago Guardian. (2014). *Proposed List of CAPE 2014 Scholarship Recipients*. Retrieved from http://www.guardian.co.tt/news/2014-10-21/proposed-list-cape-2014-scholarship-recipients

Trinidad and Tobago Ministry of Education (1960). *Concordat of 1960*. Retrieved from http://moe.edu.tt/

Trinidad and Tobago Ministry of Education. (2011). Education Sector Strategic Plan 2011-2015. Retrieved from http://planipolis.iiep.unesco.org/upload/ Trinidad%20and% 20Tobago/ Trinidad%20and%20Tobago_Strategic_plan_2011-2015.pdf

Trinidad and Tobago Ministry of Education (2013). *SEA Language Arts Examination 2013*. Retrieved from: http://moe.edu.tt/messages-releases/item/491-software-sea-math-2010-2013-past-papers

Trinidad and Tobago Ministry of Education (2014a). *Student Exams*. Retrieved from http://www.moe.gov.tt/student_exam_sea.html

Trinidad and Tobago Ministry of Education (2014b). *SEA Mathematics Examination 2014.* Retrieved from http://moe.edu.tt/messages-releases/item/491-software-sea-math-2010-2014-past-papers

Trinidad and Tobago Ministry of Finance and the Economy. (2014). *Budget Statement 2014.* Retrieved from http://finance.gov.tt/wp-content/uploads/2013/11/Budget-Statement-2014.pdf

Trinidad and Tobago Ministry of Public Administration. (2014). *Scholarship Winners 2014.* Retrieved from http://www.scholarships.gov.tt/

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. *Personnel Research Section Report No. 984.* Washington, DC: Department of the Army.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14,* 623 – 628.

Tynes-Jones, J. (2005). Applying the bio-ecological model of intelligence to the Wechsler Scales in the Bahamas. *College of the Bahamas Research Journal, 13,* 68 – 80.

van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology, 54*, 119 – 135. doi:10.1016/j.erap.2003.12.004

van IJzendoorn, M. H., Juffer, F., & Klein Poelhuis, C. W. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin, 131*, 301-316.

Vazquez-Nuttall, E. (1979). Test review. System of Multipluralistic Assessment (SOMPA). *Journal of Educational Measurement, 16*(4), 285 – 290.

Verhagen, J. & Leseman, P. (2016). How do verbal short term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology, 141,* 65 – 82. doi:10.1016/j.jecp.2015.06.015.

Vernon, P.E. 1961. *The structure of human abilities (2nd Ed.).* London: Methuen.

von Stumm, S. & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence, 48*, 30 -36.

Vrieze, S.I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), *Psychological Methods, 17*(2), 228–243

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19–27.

Watkins, M. W. (2003). IQ subtest analysis: Clinical acumen or clinical illusion. *Scientific Review of Mental Health Practice, 2*, 118 – 141.

Watkins, M.W. (2006). Structure of the Wechsler Intelligence Scale for Children--Fourth Edition among a national sample of referred students. *Psychological Assessment; 22*(4), 782 – 787. doi: 10.1037/a0020043.

Watkins, M.W. (2010). Structure of the Wechsler Intelligence Scale for Children—Fourth Edition among a national sample of referred students. *Psychological Assessment, 22*(4), 782-787.

Watkins, M. W., Canivez, G. L., James, T., James, K., & Good, R. (2013). Construct validity of the WISC-IV^{UK} with a large referred Irish sample. International *Journal of School & Educational Psychology, 1*, 102 – 111

Watkins, M. W., Glutting, J. J., & Lei, P.-W. (2007). Validity of the full scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology, 14,* 13 – 20.

Watkins, M.W. & Smith, L.G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children—Fourth Edition. *Psychological Assessment, 25*(2), 477–483. doi: 10.1037/a0031653.

Watkins, M.W., Wilson, S.M., Kotz, K.M., Carbone, M.C., & Babula T. (2006). Factor structure of the Wechsler Intelligence Scale for Children-Fourth Edition among referred students. *Educational and Psychological Measurement, 66*(6), 975–983.

Wechsler, D. (1971). Concept of collective intelligence. *American Psychologist, 26*(10), 904 – 907. http://dx.doi.org/10.1037/h0032223

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: The Psychological Corporation.

Wechsler, D. (1991). *The Wechsler intelligence scale for children—third edition*. San Antonio, TX: The Psychological Corporation.

Wechsler D. (2001a). *Wechsler Individual Achievement Test (2nd ed.).* San Antonio, TX: NCS Pearson.

Wechsler D. (2001b). *Wechsler Individual Achievement Test: (2nd ed.).*San

Antonio, TX: NCS Pearson.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence-Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003a). *Wechsler intelligence scale for children (4th ed.).* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003b). *The WISC-IV technical and interpretive manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003c). *The WISC-IV Scoring and Administration Manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003d). *The WISC-IV Stimulus Booklet.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2004a). *The Wechsler Intelligence Scale for Children, Fourth Edition Canadian Manual.* Toronto, Ontario, Canada: Psychological Corporation.

Wechsler, D. (2004b). *The Wechsler Intelligence Scale for Children, Fourth Edition (Spanish).* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2004c). *The Wechsler Intelligence Scale for Children, Fourth Edition*. London: Pearson Assessment.

Wechsler, D. (2005). *Manuel de l'Echelle d'Intelligence de Wechsler pour Enfants – 4e édition (Manual for the Wechsler Intelligence Scale for Children – fourth edition).* Paris: Editions du Centre de Psychologie Appliquée.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children–Fifth Edition.* San Antonio, TX: NCS Pearson.

Weiss, L.G. (2003). The WISC-III in the United States. In Georgas, Weiss, van de Vijver, Saklofske (Eds), *Culture and Children's Intelligence: Cross Cultural Analysis of the WISC-III* (pp. 41 – 58). London: Academic Press.

Weiss, L.G., Harris, J.G., Prifitera, A., Courville, T., Rolfhus, E., Holdnack, J.A., Saklofske, D.H., & Holdnack, J.A (2006). *WISC-IV interpretation in societal context. The essentials and beyond.* In L.G. Weiss, D.H. Saklofske, A. Prifitera & J.A. Holdnack (Eds), WISC-IV Advanced Clinical Interpretation (pp. 1 – 49). London: Elsevier.

Weiss, L. G., & Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC-III FSIQ across ethnic and gender groups. *Journal of School Psychology, 33,* 297 – 304.

Weiss, L.G., Prifitera, A., Holdnack, J.A., Saklofske, D.H., Rolfhus, E. & Coalson, D. (2006). The essentials and beyond. In L.G. Weiss, D.H. Saklofske, A. Prifitera & J.A. Holdnack (Eds), *WISC-IV Advanced Clinical Interpretation* (pp. 59 – 96). London: Elsevier.

Weiss, L. G., Prifitera, A., & Roid, G. (1993). The WISC-III and the fairness of predicting achievement across ethnic and gender groups. Advances in psychoeducational assessment; The Wechsler Intelligence Scale for Children: Third Edition. *Journal of Psychoeducational Assessment, Monograph Series,* 35 – 42.

West, S.G., Finch, J.F. & Curran, P.J. (1995). Structural equation models with nonnormal variables: problems and remedies. In R.H. Hoyle, (Ed), *Structural Equation Modeling: Concepts, Issues and Applications* (pp. 56 – 75). Newbery Park, CA: Sage.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461 – 481. Retrieved from http://dx.doi.org/ 10.1037/0033-2909.91.3.461

White, S. (2000). *Conceptual foundations of IQ testing.* Psychology, Public Policy and Law, 6(1), 33 – 43.

Williams, E. (1964). *History of the People of Trinidad and Tobago.* London: Andre Deutsch.

Willms, J.D. & Somers, M.A. (2001). Family, classroom and school effects on children's educational outcomes in Latin America. *School Effectiveness and School Improvement, 12*(4), 409 – 445.

Winne, P.H. & Nesbit, J.C. (2010). The psychology of academic achievement. *Annual Review of Psychology, 61*, 653-678.

Wonderlic, E. F. (1992). *Manual of the Wonderlic Personnel Test & Scholastic Level Exam*. IL: Wonderlic Personnel Test, Inc.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Woodworth, R. S., & Wells, F. L. (1911). Association tests. *Psychological Monograph. 57*, 1 – 80. Retrieved from http://archive.org/stream/associationtests 13ameruoft/ associationtests13ameruoft_djvu.txt

World Bank (1995). *Trinidad and Tobago: Poverty and unemployment in an oil based economy.* Report No. 14382-TR. Country Department III. Country Operations II Latin America and the Caribbean Region. Washington, US: Author.

World Bank (2015). *Data: High Income*. Retrieved from: http://data.worldbank .org/ income-level/HIC

World Economic Forum (2014). *Global Economic Reports*. Retrieved from: http://www.weforum.org/reports/global-competitiveness-report-2014-2015.

Wraw, C., Deary, I.J., Galea, C.R. & Derc, G. (2015). Intelligence in youth and health at age 50. *Intelligence, 53*, 23 – 32.

Yang, S-Y. & Sternberg, R.J. (1997). Taiwanese Chinese people's conceptions of intelligence. *Intelligence, 25*, 21 – 36.

Zeanah, C.H., Nelson, C.A., Fox, N.A., Smyke, A.T., Marshall, P., Parker, S. & Koga, S. (2003). Designing research to study the effects of institutionalization on brain and behavioral development: The Bucharest Early Intervention Project. *Development and Psychopathology, 15,* 885 – 907. doi: 10.1017.S0954579403000452.

APPENDIX A

CONSENT FORM

Study: **Investigating the validity of the WISC-IV for Trinidad and Tobago**

Dear Parent(s) Or Guardian(s):

I am writing to ask your permission for your child to participate in a Goldsmiths University of London research project designed to investigate the validity of the Wechsler Intelligence Scale for Children -4th Edition (WISC-IV) for use in Trinidad and Tobago. You may or may not know that for over 10 years the WISC-IV has been used by many psychologists throughout the country. Studies have demonstrated its validity as a measure of intelligence in American children, but as yet there has been no formal study to determine whether or not the WISC-IV is suitable for use in Trinidad and Tobago.

Trained testers will be assessing your child’s intellectual functioning using the WISC-IV. The project in which your child has been invited to participate is expected to be a positive experience for most children, one in which activities are expected to be challenging and enjoyable. Testers will take care to establish rapport with your child and ensure that they feel comfortable about being tested. If your child is unwilling to participate, then we will give them the option to withdraw even if you had previously given your consent.

Few children may experience extreme or unusual anxiety. If you know that your child is likely to be very anxious or fearful in a testing situation then we think it may be best that you do not volunteer your child for participation.

Even if you have given your consent, you may withdraw your permission at any time during the study without penalty by indicating this decision to the researcher.

Testing is expected to take between 45 minutes and 120 minutes of your child’s time and he or she will be offered breaks during the session.

Your child’s performance on the WISC-IV is considered confidential and as such results will not be shared with anyone. However, group findings will be summarized in the final report. If you request this on the demographic questionnaire, a summary of the findings of the study will be sent by email to you once the report is completed.

I would like to assure you that this study has been reviewed and received ethics clearance through a Goldsmiths University of London Departmental Ethics Committee. However, the final decision about participation is yours. Should you have any concerns or comments relating to your child’s participation in this study, please contact the **Departmental Ethics Committee** at the Psychology

Department, Goldsmiths University of London ***(*****+44 (0)20 7919 7870/7871; psychology@gold.ac.uk*****)***

I would appreciate if you would permit your child to participate in this project, as I believe it will significantly contribute towards improving the reliability and validity of psychological assessments and diagnoses in Trinidad and Tobago.

Please complete the attached permission form. Feel free to call me at ***(652-2792; 382-4945)*** or email me at ***kor_lou@hotmail.com.***

If you have any questions about the study, or if you would like additional information to assist you in reaching a decision, please feel free to call me or send me an email using the contact details in the paragraph above.

Thank you in advance for your interest and support of this project.

Yours sincerely,

Korinne Louison, M.A.,

PhD Student

Psychology Department

Goldsmiths University of London,

London, UK.

CONSENT FORM

Study: **Investigating the validity of the WISC-IV for Trinidad and Tobago**

If you and your child would like to be part of this study please sign this form.

When you sign the form, we also ask you to provide some information.

Please tick appropriate box:

Parent/Guardian

☐ **Yes, I would like my child to participate in this study.**

☐ **No, I do not want my child to participate in this study.**

Parent/Guardian: If Yes, please complete the following:

☐ I have read the Information Sheet about the study.

☐ I understand that my child does not have to take part in this study if I do not want to.

☐ I understand that my child can withdraw from the study at any time without giving a reason.

☐ I understand that my child will be asked for verbal consent by the researcher prior to any testing.

☐ I have had the opportunity to ask any questions I wish to ask.

☐ I have access to the names and telephone numbers of the research team in case I have any questions.

Will your child be available testing during the August holidays?

☐ **Yes No**

Child's Name: ____

Child's date of birth: ____

Child's School ____

Child'sClass____

Parent's/Guardian's Name: _____________________ Contact # _______

Parent's/Guardian's Signature: ________________________ Date: ____________

DEMOGRAPHIC QUESTIONNAIRE

Please answer all questions honestly. If you do not wish to answer a question, please draw a line through it.

The researcher may review the form to make sure you didn't mistakenly skip questions. Please feel to contact the researcher at the number listed below if you need any of the questions explained to you.

Q1. Name of Child: ____

Q2. Child's School ____

Q3. Child's Class
(eg. Std 5 H) ____

Q2. Child's Date of Birth:

____________	____________	____________
Day	Month	Year

Q3. Child's Gender: ☐ Male ☐ Female

Q4. Please specify your child's ethnicity (Tick one)

☐ Predominantly African descent
☐ Predominantly Amerindian descent (eg. Carib)
☐ Predominantly Chinese descent
☐ Predominantly East Indian descent
☐ Predominantly European descent
☐ Predominantly Middle Eastern descent (eg. Syrian)
☐ Mixed descent (please specify) ____________________
☐ Other (please specify) ____________________________

NATIONALITY AND LANGUAGE

Q5. Is your child a Trinidad and Tobago national? ☐ Yes ☐ No

a. If No, what is your child's nationality? ____

b. Is English your child's first language? ☐ Yes ☐ No

c. If No, what is your child's first Language? ____

d. What language is spoken in the home? ____

HOUSEHOLD COMPOSITION

Q6. How many persons live in your home? ______________________________

a. Please list names of parent(s) and or guardian(s).

1. ______________________________

Relationship to Child: ______________________________

2. ______________________________

Relationship to Child: ______________________________

EDUCATION OF PARENTS/GUARDIANS

Q7. What is the highest level of education completed?

Parent/Guardian 1

- ☐ Did not complete Primary School
- ☐ Completed Primary School
- ☐ Completed Secondary School O'Levels
- ☐ Completed Secondary School A'Levels
- ☐ Completed Associates Degree or Technical School
- ☐ Completed Bachelors
- ☐ Completed Masters__

☐ Completed PhD

Parent /Guardian 2

☐ Did not complete Primary School

☐ Completed Primary School

☐ Completed Secondary School O'Levels

☐ Completed Secondary School A'Levels

☐ Completed Associates Degree or Technical School

☐ Completed Bachelors

☐ Completed Masters__

☐ Completed PhD

Q8. What is the occupation of parents/guardian?

Parent/Guardian 1: ____

Parent/Guardian 2: ____

Q9. What is the combined household income per month?

