

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Latent profiles of children's capabilities: Measurement invariance and differential
item functioning across gender and ethnicity within a New Zealand preschool-
aged cohort**

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Education
at Massey University, Manawatū, Aotearoa New Zealand.

Hui Zeng

2025

Abstract

A robust understanding of young children's learning and development is key to unlocking their immense potential. Within education, young children's learning and development are increasingly being addressed from strengths-based, holistic, and culturally connected perspectives. Nevertheless, previous research on children's learning and development has primarily adopted a single-lens approach, highlighting developmental delays or risks associated with domain-specific deficits. The current thesis sought to better understand and account for children's learning and development in strengths-based, holistic, and culturally connected ways. Two phases of the research were designed to: 1) support a holistic approach to understanding young children's capabilities of social, emotional, language, and executive function; 2) investigate the comparability of the shared patterns of children's capabilities across gender and ethnicity, with implications for fairness in assessment and inclusion in early childhood research.

Secondary data analyses were conducted in two sequential phases, with data drawn from the Growing Up in New Zealand study 54-month dataset. The first phase of the research used latent profile analysis to identify shared patterns of children's capabilities. Data were available for six specific areas of children's capabilities: prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function. Information about children's capabilities in these areas was gathered using the following measures: prosocial behaviour and emotional symptom

subscales in the Strengths and Difficulties Questionnaire, the modified Affect Knowledge Test, the Parent Rating of Oral Language, the shortened version of Peabody Picture Vocabulary Test III, and the Luria Hand Clap task.

Latent profile analysis is often used to identify unique patterns of capabilities across a range of domains. The present research identified a 3-group model of young children's capabilities, labelled as *Emerging*, *Progressing*, and *Competent* profiles. The associations within this 3-group model were characterised by overall differences in capabilities rather than distinct patterns. The *Emerging* profile made up 9.1% of the sample, representing children with capabilities lower than their peers in other profiles. The *Progressing* profile made up 45.1% of the sample, representing children with close to average capabilities, between the *Emerging* and *Competent* profiles. The *Competent* profile made up 45.8% of the sample, representing children with the highest capabilities relative to peers.

The second phase of the research tested the measurement invariance of the selected measures through multi-group confirmatory factor analysis and examined differential item functioning for gender and ethnicity as covariates to the latent profile factors, via Multiple Indicators Multiple Causes modelling. Findings from multi-group confirmatory factor analyses established partial invariance for the prosocial behaviours and emotional symptoms subscales in the Strengths and Difficulties Questionnaire, full invariance for the modified Affect Knowledge Test and the Parent Rating of Oral Language across gender groups, and partial invariance for these four measures across

ethnic groups. Due to the nature of the data available, it was not possible to examine measurement invariance for the shortened version of the Peabody Picture Vocabulary Test III and the Luria Hand Clap task. Nonetheless, differential item functioning via Multiple Indicators Multiple Causes modelling suggested gender was a potential source of uniform differential item functioning for prosocial behaviours, emotional symptoms, emotional knowledge, and expressive language, and nonuniform differential item functioning for receptive language and executive function; however, these effects were determined to be small or negligible. Ethnicity was also identified as a potential source of uniform differential item functioning for emotional symptoms, and nonuniform differential item functioning for emotional knowledge, expressive language, receptive language, and executive function. These effects were determined to be more notable, suggesting the assessments used may not measure the constructs equivalently across ethnic groups, potentially indicating bias in measurement. While the negligible or small uniform and nonuniform differential item functioning effects across gender groups supported the comparability of the 3-group model, the more substantial effects did not support the comparability of the 3-group model across ethnic groups.

Findings from the present research offer important insights into understanding children's learning and development in strengths-based, holistic, and culturally connected ways. First, the investigation sought to use a strengths-based lens by focusing on what children could do relative to peers (e.g., capabilities) rather than framing scores from a deficit or risk-based perspective. Second, the 3-group model

centralised the 'whole learner' perspective by incorporating six different areas of children's capabilities into one investigation, highlighting the interconnected nature of these capabilities. Rather than forming distinct patterns, the model reflected overall differences in capabilities across domains, reinforcing the holistic nature of development. Third, the focus on culturally connected assessment sought to understand how appropriate the measures available were for different aspects of children's identity (i.e., gender and ethnicity). Testing measurement invariance was key to highlighting the importance of fairness in educational assessment and practices. Failure to account for differential item functioning effects may result in inappropriate conclusions about the comparability of latent profiles across observed groups, such as gender or ethnicity.

Findings offer significant implications for early childhood educational research and practices, which can inform decisions about service design and policy development. The holistic approach reflected by the 3-group model is a shift away from singular, domain-specific perspectives and research. Given the importance of strengths-based, holistic, and culturally connected teaching and learning support reflected in the early childhood curriculum in New Zealand, current research and programmes targeting isolated capabilities from risk-based perspectives may have limited or misaligned applications in early childhood education. Additionally, factors such as gender and ethnicity may introduce biases in the estimation of effects for the latent profiles identified in the present research, thereby potentially limiting their

comparability across diverse subpopulations. This underscores the importance of fairness in assessment and highlights the cautions and limitations of existing research. Taken together, the thesis highlights the importance of holistic and unbiased assessments to better understand and report on children's capabilities, which can in turn influence teaching and learning policy and practices in inclusive and equitable ways.

Acknowledgements

As I begin these acknowledgments, I have thought through where to start, whom I should express my gratitude to, and how this will shape my future. Often, it is challenging not to overlook anyone who has contributed to or supported me throughout my PhD journey, both in my studies and life in general. Undoubtedly, there are many significant contributors to these efforts. Let me quote the Chinese Confucian Philosopher Mencius, 'Everything goes right when it is at the right time, in the right place, and with the right people'.

The onset of COVID-19 in 2020 led me to Palmerston North to study in 2021, a decision that has become a blessing in disguise. While the pandemic was disastrous for many, it became the push I needed to finally pursue my long-delayed PhD, a dream I had initially planned to start in 2018 or even earlier. Being here in Palmerston North, I have been surrounded by nature and escaped from the modern metropolis's noise, hustle, and crowdedness, which has provided the ideal setting for my studies. Even though it is thousands of miles away from my home country and contrasting societal values towards many aspects of life, I have deeply felt both motivated and respected, especially as a single woman in my early-mid 30s, making this unexpected journey one of personal growth and fulfilment.

First and foremost, I am deeply grateful for my academic supervisors: Associate Professor Tara McLaughlin, Dr. Rachael Pond, and Dr. Joanne Allen. I must extend my appreciation to Tara, who has always been instrumental in my journey. Her unwavering

support and acceptance of who I am and where I come from have been invaluable, especially as I navigate my studies being an English learner and acknowledge unique challenges. She exemplifies the model of professionalism, dedication, and hard work that I aspire to emulate. Rachael has significantly influenced my approach to research by emphasising the importance of conceptual thinking and critical evaluation of theories. Her guidance has been crucial in sparking my own research ideas. Jo has been an exceptional supervisor, combining her expertise in statistical analysis with a deep understanding of child development theories. Her meticulous approach to statistical analyses has been inspiring and something I have always strived to achieve. Thank you all for guiding me through this incredible learning journey and helping me achieve my best throughout my PhD.

I am