☐ less than 4999TT

☐ 5000 – 9999TT

☐ 10000 – 19999TT

☐ 20000 – 29999 TT

☐ 30000 – 39999 TT

☐ 40000 – 49999 TT

☐ 50000 – 69999 TT

☐ 70000 – 89999 TT

☐ over 90000 TT

EXPOSURE TO AMERICAN CULTURE

Q10.	How exposed is your child to American culture via television, internet and travel?	☐ Very little (No access to television or internet. My child has never travelled to the United States)
		☐ Some (My child watches American television shows and has some idea about American culture music and events.)
		☐ Moderate (My child has access to television and the internet, has a moderate amount of knowledge about the American culture and has travelled to the United States on one or two occasions)
		☐ Very Much (My child has travelled to the United States on several occasions and is very knowledgeable about American culture through television, internet and also American family members and friends)

LEARNING DISABILITIES

Q11.	Has your child ever been diagnosed with a Learning, Speech or Behavioural Disorder that may have negatively impacted school performance?	☐ Yes	☐ No
	a. If Yes, what was the diagnosis?	☐ Reading Disorder (eg. Dylexia) ☐ Math Disorder ☐ Disorder of Written Expression ☐ Reading, Math and Writing Disorder ☐ Learning Disorder (Unspecified) ☐ Speech/Language Disorder ☐ ADHD ☐ Autism Spectrum Disorder ☐ Other (Please Specify): ____________________	

OTHER DIAGNOSES

Q12.	Has your child been diagnosed with any of the following psychological disorders? (Please tick all that apply).	☐ Clinical Depression ☐ Anxiety/Panic Disorder ☐ Psychotic Disorder/Schizophrenia ☐ Other (Please Specify): ____________________

Q13.	Has your child been diagnosed with a serious medical illness?	☐ Yes	☐ No
	a. If Yes, please specify.	______________________________	
Q14.	Has your child been diagnosed with a visual impairment?	☐ Yes	☐ No
	a. If Yes, please specify.	______________________________	
Q15.	Has your child been diagnosed with a hearing impairment?	☐ Yes	☐ No
	a. If Yes, please specify.	______________________________	

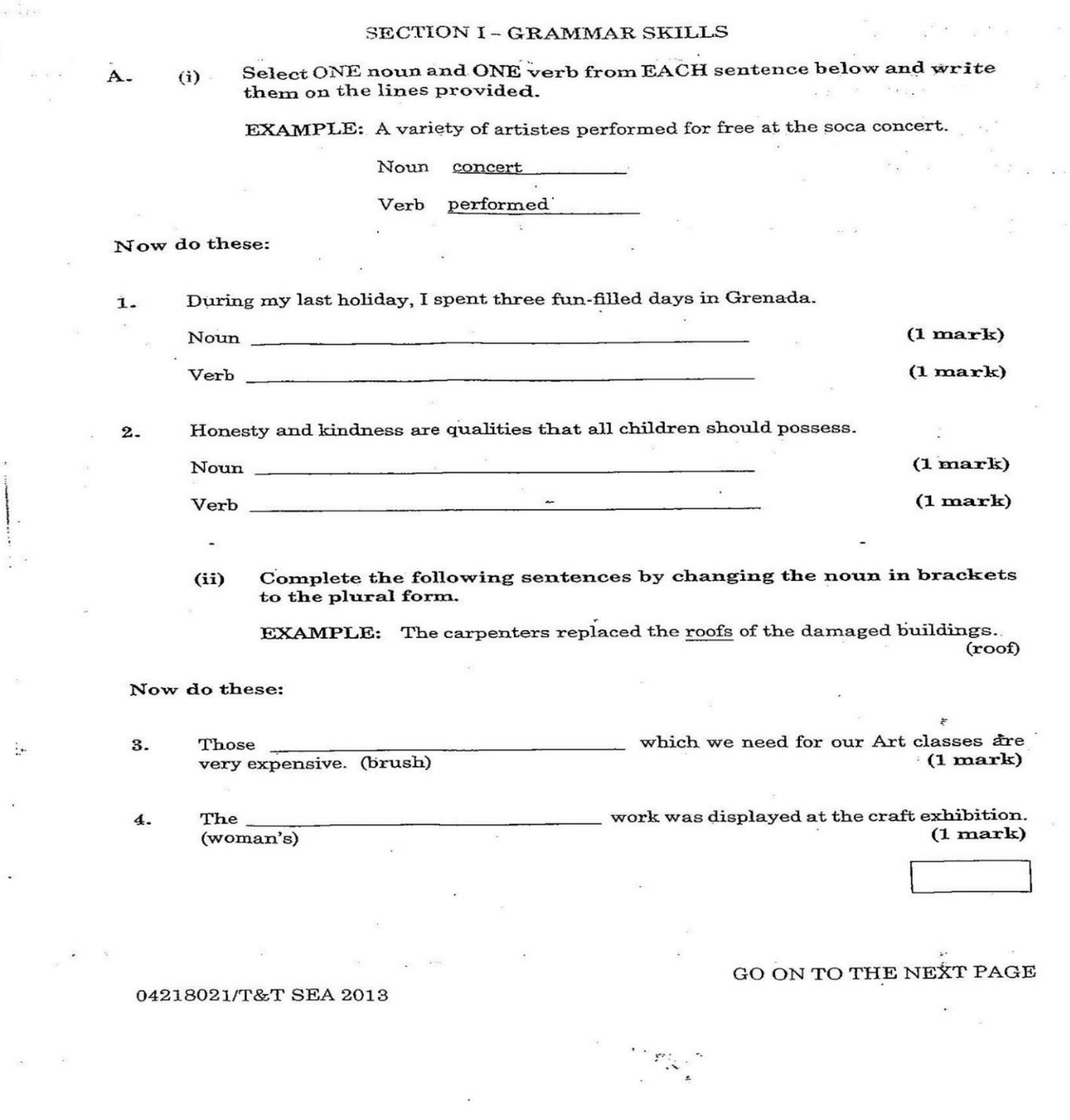
Q16. Does your child have a physical disability? ☐ Yes ☐ No

a. If Yes, please specify. ____________________

Thank you for participating in this study. If you have any questions please contact the lead researcher, Korinne Louison at the following numbers : (652-2792; 382-4945)

APPENDIX B

SEA LANGUAGE ARTS EXAMINATION (MOE, 2013)



SECTION I – GRAMMAR SKILLS

A. **(i)** **Select ONE noun and ONE verb from EACH sentence below and write them on the lines provided.**

EXAMPLE: A variety of artistes performed for free at the soca concert.

Noun concert

Verb performed

Now do these:

1. During my last holiday, I spent three fun-filled days in Grenada.

Noun ______ **(1 mark)**

Verb ______ **(1 mark)**

2. Honesty and kindness are qualities that all children should possess.

Noun ______ **(1 mark)**

Verb ______ **(1 mark)**

(ii) **Complete the following sentences by changing the noun in brackets to the plural form.**

EXAMPLE: The carpenters replaced the roofs of the damaged buildings. (roof)

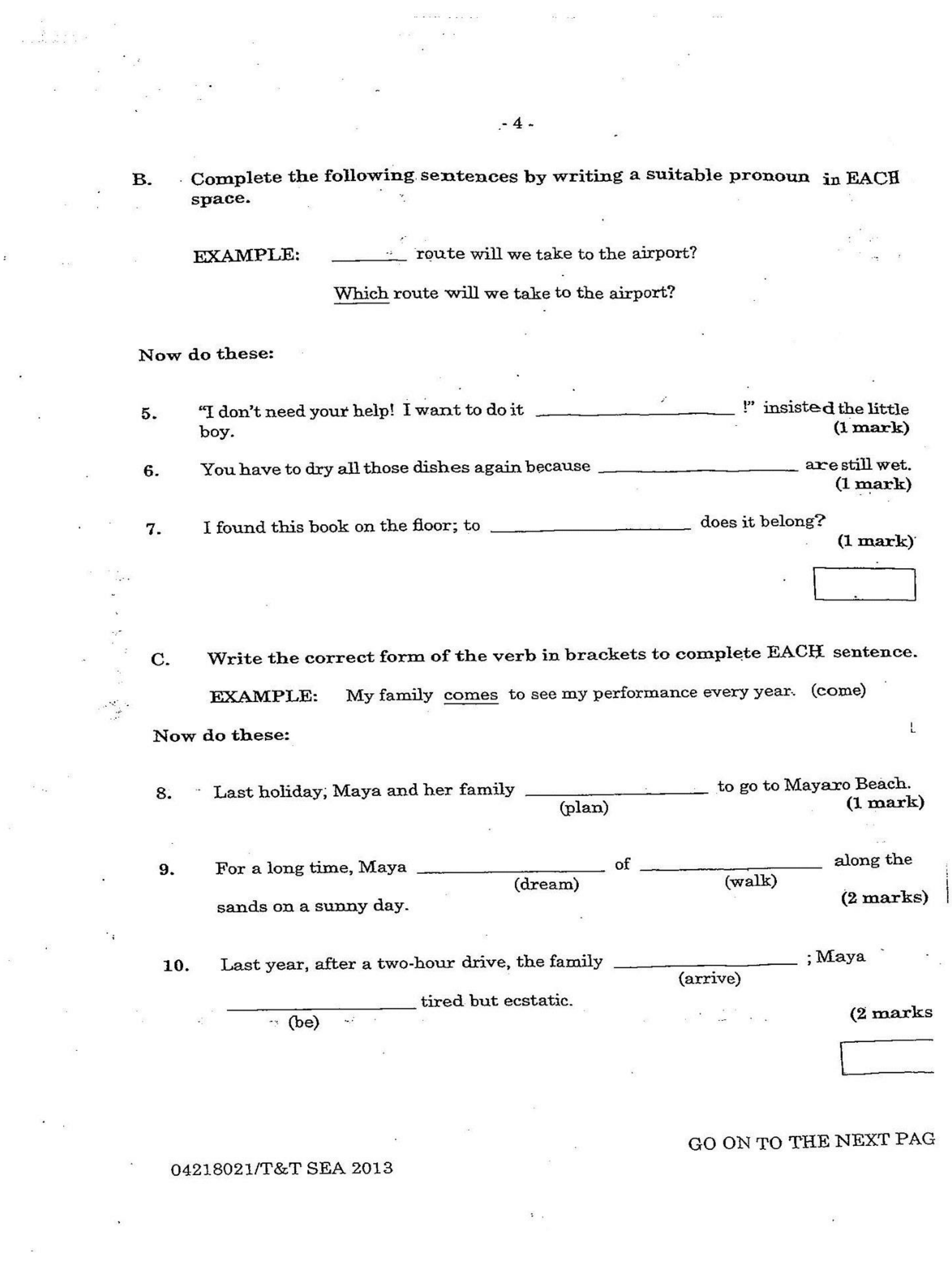
Now do these:

3. Those ______ which we need for our Art classes are very expensive. (brush) **(1 mark)**

4. The ______ work was displayed at the craft exhibition. (woman's) **(1 mark)**

GO ON TO THE NEXT PAGE

04218021/T&T SEA 2013



B. **Complete the following sentences by writing a suitable pronoun in EACH space.**

EXAMPLE: ________ route will we take to the airport?

Which route will we take to the airport?

Now do these:

5. "I don't need your help! I want to do it ________________ !" insisted the little boy. **(1 mark)**

6. You have to dry all those dishes again because ________________ are still wet. **(1 mark)**

7. I found this book on the floor; to ________________ does it belong? **(1 mark)**

C. **Write the correct form of the verb in brackets to complete EACH sentence.**

EXAMPLE: My family comes to see my performance every year. (come)

Now do these:

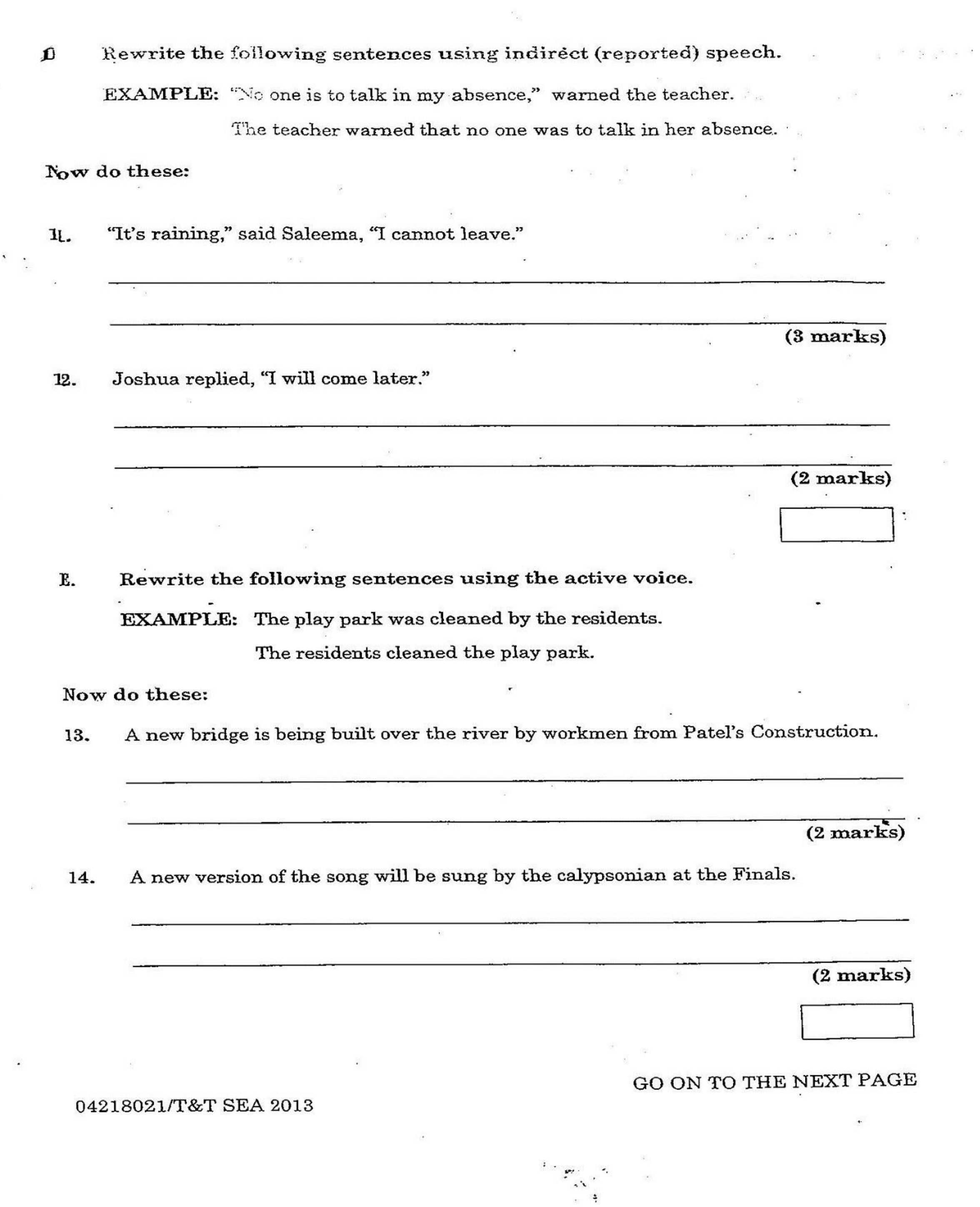
8. Last holiday, Maya and her family ________________ (plan) to go to Mayaro Beach. **(1 mark)**

9. For a long time, Maya ________________ (dream) of ________________ (walk) along the sands on a sunny day. **(2 marks)**

10. Last year, after a two-hour drive, the family ________________ (arrive); Maya ________________ (be) tired but ecstatic. **(2 marks**

GO ON TO THE NEXT PAG

04218021/T&T SEA 2013



D **Rewrite the following sentences using indirect (reported) speech.**

EXAMPLE: "No one is to talk in my absence," warned the teacher.
The teacher warned that no one was to talk in her absence.

Now do these:

11. "It's raining," said Saleema, "I cannot leave."

____

____

(3 marks)

12. Joshua replied, "I will come later."

____

____

(2 marks)

E. **Rewrite the following sentences using the active voice.**

EXAMPLE: The play park was cleaned by the residents.
The residents cleaned the play park.

Now do these:

13. A new bridge is being built over the river by workmen from Patel's Construction.

____

____

(2 marks)

14. A new version of the song will be sung by the calypsonian at the Finals.

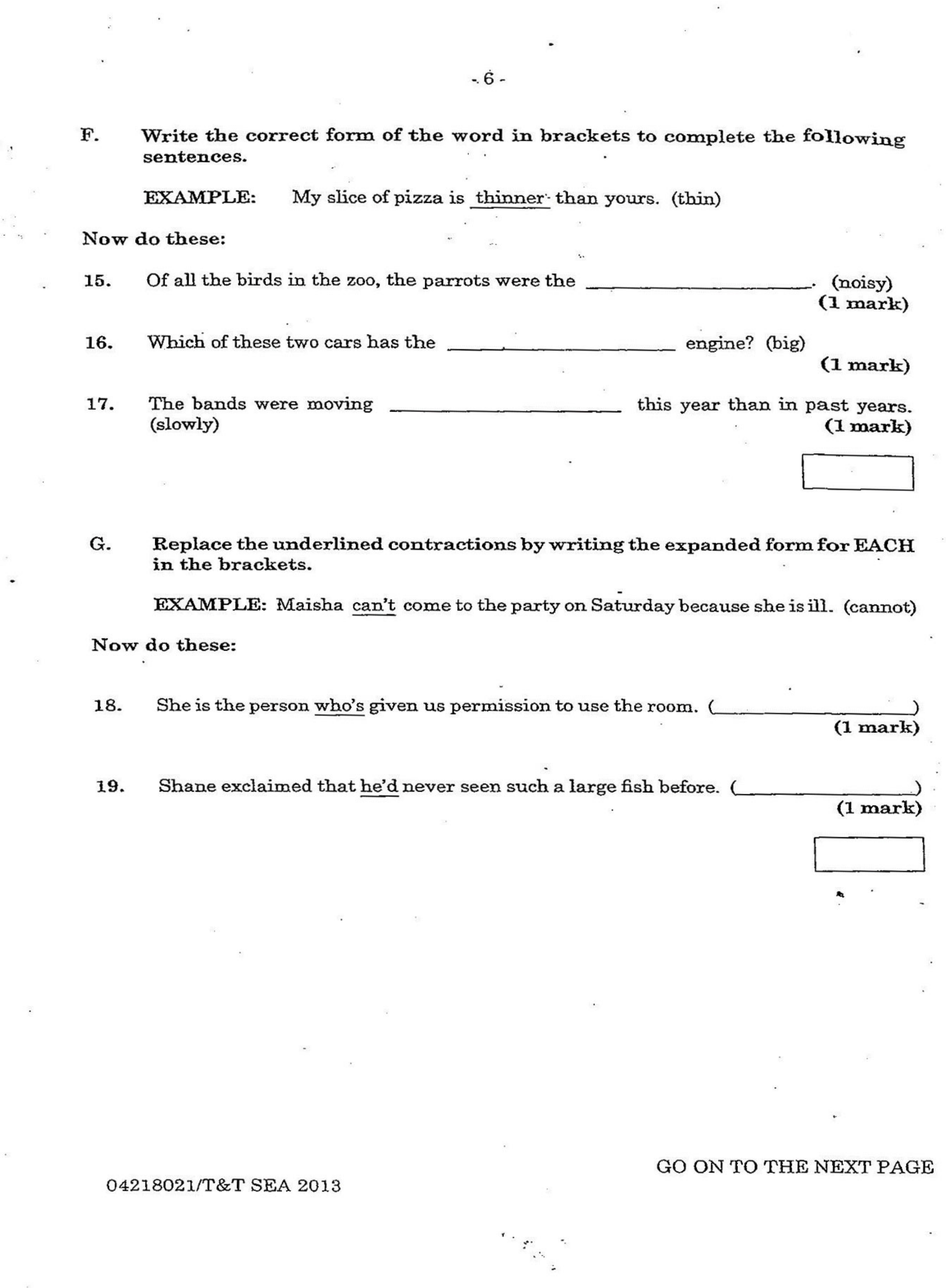
____

____

(2 marks)

GO ON TO THE NEXT PAGE

04218021/T&T SEA 2013



- 6 -

F. Write the correct form of the word in brackets to complete the following sentences.

EXAMPLE: My slice of pizza is thinner than yours. (thin)

Now do these:

15. Of all the birds in the zoo, the parrots were the ____________________. (noisy) **(1 mark)**

16. Which of these two cars has the ____________________ engine? (big) **(1 mark)**

17. The bands were moving ____________________ this year than in past years. (slowly) **(1 mark)**

G. Replace the underlined contractions by writing the expanded form for EACH in the brackets.

EXAMPLE: Maisha can't come to the party on Saturday because she is ill. (cannot)

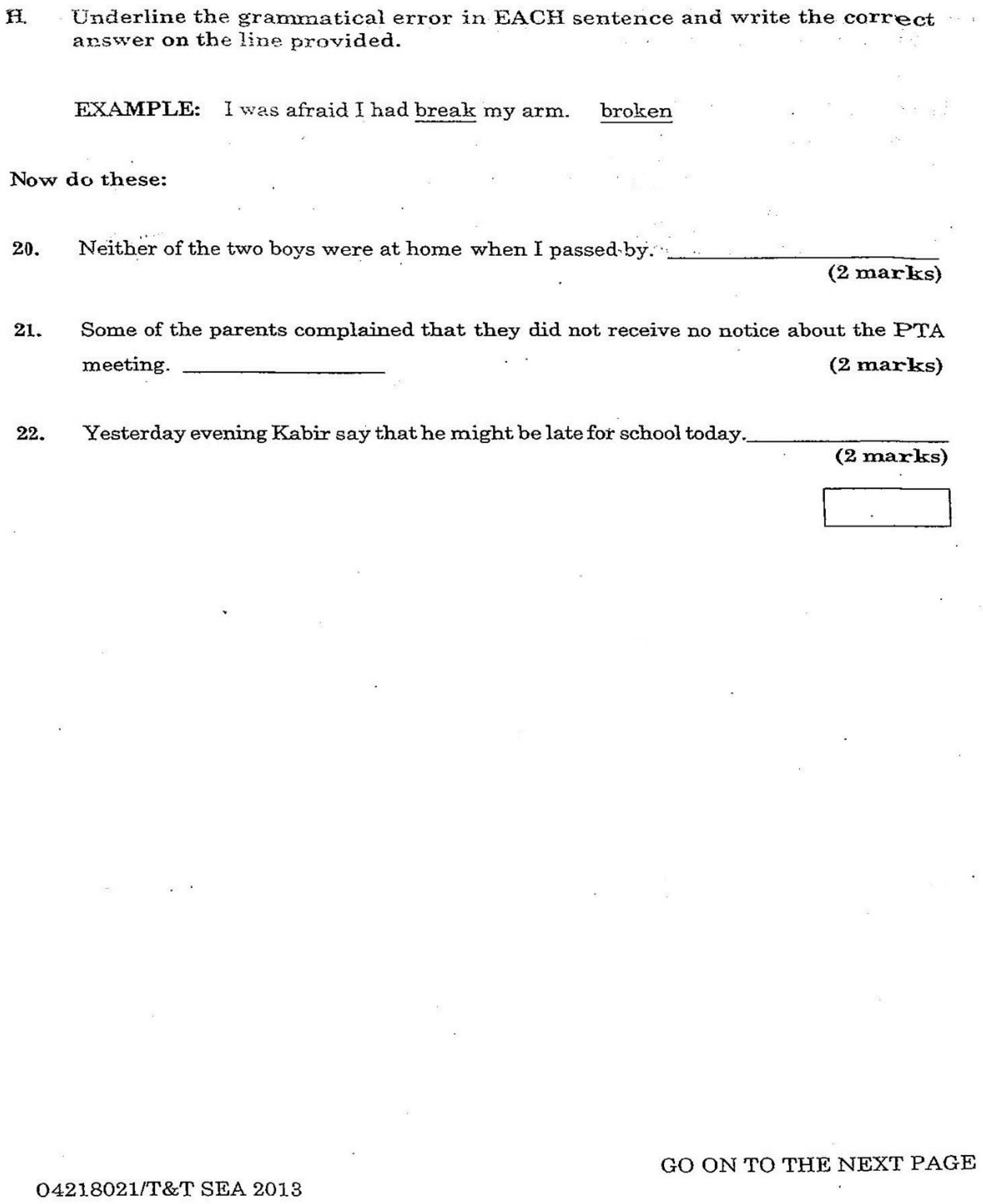
Now do these:

18. She is the person who's given us permission to use the room. (____________________) **(1 mark)**

19. Shane exclaimed that he'd never seen such a large fish before. (____________________) **(1 mark)**

GO ON TO THE NEXT PAGE

04218021/T&T SEA 2013



H. **Underline the grammatical error in EACH sentence and write the correct answer on the line provided.**

EXAMPLE: I was afraid I had break my arm. broken

Now do these:

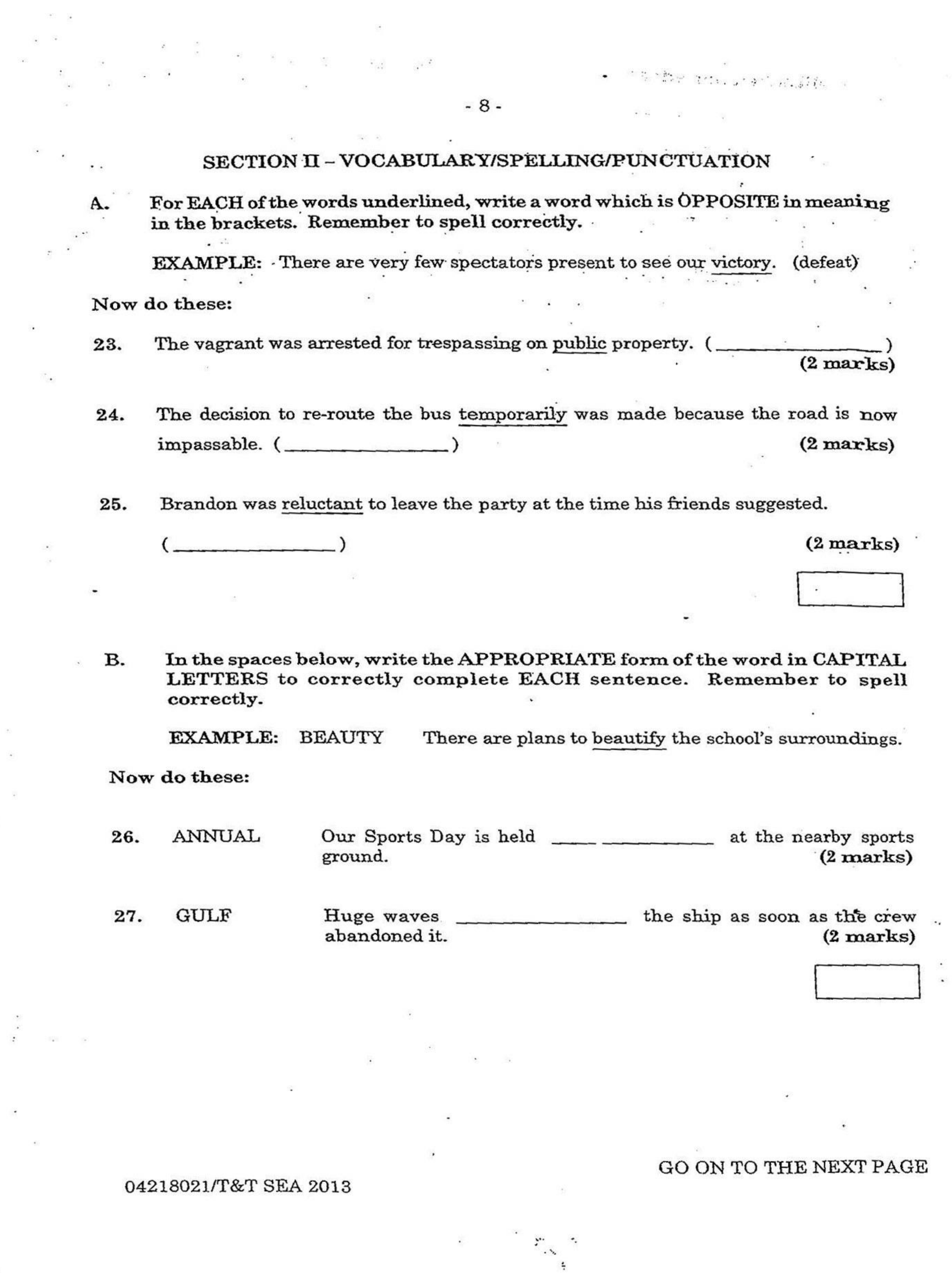
20. Neither of the two boys were at home when I passed by. ____________________
(2 marks)

21. Some of the parents complained that they did not receive no notice about the PTA meeting. ________________ **(2 marks)**

22. Yesterday evening Kabir say that he might be late for school today.______________
(2 marks)

GO ON TO THE NEXT PAGE

04218021/T&T SEA 2013



SECTION II – VOCABULARY/SPELLING/PUNCTUATION

A. **For EACH of the words underlined, write a word which is OPPOSITE in meaning in the brackets. Remember to spell correctly.**

EXAMPLE: There are very few spectators present to see our victory. (defeat)

Now do these:

23. The vagrant was arrested for trespassing on public property. (________________) **(2 marks)**

24. The decision to re-route the bus temporarily was made because the road is now impassable. (________________) **(2 marks)**

25. Brandon was reluctant to leave the party at the time his friends suggested.

(________________) **(2 marks)**

B. **In the spaces below, write the APPROPRIATE form of the word in CAPITAL LETTERS to correctly complete EACH sentence. Remember to spell correctly.**

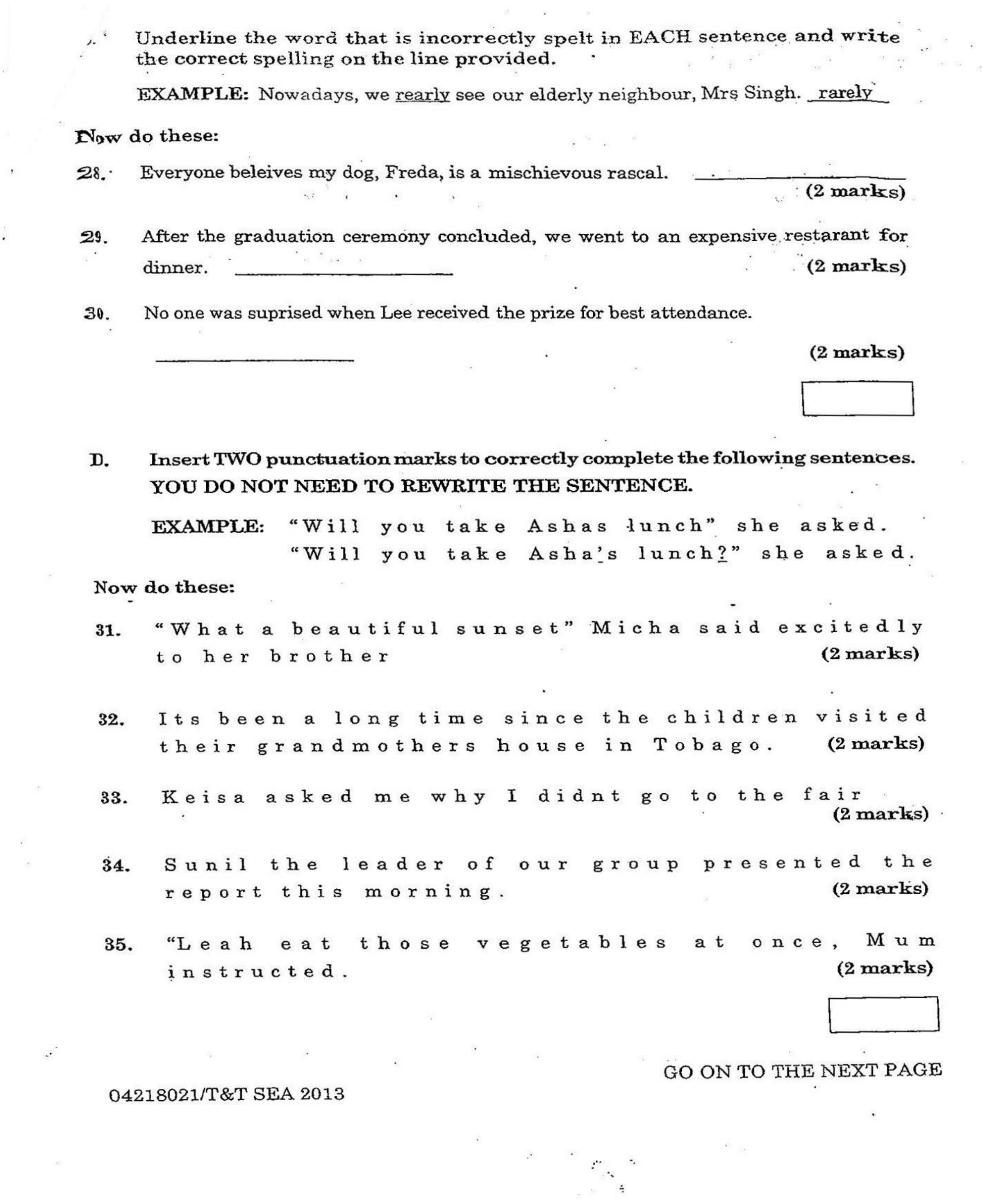
EXAMPLE: BEAUTY There are plans to beautify the school's surroundings.

Now do these:

26. ANNUAL Our Sports Day is held ________________ at the nearby sports ground. **(2 marks)**

27. GULF Huge waves ________________ the ship as soon as the crew abandoned it. **(2 marks)**

GO ON TO THE NEXT PAGE



Underline the word that is incorrectly spelt in EACH sentence and write the correct spelling on the line provided.

EXAMPLE: Nowadays, we rearly see our elderly neighbour, Mrs Singh. rarely

Now do these:

28. Everyone beleives my dog, Freda, is a mischievous rascal. ____________________ **(2 marks)**

29. After the graduation ceremony concluded, we went to an expensive restarant for dinner. ____________________ **(2 marks)**

30. No one was suprised when Lee received the prize for best attendance.

____________________ **(2 marks)**

D. Insert TWO punctuation marks to correctly complete the following sentences.
YOU DO NOT NEED TO REWRITE THE SENTENCE.

EXAMPLE: "Will you take Ashas lunch" she asked.
"Will you take Asha's lunch?" she asked.

Now do these:

31. "What a beautiful sunset" Micha said excitedly to her brother **(2 marks)**

32. Its been a long time since the children visited their grandmothers house in Tobago. **(2 marks)**

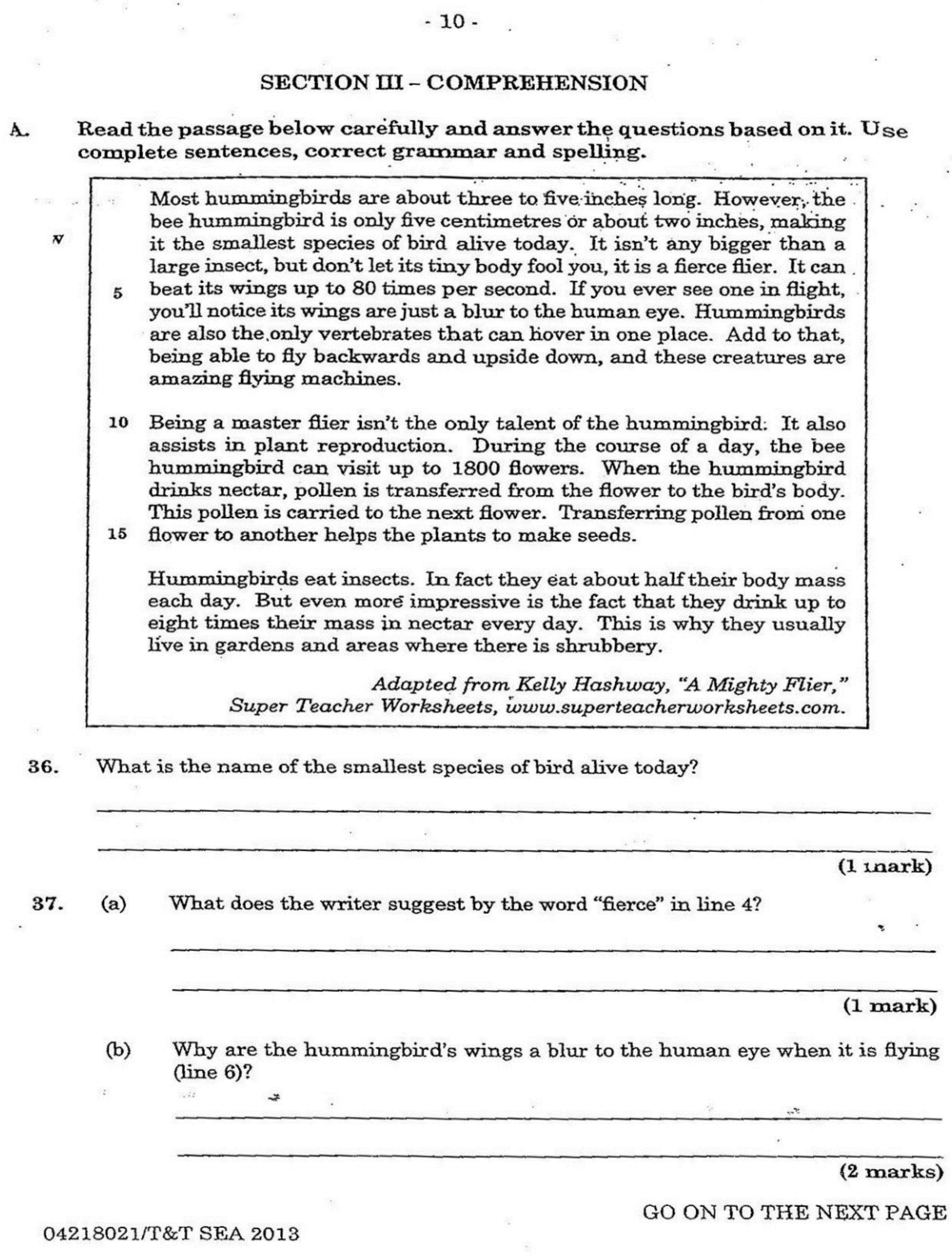
33. Keisa asked me why I didnt go to the fair **(2 marks)**

34. Sunil the leader of our group presented the report this morning. **(2 marks)**

35. "Leah eat those vegetables at once, Mum instructed. **(2 marks)**

GO ON TO THE NEXT PAGE

04218021/T&T SEA 2013



SECTION III – COMPREHENSION

A. Read the passage below carefully and answer the questions based on it. Use complete sentences, correct grammar and spelling.

> Most hummingbirds are about three to five inches long. However, the bee hummingbird is only five centimetres or about two inches, making it the smallest species of bird alive today. It isn't any bigger than a large insect, but don't let its tiny body fool you, it is a fierce flier. It can beat its wings up to 80 times per second. If you ever see one in flight, you'll notice its wings are just a blur to the human eye. Hummingbirds are also the only vertebrates that can hover in one place. Add to that, being able to fly backwards and upside down, and these creatures are amazing flying machines.
>
> Being a master flier isn't the only talent of the hummingbird. It also assists in plant reproduction. During the course of a day, the bee hummingbird can visit up to 1800 flowers. When the hummingbird drinks nectar, pollen is transferred from the flower to the bird's body. This pollen is carried to the next flower. Transferring pollen from one flower to another helps the plants to make seeds.
>
> Hummingbirds eat insects. In fact they eat about half their body mass each day. But even more impressive is the fact that they drink up to eight times their mass in nectar every day. This is why they usually live in gardens and areas where there is shrubbery.
>
> *Adapted from Kelly Hashway, "A Mighty Flier,"*
> *Super Teacher Worksheets, www.superteacherworksheets.com.*

36. What is the name of the smallest species of bird alive today?

(1 mark)

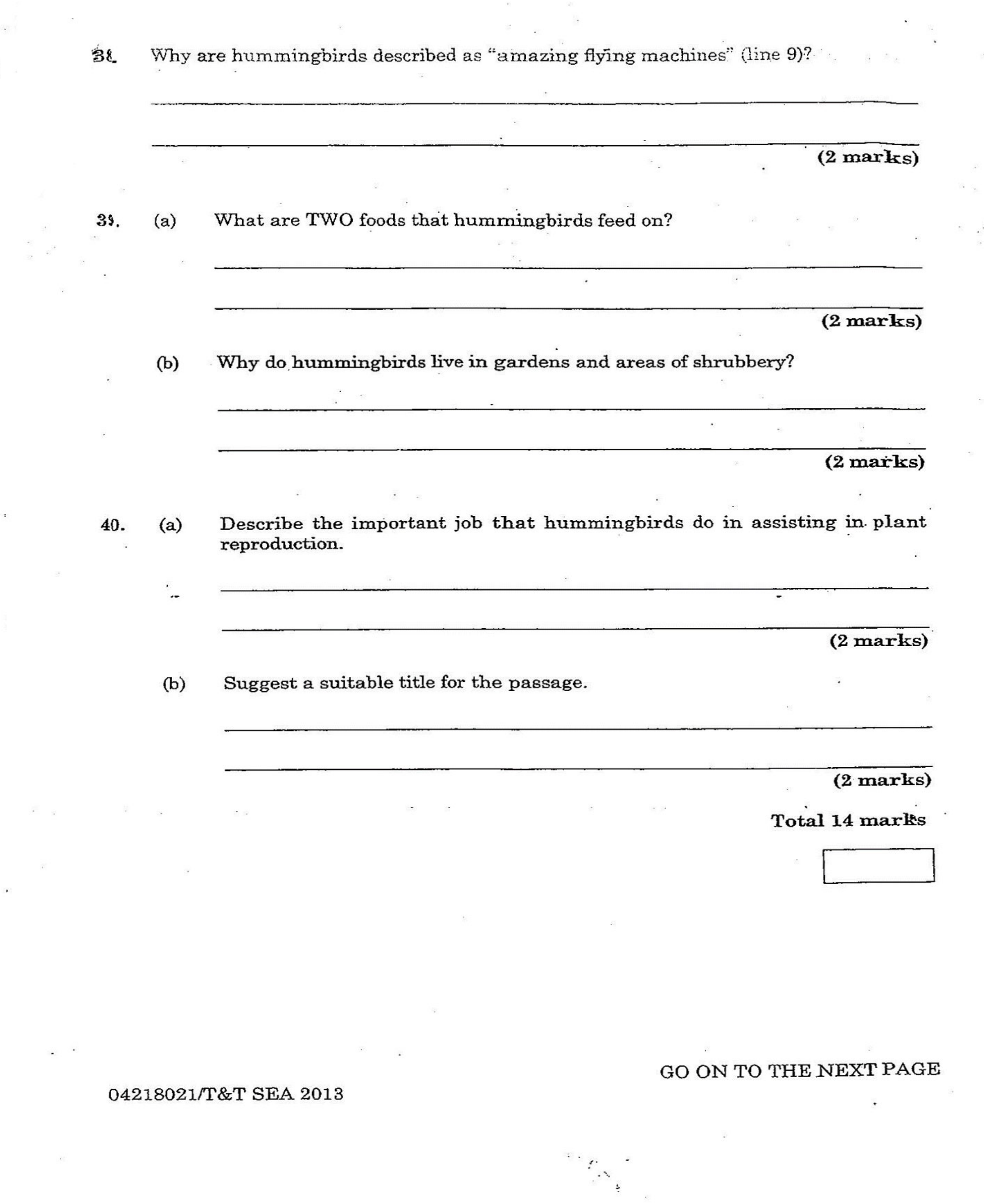
37. (a) What does the writer suggest by the word "fierce" in line 4?

(1 mark)

(b) Why are the hummingbird's wings a blur to the human eye when it is flying (line 6)?

(2 marks)

GO ON TO THE NEXT PAGE



38. Why are hummingbirds described as "amazing flying machines" (line 9)?

____

____

(2 marks)

39. (a) What are TWO foods that hummingbirds feed on?

____

____

(2 marks)

(b) Why do hummingbirds live in gardens and areas of shrubbery?

____

____

(2 marks)

40. **(a)** **Describe the important job that hummingbirds do in assisting in plant reproduction.**

____

____

(2 marks)

(b) Suggest a suitable title for the passage.

____

____

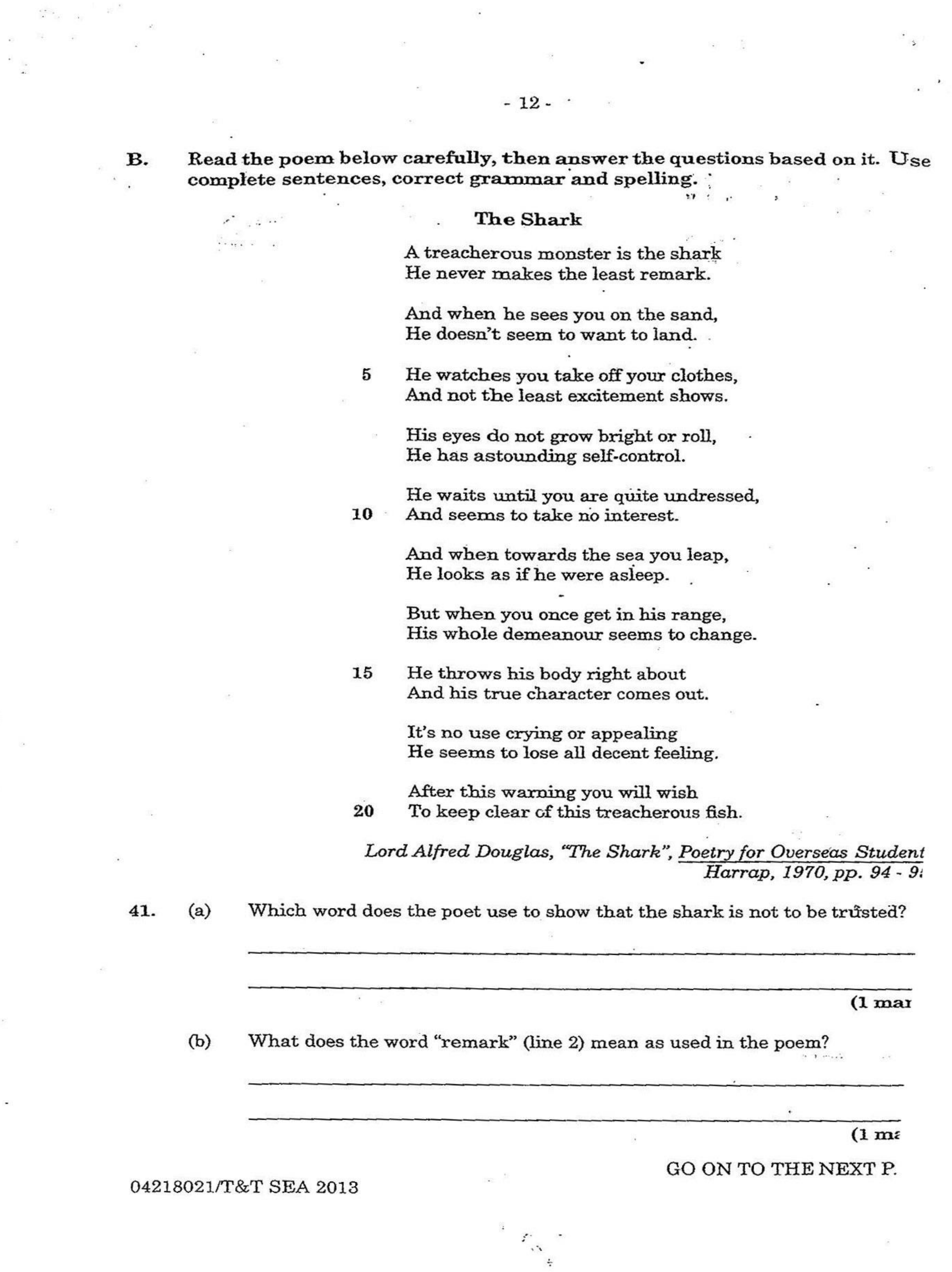
(2 marks)

Total 14 marks

☐

GO ON TO THE NEXT PAGE

04218021/T&T SEA 2013



- 12 -

B. **Read the poem below carefully, then answer the questions based on it. Use complete sentences, correct grammar and spelling.**

The Shark

A treacherous monster is the shark
He never makes the least remark.

And when he sees you on the sand,
He doesn't seem to want to land.

He watches you take off your clothes,
And not the least excitement shows.

His eyes do not grow bright or roll,
He has astounding self-control.

He waits until you are quite undressed,
And seems to take no interest.

And when towards the sea you leap,
He looks as if he were asleep.

But when you once get in his range,
His whole demeanour seems to change.

He throws his body right about
And his true character comes out.

It's no use crying or appealing
He seems to lose all decent feeling.

After this warning you will wish
To keep clear of this treacherous fish.

Lord Alfred Douglas, "The Shark", Poetry for Overseas Student
Harrap, 1970, pp. 94 - 9

41. (a) Which word does the poet use to show that the shark is not to be trusted?

____

____

(1 mai

(b) What does the word "remark" (line 2) mean as used in the poem?

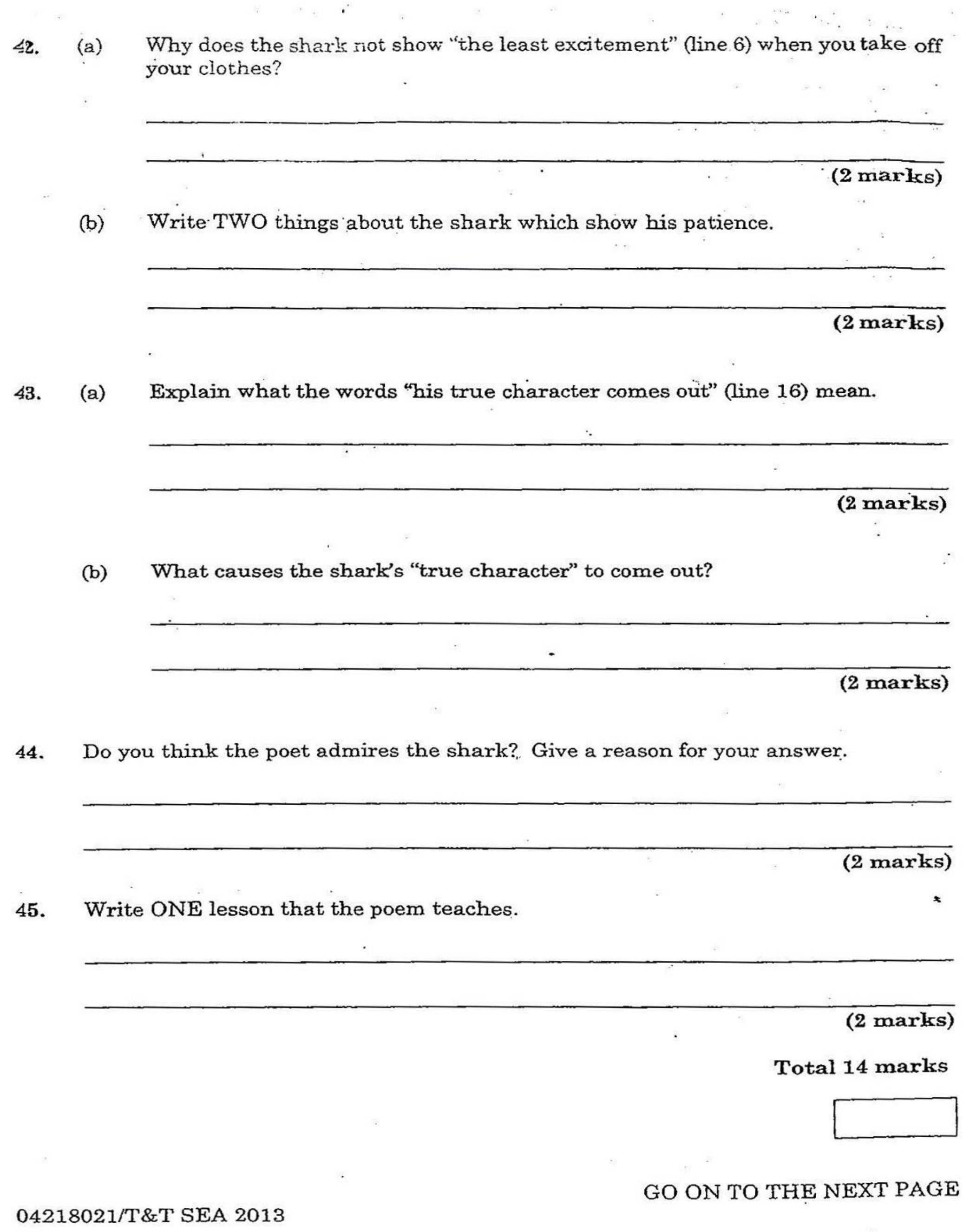
____

____

(1 ma

GO ON TO THE NEXT P.

04218021/T&T SEA 2013



42. (a) Why does the shark not show "the least excitement" (line 6) when you take off your clothes?

____

____

(2 marks)

(b) Write TWO things about the shark which show his patience.

____

____

(2 marks)

43. (a) Explain what the words "his true character comes out" (line 16) mean.

____

____

(2 marks)

(b) What causes the shark's "true character" to come out?

____

____

(2 marks)

44. Do you think the poet admires the shark? Give a reason for your answer.

____

____

(2 marks)

45. Write ONE lesson that the poem teaches.

____

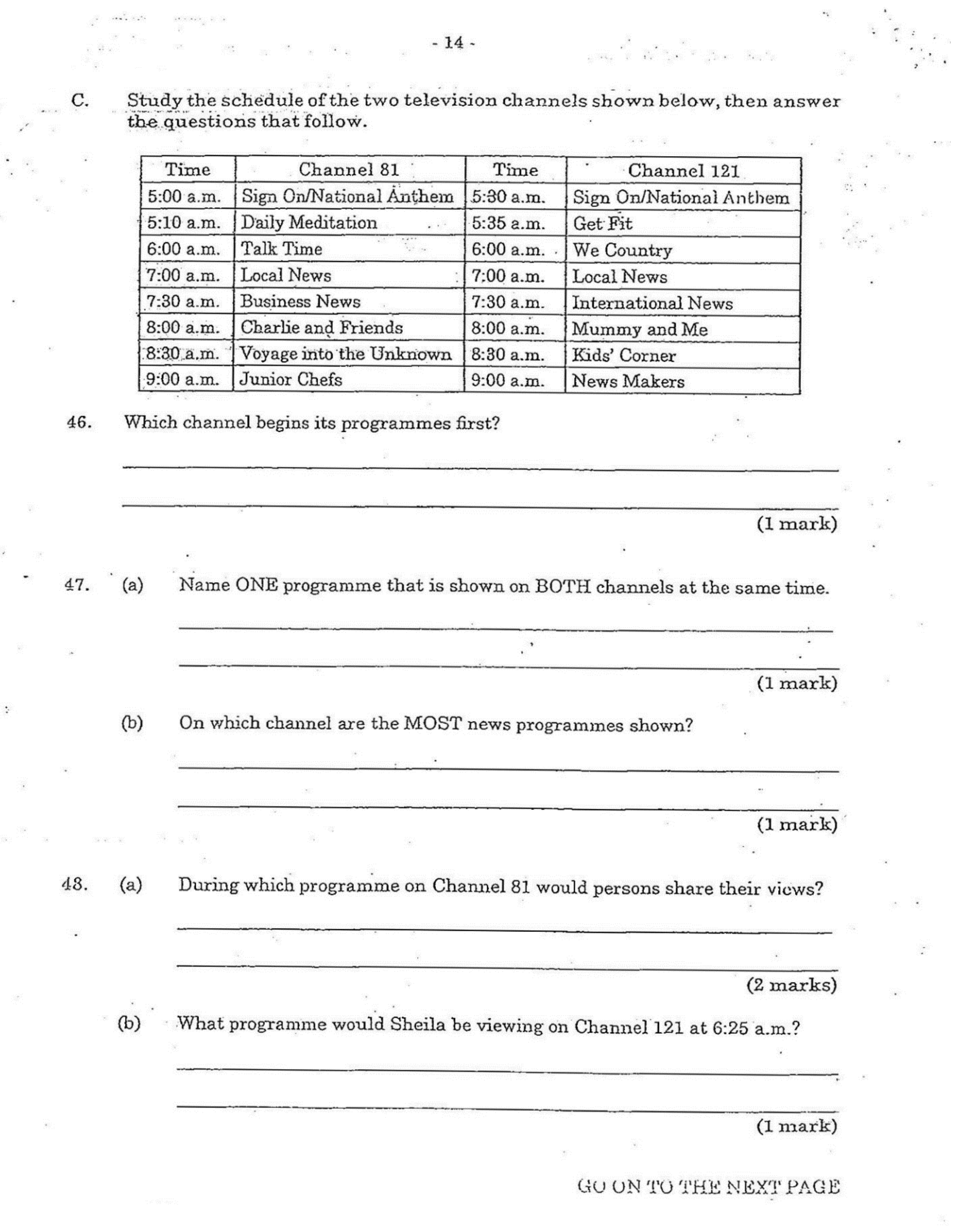
____

(2 marks)

Total 14 marks

GO ON TO THE NEXT PAGE

04218021/T&T SEA 2013



- 14 -

C. Study the schedule of the two television channels shown below, then answer the questions that follow.

Time	Channel 81	Time	Channel 121
5:00 a.m.	Sign On/National Anthem	5:30 a.m.	Sign On/National Anthem
5:10 a.m.	Daily Meditation	5:35 a.m.	Get Fit
6:00 a.m.	Talk Time	6:00 a.m.	We Country
7:00 a.m.	Local News	7:00 a.m.	Local News
7:30 a.m.	Business News	7:30 a.m.	International News
8:00 a.m.	Charlie and Friends	8:00 a.m.	Mummy and Me
8:30 a.m.	Voyage into the Unknown	8:30 a.m.	Kids' Corner
9:00 a.m.	Junior Chefs	9:00 a.m.	News Makers

46. Which channel begins its programmes first?

(1 mark)

47. (a) Name ONE programme that is shown on BOTH channels at the same time.

(1 mark)

(b) On which channel are the MOST news programmes shown?

(1 mark)

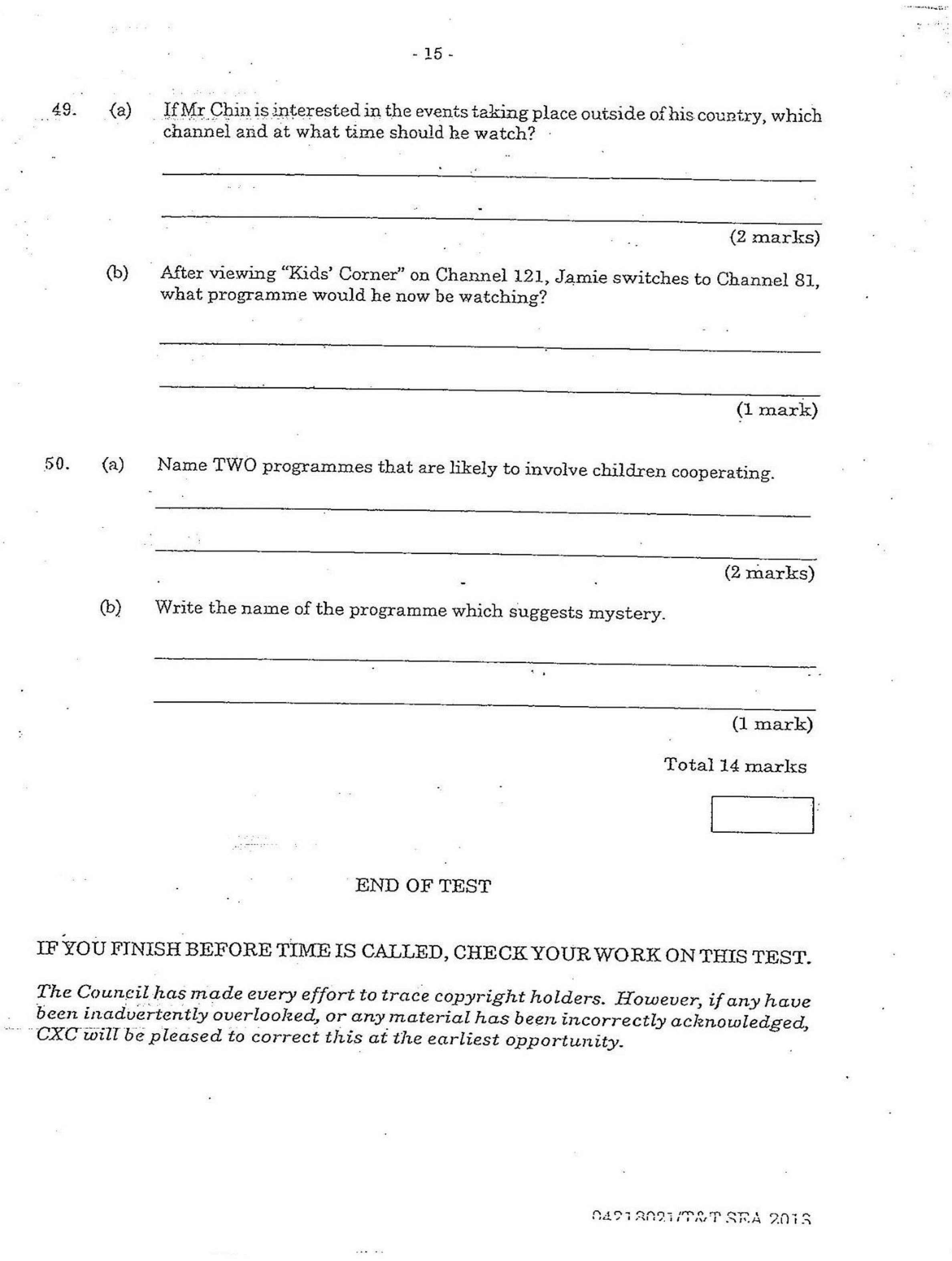
48. (a) During which programme on Channel 81 would persons share their views?

(2 marks)

(b) What programme would Sheila be viewing on Channel 121 at 6:25 a.m.?

(1 mark)

GO ON TO THE NEXT PAGE



49. (a) If Mr Chin is interested in the events taking place outside of his country, which channel and at what time should he watch?

(2 marks)

(b) After viewing "Kids' Corner" on Channel 121, Jamie switches to Channel 81, what programme would he now be watching?

(1 mark)

50. (a) Name TWO programmes that are likely to involve children cooperating.

(2 marks)

(b) Write the name of the programme which suggests mystery.

(1 mark)

Total 14 marks

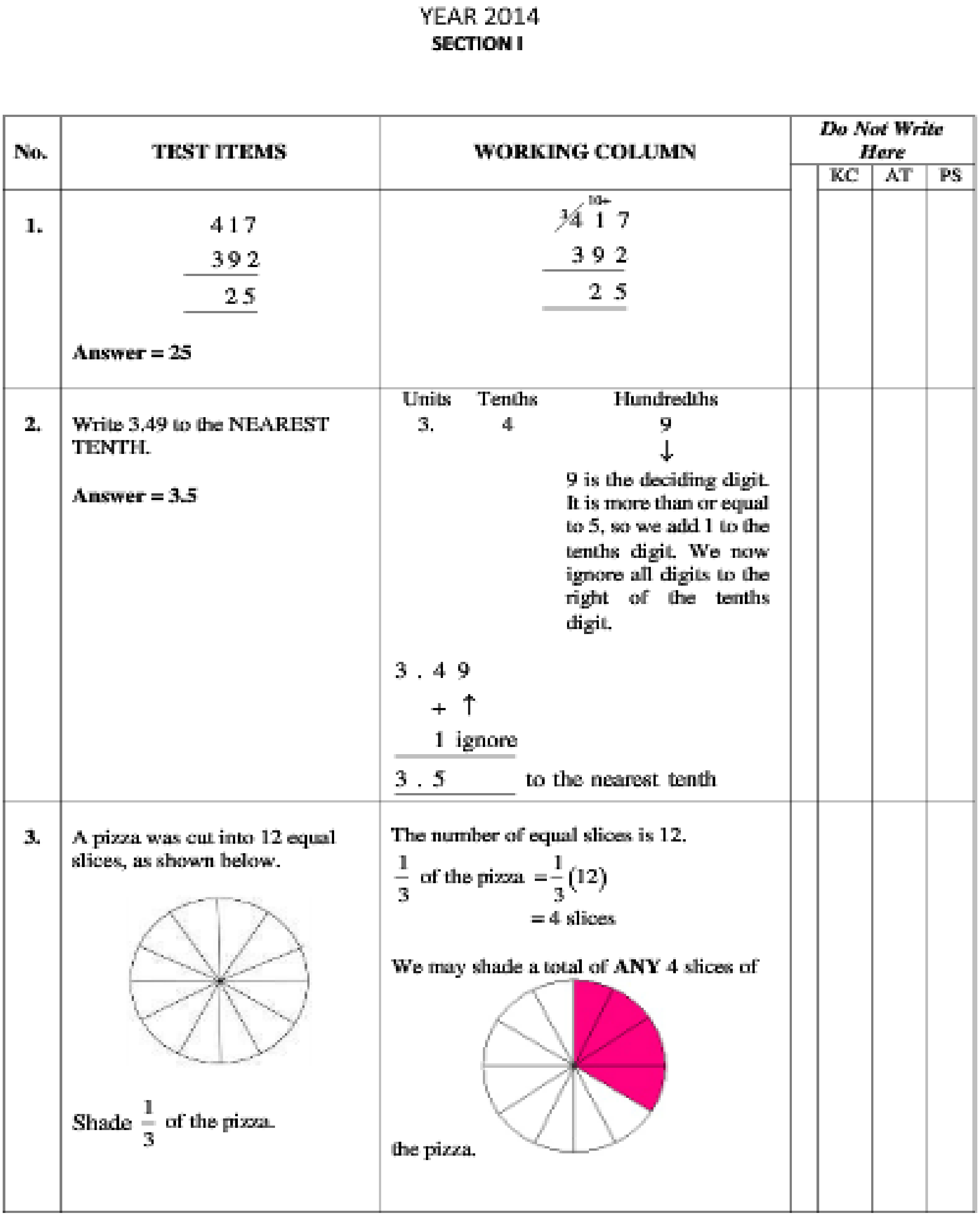
END OF TEST

IF YOU FINISH BEFORE TIME IS CALLED, CHECK YOUR WORK ON THIS TEST.

The Council has made every effort to trace copyright holders. However, if any have been inadvertently overlooked, or any material has been incorrectly acknowledged, CXC will be pleased to correct this at the earliest opportunity.

APPENDIX C

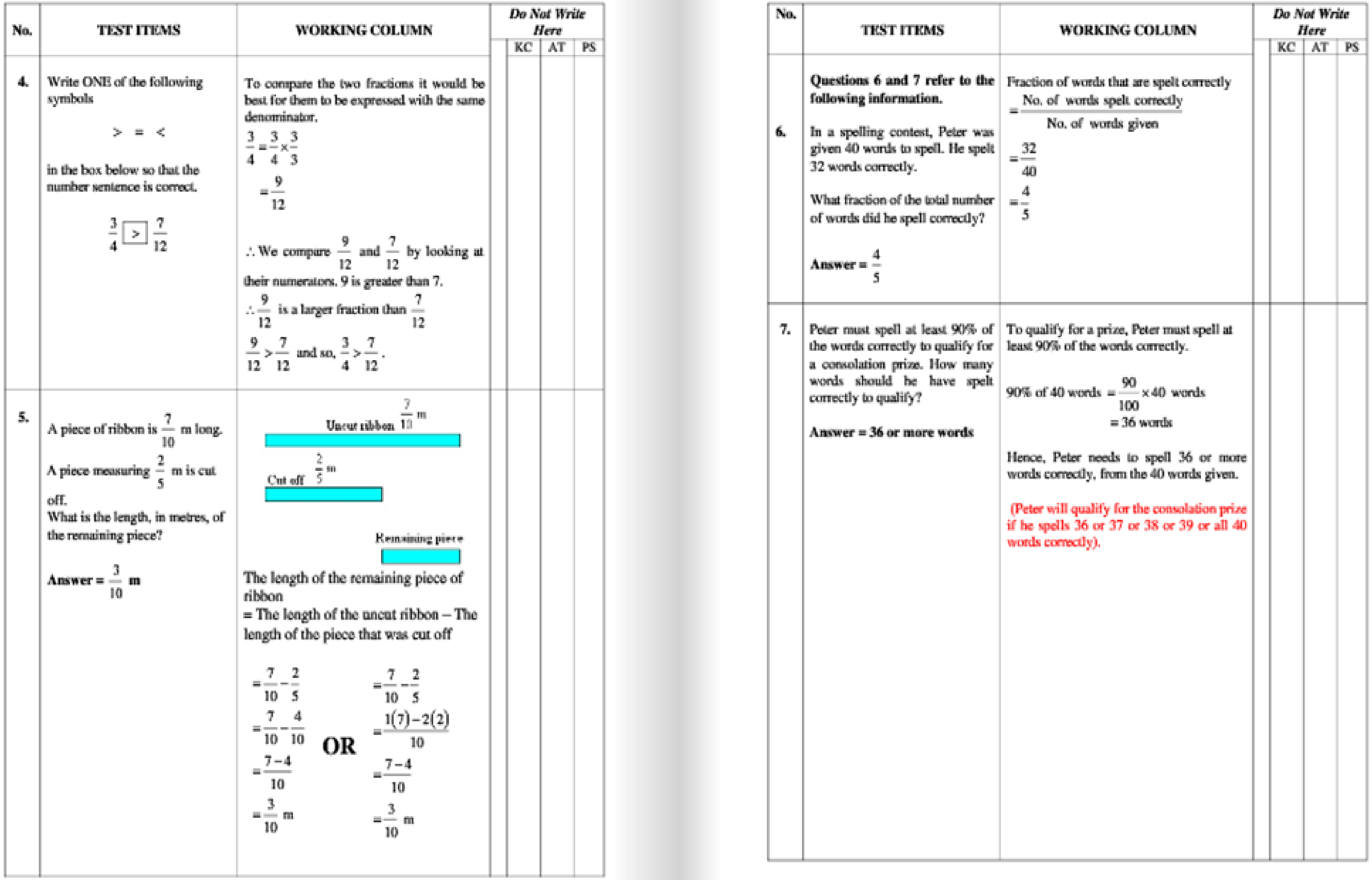
SEA Mathematics Examination (MOE, 2014)



YEAR 2014

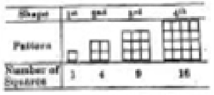
SECTION I

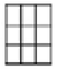
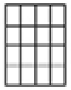
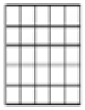
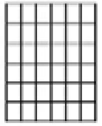
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*			
				KC	AT	PS
1.	417 392 25 **Answer = 25**	$\overset{3}{\cancel{4}}\ \overset{10+}{1}\ 7$ 3 9 2 2 5				
2.	Write 3.49 to the NEAREST TENTH. **Answer = 3.5**	Units Tenths Hundredths 3. 4 9 ↓ 9 is the deciding digit. It is more than or equal to 5, so we add 1 to the tenths digit. We now ignore all digits to the right of the tenths digit. 3 . 4 9 + ↑ 1 ignore 3 . 5 to the nearest tenth				
3.	A pizza was cut into 12 equal slices, as shown below. Shade $\frac{1}{3}$ of the pizza.	The number of equal slices is 12. $\frac{1}{3}$ of the pizza $= \frac{1}{3}(12)$ $= 4$ slices We may shade a total of **ANY** 4 slices of the pizza.				



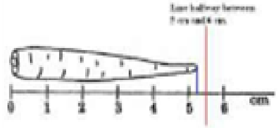
No.	TEST ITEMS	WORKING COLUMN	Do Not Write Here		
			KC	AT	PS
4.	Write ONE of the following symbols > = < in the box below so that the number sentence is correct. $\frac{3}{4} \boxed{>} \frac{7}{12}$	To compare the two fractions it would be best for them to be expressed with the same denominator. $\frac{3}{4} = \frac{3}{4} \times \frac{3}{3}$ $= \frac{9}{12}$ $\therefore$ We compare $\frac{9}{12}$ and $\frac{7}{12}$ by looking at their numerators. 9 is greater than 7. $\therefore \frac{9}{12}$ is a larger fraction than $\frac{7}{12}$ $\frac{9}{12} > \frac{7}{12}$ and so, $\frac{3}{4} > \frac{7}{12}$.			
5.	A piece of ribbon is $\frac{7}{10}$ m long. A piece measuring $\frac{2}{5}$ m is cut off. What is the length, in metres, of the remaining piece? **Answer =** $\frac{3}{10}$ **m**	 The length of the remaining piece of ribbon = The length of the uncut ribbon – The length of the piece that was cut off $= \frac{7}{10} - \frac{2}{5}$ $= \frac{7}{10} - \frac{4}{10}$ $= \frac{7-4}{10}$ $= \frac{3}{10}$ m **OR** $= \frac{7}{10} - \frac{2}{5}$ $= \frac{1(7)-2(2)}{10}$ $= \frac{7-4}{10}$ $= \frac{3}{10}$ m			

No.	TEST ITEMS	WORKING COLUMN	Do Not Write Here		
			KC	AT	PS
6.	**Questions 6 and 7 refer to the following information.** In a spelling contest, Peter was given 40 words to spell. He spelt 32 words correctly. What fraction of the total number of words did he spell correctly? **Answer =** $\frac{4}{5}$	Fraction of words that are spelt correctly $= \frac{\text{No. of words spelt correctly}}{\text{No. of words given}}$ $= \frac{32}{40}$ $= \frac{4}{5}$			
7.	Peter must spell at least 90% of the words correctly to qualify for a consolation prize. How many words should he have spelt correctly to qualify? **Answer = 36 or more words**	To qualify for a prize, Peter must spell at least 90% of the words correctly. 90% of 40 words $= \frac{90}{100} \times 40$ words = 36 words Hence, Peter needs to spell 36 or more words correctly, from the 40 words given. (Peter will qualify for the consolation prize if he spells 36 or 37 or 38 or 39 or all 40 words correctly).			

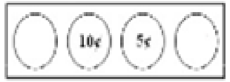
<table>
<tr><th rowspan="2">No.</th><th rowspan="2">TEST ITEMS</th><th rowspan="2">WORKING COLUMN</th><th colspan="3">Do Not Write Here</th></tr>
<tr><th>KC</th><th>AT</th><th>PS</th></tr>
<tr><td>8.</td><td>The first four shapes in a pattern are shown below. EACH shape is made from squares of the same size.


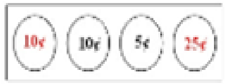
How many squares would form the 6th shape?
Answer = 36 squares</td><td><table>
<tr><th>Shape</th><th>Pattern</th><th>No. of Squares</th></tr>
<tr><td>1st</td><td></td><td>1
(1×1)</td></tr>
<tr><td>2nd</td><td></td><td>4
(2×2)</td></tr>
<tr><td>3rd</td><td></td><td>9
(3×3)</td></tr>
<tr><td>4th</td><td></td><td>16
(4×4)</td></tr>
<tr><td>5th</td><td></td><td>25
(5×5)</td></tr>
<tr><td>6th</td><td></td><td>36
(6×6)</td></tr>
</table>
Therefore, in the 6th shape we can expect 36 squares.</td><td></td><td></td><td></td></tr>
<tr><td>9.</td><td>Express 1 litre in cm^3
1 litre = 1000 cm^3</td><td>1 litre = 1000 ml
1 ml = 1 cm^3 (or 1 c.c)
$\therefore$ 1 litre = 1000 cm^3</td><td></td><td></td><td></td></tr>
<tr><td>10.</td><td>Express 120 minutes in hours
120 minutes = 2 hours</td><td>1 hour = 60 minutes
60 minutes = 1 hour
$1 \text{ minute} = \frac{1}{60} \text{ hour}$
$120 \text{ minutes} = \frac{1}{60}\times120 \text{ hours}$
$= 2 \text{ hours}$</td><td></td><td></td><td></td></tr>
</table>

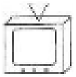
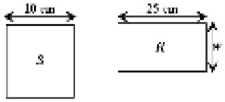
<table>
<tr><th rowspan="2">No.</th><th rowspan="2">TEST ITEMS</th><th rowspan="2">WORKING COLUMN</th><th colspan="3">Do Not Write Here</th></tr>
<tr><th>KC</th><th>AT</th><th>PS</th></tr>
<tr><td>11.</td><td>The length of a carrot is measured below.

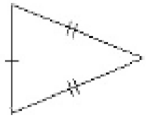
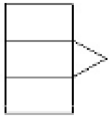
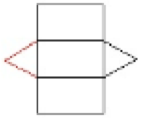
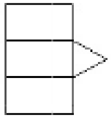
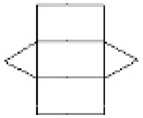

What is its length to the NEAREST centimetre?
Answer = 5 cm</td><td>

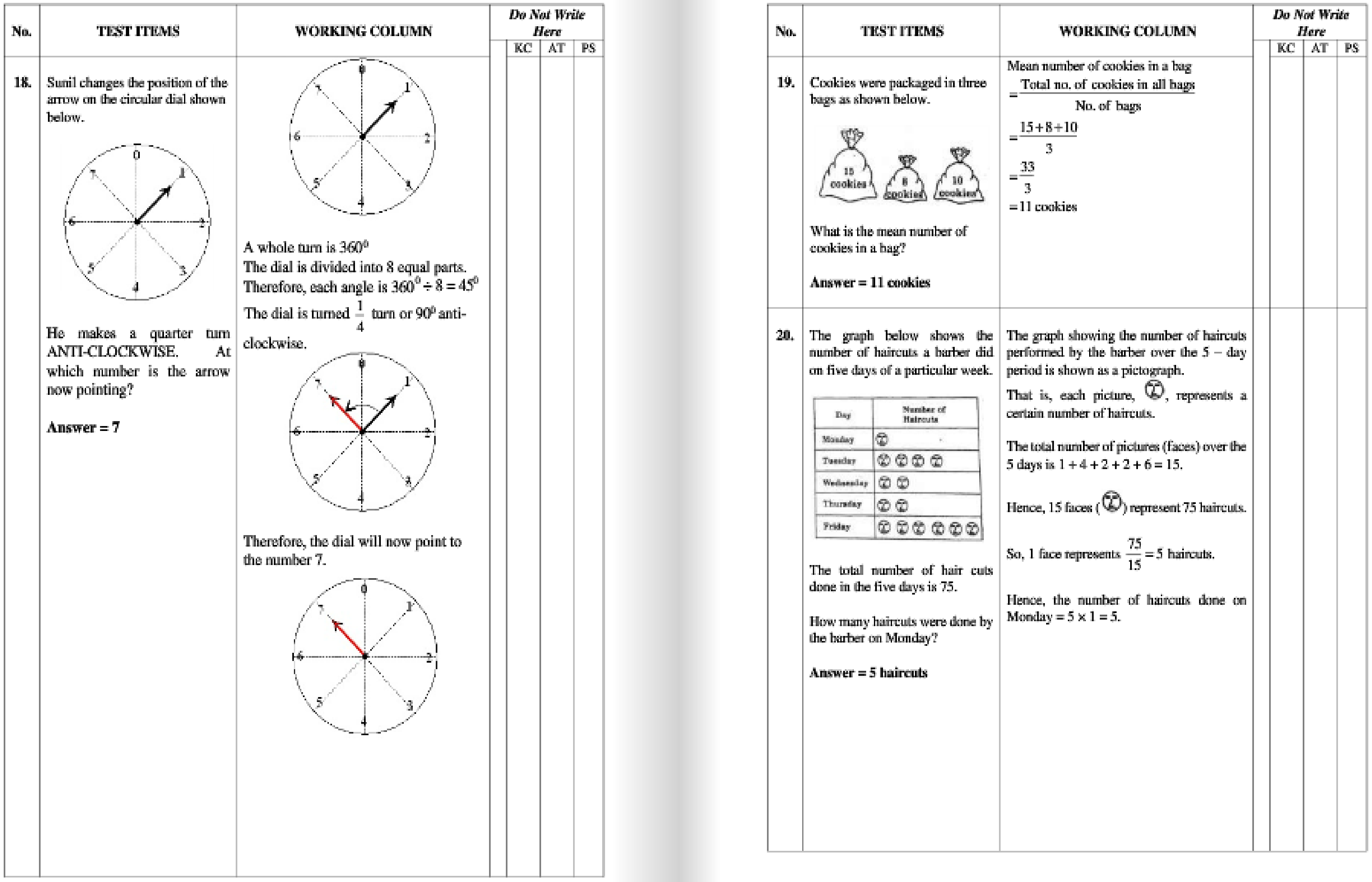
The end of the carrot (indicated by the blue line) lies before the halfway mark (shown red) between 5 cm and 6 cm.
Hence, the length of the carrot is 5 cm, when measured to the nearest cm.</td><td></td><td></td><td></td></tr>
<tr><td>12.</td><td>A square sheet of paper has sides of 11 cm.
What is its area?
Answer = 121 cm^2</td><td>

Area of the square sheet of paper
= Length of side × Length of side
= 11 cm × 11 cm
= 121 cm^2</td><td></td><td></td><td></td></tr>
<tr><td>13.</td><td>Shari has 4 coins on her desk. They have a total value of 50¢. The value of two coins is shown in the diagram below.


Write the correct value on EACH of the other 2 coins.
Answer = One 10¢ and One 25¢ as shown in red

 </td><td>The total value of all 4 coins in 50¢. We are shown:
1 coin with a value of 10¢ and
1 coin with a value of 5¢.
The value of these two coins together
= 10¢ + 5¢
=15¢
Hence, the value of the remaining two coins
= 50¢ - 15¢
= 35¢
Coins are made in the values of 1¢, 5¢, 10¢, 25¢ and 50¢. Two coins must have a total value of 35¢. Therefore, they must be one 10¢ and one 25¢.</td><td></td><td></td><td></td></tr>
</table>

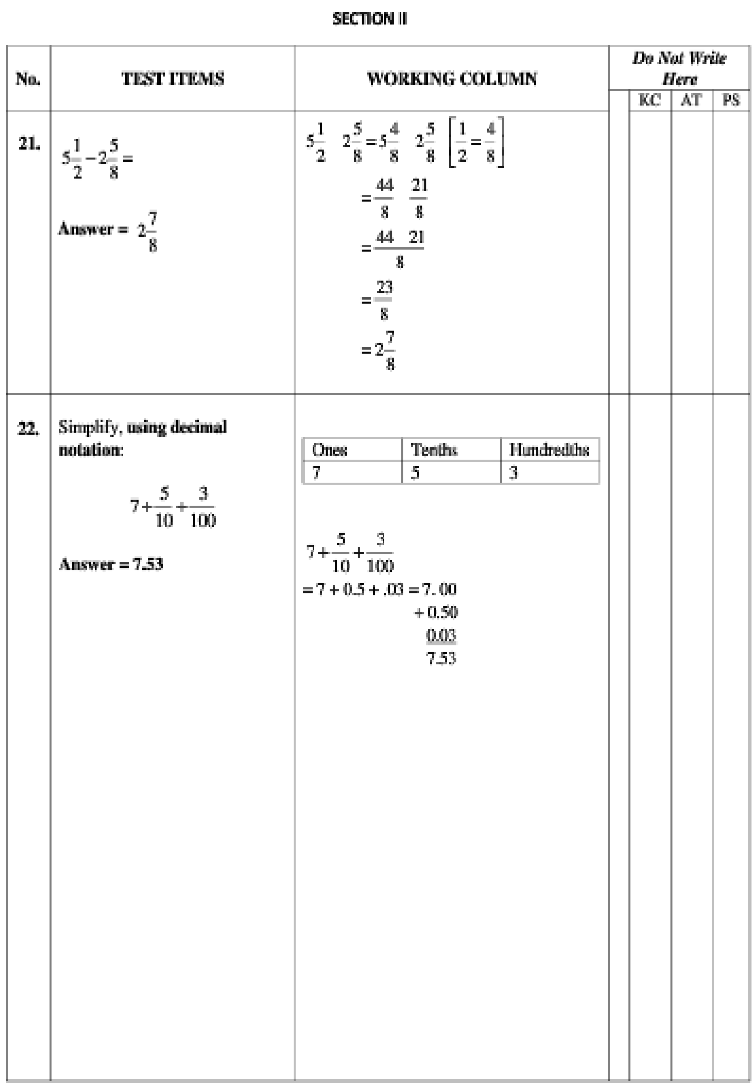
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
14.	A sales clerk is preparing a tag to show the selling price for a television.  Complete the tag below. **Cost Price** $1 740.00 **Discount** $ 174.00 **Selling Price** **Answer:** **Cost Price** $1 740.00 **Discount** $ 174.00 **Selling Price** **$1 566.00**	Selling price = Cost price – Discount = $1 740 –$ 1 74 $1 5 66 (It is better if, $1 740, be referred to as the '**Marked Price**', since '**Cost Price**' is actually the price that is paid for an item)			
15.	A square, labelled *S*, and a rectangle, labelled *R*, are shown below. **(The shapes are not drawn to scale.)**  Both shapes have the **same** area. Calculate the width, *w*, of the rectangle *R*? **Answer = 4 cm**	The area of the square, *S*, is the same as the area of the rectangle, *R*. Area of the square $= 10\text{ cm} \times 10\text{ cm}$ $= 100\text{ cm}^2$ Area of the rectangle $= 25 \times w$ Hence, $25 \times w = 100$ $w = 100 \div 25$ $= 4\text{ cm}$ Hence the width of the rectangle = 4 cm.			

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
16.	The triangle below has ONLY two sides of equal length.  What type of triangle is it? **Answer: Isosceles**	A triangle with only two equal sides is called isosceles. Such a triangle may also be identified by having only two equal angles.			
17.	Complete the drawing below to show the net of a triangular-based prism.  **Answer:** 	The incomplete net of a triangular – based prism given is  The prism would have three (3) equal rectangular faces and two (2) equal triangular faces. Hence, the completed net would look like:  When folded the solid figure would look like: 			



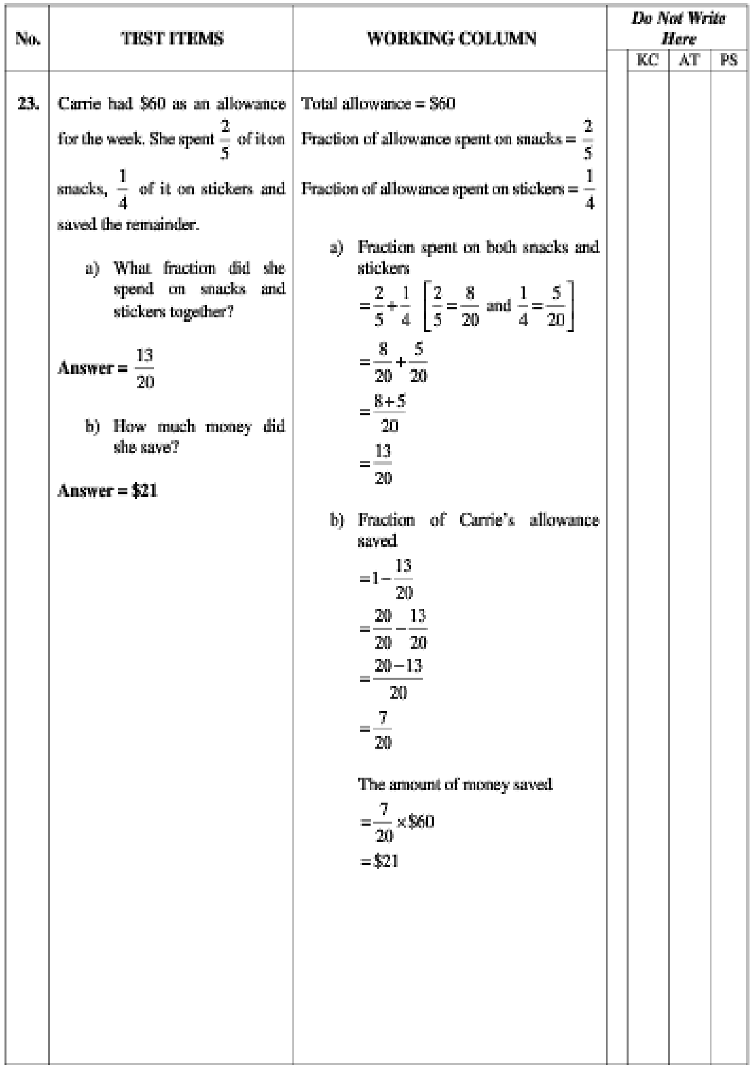
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
18.	Sunil changes the position of the arrow on the circular dial shown below. He makes a quarter turn ANTI-CLOCKWISE. At which number is the arrow now pointing? **Answer = 7**	A whole turn is 360^0 The dial is divided into 8 equal parts. Therefore, each angle is $360^0 \div 8 = 45^0$ The dial is turned $\frac{1}{4}$ turn or 90^0 anti-clockwise. Therefore, the dial will now point to the number 7.			

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
19.	Cookies were packaged in three bags as shown below. What is the mean number of cookies in a bag? **Answer = 11 cookies**	Mean number of cookies in a bag $= \frac{\text{Total no. of cookies in all bags}}{\text{No. of bags}}$ $= \frac{15+8+10}{3}$ $= \frac{33}{3}$ $= 11$ cookies			
20.	The graph below shows the number of haircuts a barber did on five days of a particular week. Day — Number of Haircuts Monday ☺ Tuesday ☺ ☺ ☺ ☺ Wednesday ☺ ☺ Thursday ☺ ☺ Friday ☺ ☺ ☺ ☺ ☺ ☺ The total number of hair cuts done in the five days is 75. How many haircuts were done by the barber on Monday? **Answer = 5 haircuts**	The graph showing the number of haircuts performed by the barber over the 5 – day period is shown as a pictograph. That is, each picture, ☺, represents a certain number of haircuts. The total number of pictures (faces) over the 5 days is $1 + 4 + 2 + 2 + 6 = 15$. Hence, 15 faces (☺) represent 75 haircuts. So, 1 face represents $\frac{75}{15} = 5$ haircuts. Hence, the number of haircuts done on Monday $= 5 \times 1 = 5$.			

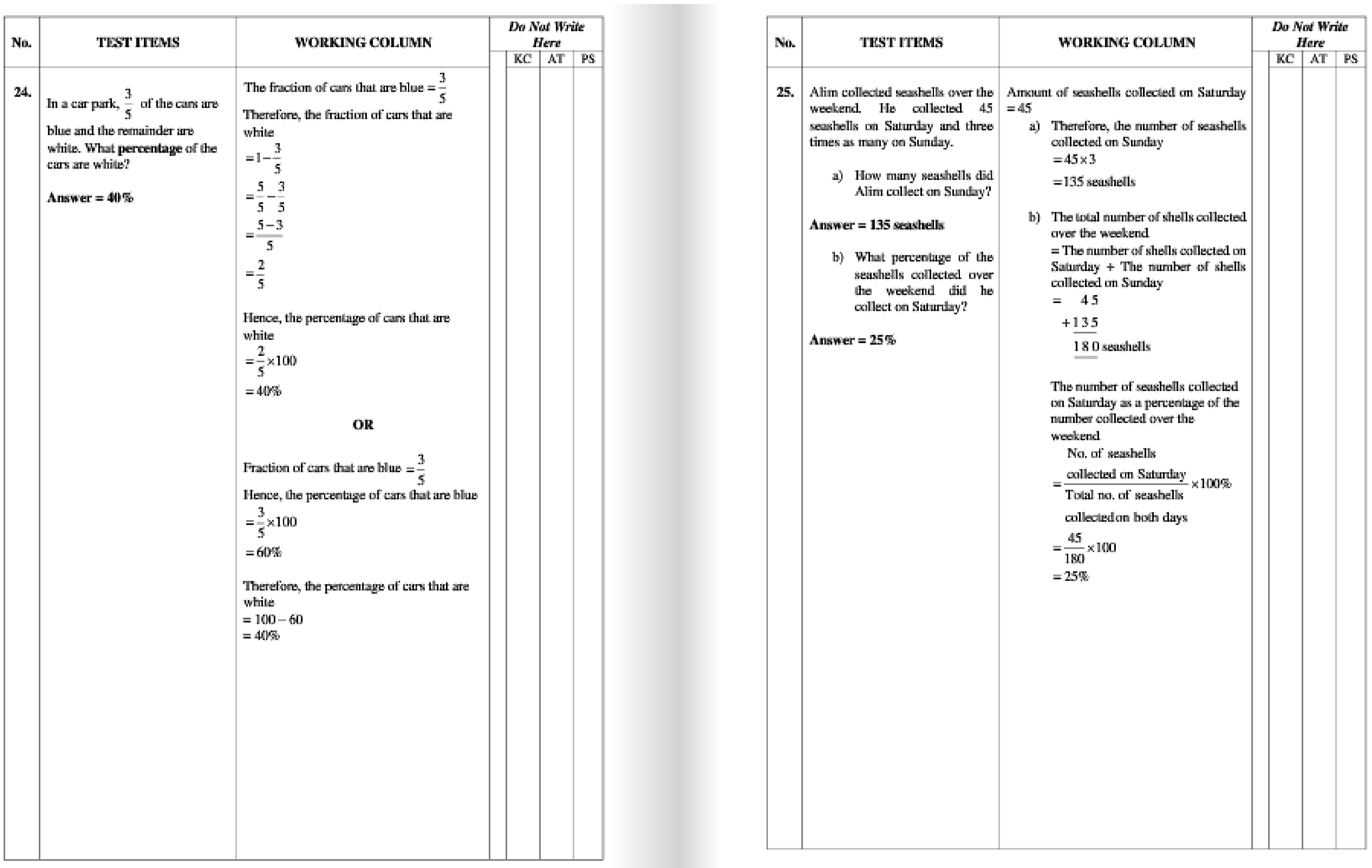


SECTION II

No.	TEST ITEMS	WORKING COLUMN	Do Not Write Here		
			KC	AT	PS
21.	$5\frac{1}{2}-2\frac{5}{8}=$ **Answer =** $2\frac{7}{8}$	$5\frac{1}{2}\ \ 2\frac{5}{8}=5\frac{4}{8}\ \ 2\frac{5}{8}\ \left[\frac{1}{2}=\frac{4}{8}\right]$ $=\frac{44}{8}\ \ \frac{21}{8}$ $=\frac{44\ \ 21}{8}$ $=\frac{23}{8}$ $=2\frac{7}{8}$			
22.	Simplify, **using decimal notation:** $7+\frac{5}{10}+\frac{3}{100}$ **Answer = 7.53**	Ones | Tenths | Hundredths 7 | 5 | 3 $7+\frac{5}{10}+\frac{3}{100}$ $=7+0.5+.03=7.00$ $+0.50$ $\underline{0.03}$ 7.53			



No.	TEST ITEMS	WORKING COLUMN	Do Not Write Here		
			KC	AT	PS
23.	Carrie had \$60 as an allowance for the week. She spent $\frac{2}{5}$ of it on snacks, $\frac{1}{4}$ of it on stickers and saved the remainder. a) What fraction did she spend on snacks and stickers together? **Answer =** $\frac{13}{20}$ b) How much money did she save? **Answer = \$21**	Total allowance = \$60 Fraction of allowance spent on snacks = $\frac{2}{5}$ Fraction of allowance spent on stickers = $\frac{1}{4}$ a) Fraction spent on both snacks and stickers $=\frac{2}{5}+\frac{1}{4}\ \left[\frac{2}{5}=\frac{8}{20}\text{ and }\frac{1}{4}=\frac{5}{20}\right]$ $=\frac{8}{20}+\frac{5}{20}$ $=\frac{8+5}{20}$ $=\frac{13}{20}$ b) Fraction of Carrie's allowance saved $=1-\frac{13}{20}$ $=\frac{20}{20}-\frac{13}{20}$ $=\frac{20-13}{20}$ $=\frac{7}{20}$ The amount of money saved $=\frac{7}{20}\times\$60$ $=\$21$			



No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
24.	In a car park, $\frac{3}{5}$ of the cars are blue and the remainder are white. What **percentage** of the cars are white? **Answer = 40%**	The fraction of cars that are blue $=\frac{3}{5}$ Therefore, the fraction of cars that are white $=1-\frac{3}{5}$ $=\frac{5}{5}-\frac{3}{5}$ $=\frac{5-3}{5}$ $=\frac{2}{5}$ Hence, the percentage of cars that are white $=\frac{2}{5}\times 100$ $=40\%$ **OR** Fraction of cars that are blue $=\frac{3}{5}$ Hence, the percentage of cars that are blue $=\frac{3}{5}\times 100$ $=60\%$ Therefore, the percentage of cars that are white $=100-60$ $=40\%$			

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
25.	Alim collected seashells over the weekend. He collected 45 seashells on Saturday and three times as many on Sunday. a) How many seashells did Alim collect on Sunday? **Answer = 135 seashells** b) What percentage of the seashells collected over the weekend did he collect on Saturday? **Answer = 25%**	Amount of seashells collected on Saturday $=45$ a) Therefore, the number of seashells collected on Sunday $=45\times 3$ $=135$ seashells b) The total number of shells collected over the weekend = The number of shells collected on Saturday + The number of shells collected on Sunday $=\ \ 45$ $+135$ 180 seashells The number of seashells collected on Saturday as a percentage of the number collected over the weekend $=\frac{\text{No. of seashells collected on Saturday}}{\text{Total no. of seashells collected on both days}}\times 100\%$ $=\frac{45}{180}\times 100$ $=25\%$			

<table>
<tr><th rowspan="2">No.</th><th rowspan="2">TEST ITEMS</th><th rowspan="2">WORKING COLUMN</th><th colspan="3">Do Not Write Here</th></tr>
<tr><th>KC</th><th>AT</th><th>PS</th></tr>
<tr><td>26.</td><td>The table below shows the points scored for hitting three different coloured targets in a video game.
<table><tr><th>Colour</th><th>No. of Points</th></tr><tr><td>White</td><td>2</td></tr><tr><td>Blue</td><td>3</td></tr><tr><td>Gold</td><td>5</td></tr></table>
Ana scored 53 points in playing the game. She hit the blue target 4 times and the white target 3 times.

How many times did she hit the GOLD target?

Answer = 7 times</td><td>When Ana hit the blue target 4 times she would have scored $3 \times 4 = 12$ points.

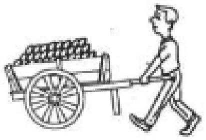
When Ana hit the white target 3 times she would have scored $2 \times 3 = 6$ points.

Hence, by hitting the blue targets and the white targets, Ana scored $12 + 6 = 18$ points.

Ana scored a total of 53 points.
Hence the number of points scored by Ana on hitting the gold target is $53 - 18 = 35$ points.

Each time Ana hits the gold target she scores 5 points.

For a total of 35 points, she would have hit the gold target
$= \frac{35}{5}$
$= 7$ times</td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><th rowspan="2">No.</th><th rowspan="2">TEST ITEMS</th><th rowspan="2">WORKING COLUMN</th><th colspan="3">Do Not Write Here</th></tr>
<tr><th>KC</th><th>AT</th><th>PS</th></tr>
<tr><td>27.</td><td>A fruit vendor transported 360 mangoes to the market. While transporting the mangoes, 10% of them were damaged.

a) How many mangoes were damaged?

Answer = 36 mangoes

b) The mangoes that were not damaged were packed into boxes of 12. How many boxes were used to pack these mangoes?

Answer = 27 boxes</td><td>The number of mangoes being transported to the market = 360
a) Percentage of mangoes damaged while being transported = 10%

The number of damaged mangoes
=10% of 360
$= \frac{10}{100} \times 360$
=36 mangoes

b) The number of mangoes that were not damaged
= The number of mangoes transported to the market − The number of damaged mangoes
= 360 − 36
= 324 mangoes

324 mangoes are to be packed in boxes of 12

The number of boxes used
$= \frac{324}{12}$
= 27 boxes

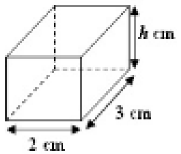
$$\begin{array}{r} 27 \\ 12\overline{)324} \\ -24 \\ \hline 84 \\ -84 \\ \hline 0 \end{array}$$</td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><th rowspan="2">No.</th><th rowspan="2">TEST ITEMS</th><th rowspan="2">WORKING COLUMN</th><th colspan="3">Do Not Write Here</th></tr>
<tr><th>KC</th><th>AT</th><th>PS</th></tr>
<tr><td>28.</td><td>Five years ago, Paul was $\frac{3}{8}$ his father's age. Paul's father is now 37 years old. How old is Paul now?

Answer = 17 years</td><td>Paul's father is now 37 years old.
Five years ago, Paul's father would have been 37 – 5 = 32 years old.

Therefore, Paul was $\frac{3}{8}$ of his father's age when his father was 32.

Paul's age at that time (5 years ago) would have been $\frac{3}{8} \times 32 = 12$ years.

Now, Paul's age $= 12 + 5$
$= 17$ years</td><td></td><td></td><td></td></tr>
<tr><td>29.</td><td>The volume of the cuboid shown below is 48 cm^3. The length of the cuboid is 3 cm, the width is 2 cm and the height is h cm.

Calculate the value of h.

Answer = 8</td><td>Volume of cuboid
$=$ Length $\times$ Width $\times$ Height
$= 48$ cm^3

Hence,
3 2 $h = 48$
$6h = 48$
$h = 48$ 6
$h = 8$</td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><th rowspan="2">No.</th><th rowspan="2">TEST ITEMS</th><th rowspan="2">WORKING COLUMN</th><th colspan="3">Do Not Write Here</th></tr>
<tr><th>KC</th><th>AT</th><th>PS</th></tr>
<tr><td>30.</td><td>Burns ran the following distances over a 2 – week period while training for the Olympics.
<table><tr><th>Week</th><th>Distance Ran</th></tr><tr><td>1</td><td>3 km 800 m</td></tr><tr><td>2</td><td>2.75 km</td></tr></table>
What is the TOTAL distance covered by Burns over the 2 weeks?

Answer = 6 km 550 m</td><td>Distance ran by Burns in week 1 = 3 km 800 m
Distance ran by Burns in week 2 = 2.75 km

1 km = 1000 m
Therefore 0.75 km $= 0.75 \times 1000$ m
$= 750$ m
Therefore, in week 2, Burns ran a distance of 2 km 750 m.

The total distance ran by Burns, over the two week period
= 3 km 800 m + 2 km 750 m

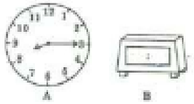
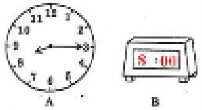
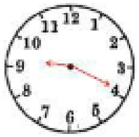
km m
$^{+1}$3 800
+ 2 750
6 550m

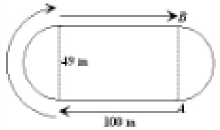
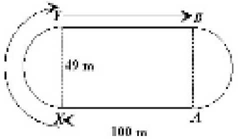
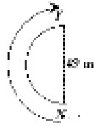
800 m
+ 750 m
1550 m =1km 550m</td><td></td><td></td><td></td></tr>
<tr><td>31.</td><td>Mr. Lee borrowed $8 000 from the bank to buy a used car. He paid simple interest at a rate of 12% per year for a period of 3 years.

How much simple interest did Mr. Lee pay?

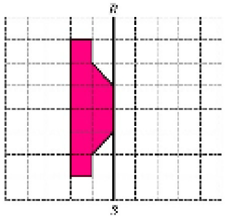
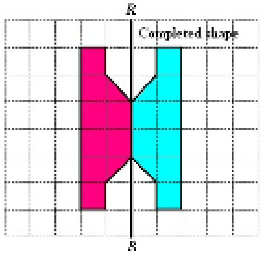
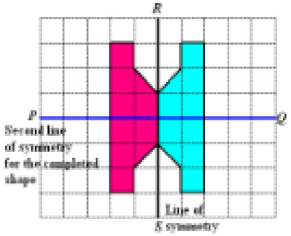
Answer = $2 880</td><td>Simple Interest
$= \frac{\text{Principal} \times \text{Rate} \times \text{Time}}{100}$
$= \frac{\$8000 \times 12 \times 3}{100}$
$= \$2880$</td><td></td><td></td><td></td></tr>
</table>

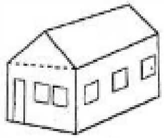
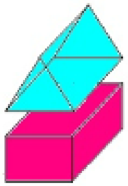
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
32.	Jane bought the three items, shown below, at the supermarket.  a) What is the TOTAL cost of the 3 items? **Answer = $7.65** b) Calculate Jane's EXACT change if she paid for the items with a $20.00 bill. **Answer = $12.35**	a) Cost of 1 notebook = $1.90 Cost of 1 comb = $2.10 Cost of 1 pair of scissors = $3.65 Total cost =$7.65 a) Change from $20.00 will be The amount that Jane paid – The cost of all the items = $ 20.00 – $ 7.65 $1 2. 35			

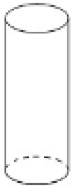
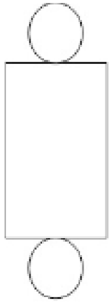
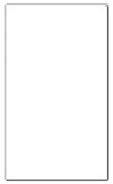
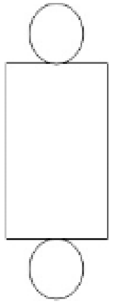
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
33.	The time shown on Clock A is 15 minutes AHEAD of the correct time.  a) Insert the correct time on Clock B. **Answer = 8:00** b) At the end of the next hour, Clock A gained an additional 5 minutes ahead of the correct time. What time will be shown on Clock A? **Answer = 9:20**	a) The time shown on Clock A is a quarter past 8 **OR** 15 minutes after 8 o'clock. Since the time shown is 15 minutes ahead of the correct time, then the correct time is 8 : 1 5 – : 1 5 8 : 00 The correct time shown on Clock B should be 8:00, since B is a digital clock.  b) Clock A shows the incorrect time of 8:15. In one hour time, Clock A shows an additional 5 minutes ahead. Hence, Clock A will show: 8 : 1 5 + 1 : 0 5 9: 2 0 			

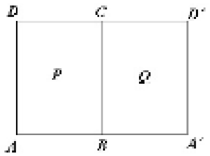
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
34.	The diagram below shows a cycling track consisting of a rectangle and two semi-circles.  A cyclist starts at point A and cycles in the direction of the arrows to point B. What distance did he cover? $\left(\pi = \frac{22}{7}\right)$ **Answer = 277 m**	 The points X and Y are named on the figure for convenience. Since the arrow shows the direction of the cyclist from A to B, we can say the cyclist rides from A to X, X to Y and then Y to B. From A to X is 100 m.  From X to Y is a semi-circle of diameter 49 m.  The distance from X to Y is one half the circumference of the circle $= \frac{1}{2}(\text{Diameter} \times \pi)$ $= \frac{1}{2}\left(49 \times \frac{22}{7}\right)$ $= 77$ m From Y to B is 100 m.  Therefore the total distance covered by the cyclist $= 100 + 77 + 100$ $= 277$ m			

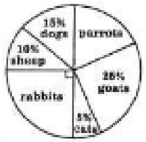
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
35.	The school cafeteria, bought 5 dozen silly bands at \$15 per dozen and sold them for \$2 EACH. a) What was the profit, in dollars, made by the school cafeteria? **Answer = \$45** b) Calculate the profit as a percentage of the cost price. **Answer = 60%**	a) The cost of one dozen silly bands = \$15 Therefore, the cost of 5 dozen silly bands $= \$15 \times 5$ $= \$75$ The selling price of 1 silly band = \$2. Hence, the selling price of all 5 dozen silly bands $= 5 \times 12 \times \$2$ $= \$120$ The profit made $= \text{Selling price} - \text{Cost price}$ $= \$120 - \$75$ $= \$45$ b) Profit as a percentage of the cost price $= \frac{\text{Profit}}{\text{Cost price}} \times 100\%$ $= \frac{45}{75} \times 100\%$ $= 60\%$			

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
36.	The diagram below shows an incomplete plane shape and one of its lines of symmetry, *RS*.  a) Complete the drawing of the shape. b) Draw another line of symmetry, *PQ*, for the completed shape.	a)  b)  			

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
37.	Use the diagram below to answer the questions that follow.  State the names of: a) ONE plane shape in the diagram above. **Answer: Rectangle,** **(Also - triangle or pentagon or square)** b) ONE solid that would be needed to make a model of the building. **Answer: A Cuboid** **(or a prism)**	a) In the diagram shown, one plane shape is a rectangle (the door, the four sides of the house and the two sides of the roof). There are other plane shapes such as a square (the windows). If we consider the front and the back of the house as a complete face, then we also have 5 sided shapes or pentagons as another plane shape. If we consider the dotted line as dividing the front face of the house, then we have a triangle as another plane shape. b) In order to make a model to the building we would need a cuboid. A prism can be attached to the top to complete the model (A prism is another solid that would be needed to complete the model. It would be placed at the top of the cuboid, with both solids having no overlap). 			

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
38.	The diagram below shows a cylindrical package closed at both ends.  a) How many faces does the package have? **Answer: 3 faces** b) Draw the net of the package. 	a) The cylinder has two circular faces (top and bottom) and a curved face. Hence, the cylinder has 3 faces. b) The net of the curved face only is a rectangle.  The net of the complete package must include the base and top. These are both circles, so the complete net is shown below. 			

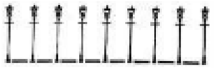
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
39.	The shape $ABCD$ below is moved from its position at P to the position at Q so that Corner A is now at A' and Corner D is now at D'.  a) What is the name of this movement? **Answer: Reflection or flip** b) Describe the movement in (a) FULLY. **Answer: $BCDA$ is reflected in the line BC to produce the image $BCD'A'$.** c) Under the same movement in (a), describe what happens to Corner B? **Answer: Corner B remained in the same place. We can say that Corner B is invariant.**	a) In the movement, B and C remain in the same position. $BCD'A'$ is the same size as $BCDA$. The movement is a reflection or a 'flip' about the line BC. b) BC acts as the line of reflection. The shape $BCDA$ is reflected in the line BC to produce the image $BCD'A'$. c) The Corner B remained in the same place and is unaltered or unchanged. In mathematics, we can say that Corner B is invariant. (The same can be said for Corner C).			

<table>
<tr><th rowspan="2">No.</th><th rowspan="2">TEST ITEMS</th><th rowspan="2">WORKING COLUMN</th><th colspan="3"><i>Do Not Write Here</i></th></tr>
<tr><th>KC</th><th>AT</th><th>PS</th></tr>
<tr><td>40.</td><td>The pie chart below shows the favourite pets of the pupils in Standard 3.


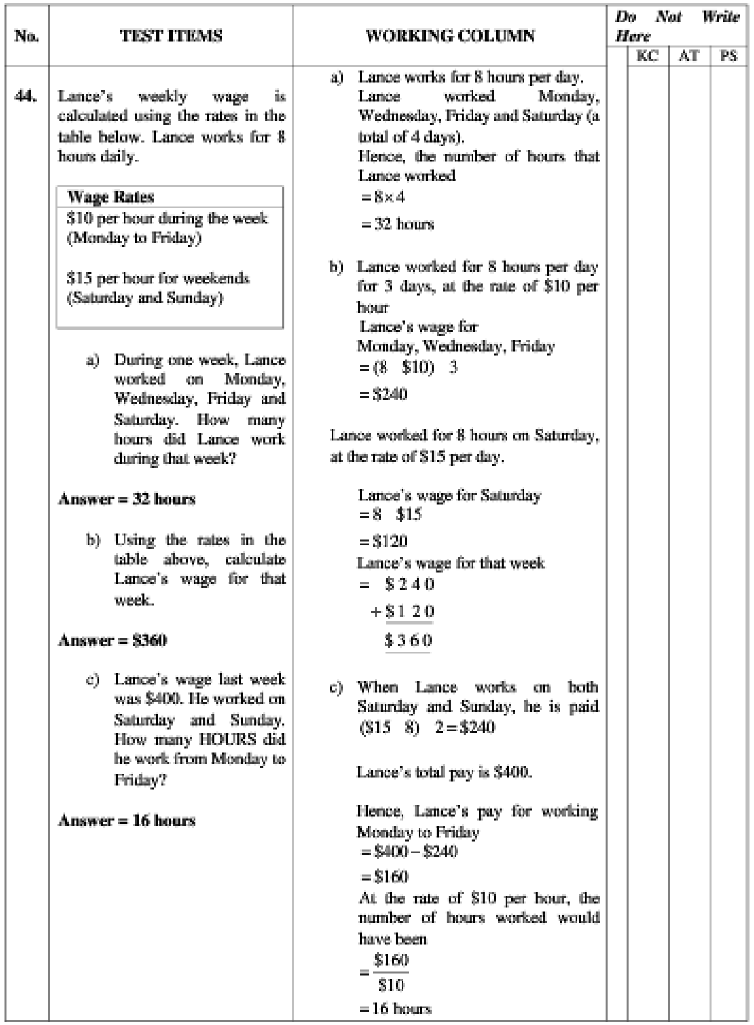
a) Which TWO pets are equally favoured?
Answer: Rabbits and goats
b) What percentage of the pupils in Standard 3 favour parrots?
Answer: 20%</td><td>a) From the pie chart we can list a table to show the percentage of each sector.
<table><tr><td>Cats</td><td>5%</td></tr><tr><td>Dogs</td><td>15%</td></tr><tr><td>Sheep</td><td>10%</td></tr><tr><td>Goats</td><td>25%</td></tr><tr><td>Rabbits</td><td>$\frac{90^o}{360^o}\times 100 = 25\%$</td></tr><tr><td>Parrots</td><td>$100 \quad (5+15+10+25+25)$
$= 20\%$</td></tr></table>
Hence, rabbits and goats are equally favoured since they both have the same percentage of pupils who favour them.
b) Since the sum of the percentage of all the sectors must total 100, then the percentage of pupils who favour parrots
$=100-(5+15+10+25+25)$
$=20\%$</td><td></td><td></td><td></td></tr>
</table>

SECTION III

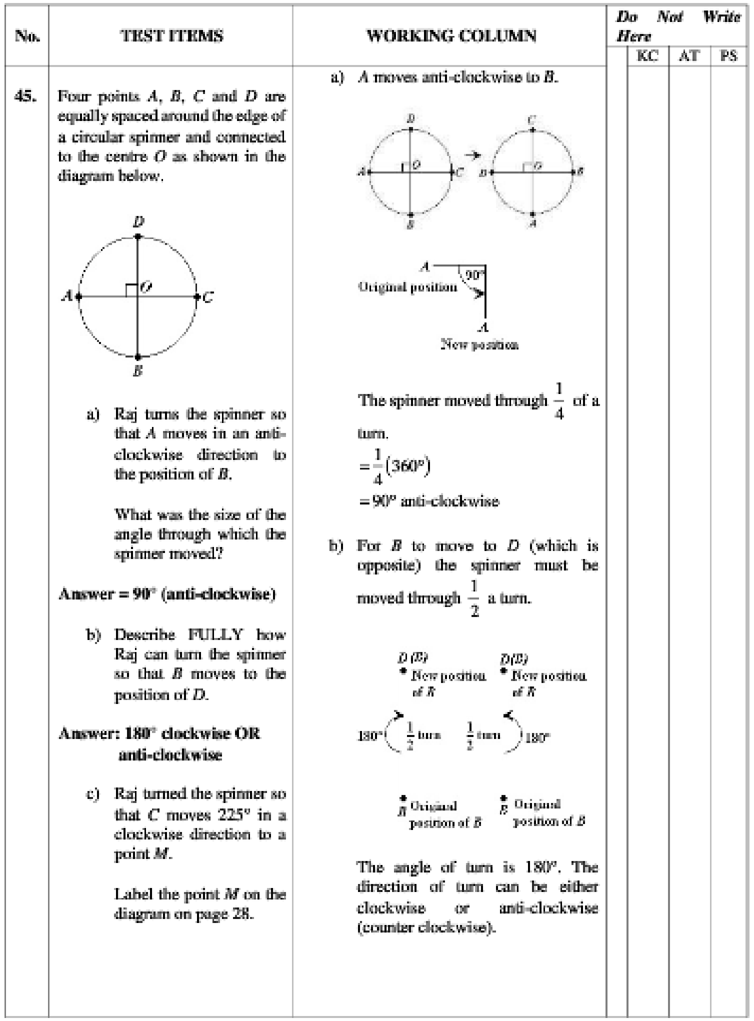
<table>
<tr><th rowspan="2">No.</th><th rowspan="2">TEST ITEMS</th><th rowspan="2">WORKING COLUMN</th><th colspan="3"><i>Do Not Write Here</i></th></tr>
<tr><th>KC</th><th>AT</th><th>PS</th></tr>
<tr><td>41.</td><td>A fruit vendor has 160 fruits in his stall. Of these, $\frac{3}{8}$ are mangoes and 20% are plums. The remainder is avocados.
a) How many mangoes does he have?
Answer = 60 mangoes
b) Express the number of fruits that are plums as a DECIMAL fraction.
Answer = 0.2
c) Calculate the number of avocados in his stall.
Answer = 68 avocados</td><td>a) Total number of fruits = 160
$\frac{3}{8}$ of the fruits are mangoes.
Therefore the number of mangoes
$=\frac{3}{8}\times 160$
$= 60$ mangoes
b) 20% of the fruits are plums.
To express 20% as a decimal:
$20\% \quad \frac{20}{100}=\frac{2}{10}=0.2$
The number of fruits that is plums as a decimal fraction, is 0.2.
c) The number of fruits this is plums
= 20% of 160
$=\frac{20}{100} \quad 160$
$= 32$
Besides mangoes and plums, the remainder of fruits is avocados.
Therefore the number of avocados
= No. of fruits −
(No. of mangoes + No. of plums)
$=160-(60+32)$
= 68 avocados</td><td></td><td></td><td></td></tr>
</table>

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*			
				KC	AT	PS
42.	Nine light posts are evenly spaced along a highway. A total of 144 plastic pipes of the same length is placed EQUALLY between the 9 posts.  a) How many pipes are placed between the first and second posts? **Answer = 18 pipes** b) Each pipe is 7 m long. The pipes are connected end-to-end (just touching each other) between the posts. What is the distance between the first and second posts? **Answer = 126 m**	a) Since there are 9 posts and the pipes are equally spaced between the posts, then the pipes are equally placed between 8 spaces. 144 pipes are spaced equally between the 8 spaces which lie between the posts. Therefore, between each set of posts there would be $\frac{144}{8}=18$ pipes. b) Length of each pipe = 7 m 18 pipes are placed, end-to-end, between the 1st and 2nd posts. ASSUMING that the pipes are all straight and that they lie in a straight line, the distance between the 1st and 2nd post will be the total length of all 18 pipes. $=18\times 7$ m $=126$ m Hence, the distance between the 1st and 2nd posts is 126 m.				

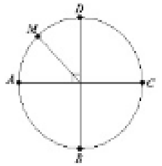
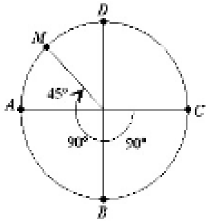
No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*			
				KC	AT	PS
43.	An applicance store offers $33\frac{1}{3}\%$ discount on cash purchases. The original price on a refrigerator is $8 400. a) What is the discounted price on the refrigerator? **Answer = $5 600** b) VAT at 15% is calculated on the discounted price. How much is the VAT? **Answer = $840** c) What is the FINAL cost of the refrigerator when purchased for cash? **Answer: $6 440**	a) Original price of refrigerator = $8400. Discount $=33\frac{1}{3}\%$ $=33\frac{1}{3}\%=\frac{1}{3}$ Discount $=\frac{1}{3}\ \$8400$ $=\$2800$ The discounted price = Original price − Discount $=\ \$8400$ $-\$2800$ $\$5600$ b) VAT = 15% of the discounted price $\therefore \text{VAT}=\frac{15}{100}\times\$5600$ $=15\times\$56$ $=\$840$ c) The final cost of the refrigerator = Discounted price + VAT = $5 600 + $840 = $6 440				

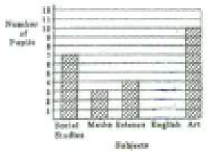


No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
44.	Lance's weekly wage is calculated using the rates in the table below. Lance works for 8 hours daily. **Wage Rates** $10 per hour during the week (Monday to Friday) $15 per hour for weekends (Saturday and Sunday) a) During one week, Lance worked on Monday, Wednesday, Friday and Saturday. How many hours did Lance work during that week? **Answer = 32 hours** b) Using the rates in the table above, calculate Lance's wage for that week. **Answer = $360** c) Lance's wage last week was $400. He worked on Saturday and Sunday. How many HOURS did he work from Monday to Friday? **Answer = 16 hours**	a) Lance works for 8 hours per day. Lance worked Monday, Wednesday, Friday and Saturday (a total of 4 days). Hence, the number of hours that Lance worked $=8\times4$ $=32$ hours b) Lance worked for 8 hours per day for 3 days, at the rate of $10 per hour Lance's wage for Monday, Wednesday, Friday = (8 $10) 3 = $240 Lance worked for 8 hours on Saturday, at the rate of $15 per day. Lance's wage for Saturday = 8 $15 = $120 Lance's wage for that week = $ 2 4 0 + $ 1 2 0 $ 3 6 0 c) When Lance works on both Saturday and Sunday, he is paid ($15 8) 2 = $240 Lance's total pay is $400. Hence, Lance's pay for working Monday to Friday = $400 – $240 = $160 At the rate of $10 per hour, the number of hours worked would have been = $160 / $10 = 16 hours			



No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
45.	Four points *A*, *B*, *C* and *D* are equally spaced around the edge of a circular spinner and connected to the centre *O* as shown in the diagram below. D A O C B a) Raj turns the spinner so that *A* moves in an anti-clockwise direction to the position of *B*. What was the size of the angle through which the spinner moved? **Answer = 90° (anti-clockwise)** b) Describe FULLY how Raj can turn the spinner so that *B* moves to the position of *D*. **Answer: 180° clockwise OR anti-clockwise** c) Raj turned the spinner so that *C* moves 225° in a clockwise direction to a point *M*. Label the point *M* on the diagram on page 28.	a) *A* moves anti-clockwise to *B*. D / A O C / B → C / D O B / A A 90° Original position A New position The spinner moved through $\frac{1}{4}$ of a turn. $=\frac{1}{4}(360^\circ)$ $=90^\circ$ anti-clockwise b) For *B* to move to *D* (which is opposite) the spinner must be moved through $\frac{1}{2}$ a turn. D (B) New position of B D(B) New position of B 180° $\frac{1}{2}$ turn $\frac{1}{2}$ turn 180° B Original position of B B Original position of B The angle of turn is 180°. The direction of turn can be either clockwise or anti-clockwise (counter clockwise).			

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
		c) The spinner is turned so that *C* moves 225° clockwise to *A*.  The angles rotated from *C* to *B* + from *B* to *A* + from *A* to *M* $=90^\circ+90^\circ+45^\circ$ $=225^\circ$ Therefore *M* is situated halfway between *A* and *D* along the circle.			

No.	TEST ITEMS	WORKING COLUMN	*Do Not Write Here*		
			KC	AT	PS
46.	The incomplete bar graph below shows the favourite subjects of the 30 pupils in a Standard 5 class.  a) How many **more** pupils favour Social Studies than Maths? **Answer = 4 pupils** b) What percentage of the class chose Maths as their favourite subject? **Answer = 10%** c) How many pupils chose English as their favourite subject? **Answer = 6 pupils** d) Complete the graph on page 30 by drawing the bar to represent the number of pupils whose favourite subject is English.	a) The number of students who favour Social Studies = 7 The number of students who favour Maths = 3 Hence, 7-3 = 4 more students favour Social Studies than Maths b) The total number of pupils in the class = 30 Percentage of pupils who favour Maths $=\frac{\text{No. of pupils who favour Maths}}{\text{Total no. of pupils}}$ $=\frac{3}{30}\times100\%$ $=10\%$ c) From the bar graph, there is no bar drawn, showing the number of pupils who favour English. Number of pupils who favour English = Number of students in the class – Number of students who favour the remaining subjects $=30-(7+3+4+10)$ $=30-24$ $=6$ d) The completed bar graph showing the number of students who favour English will be: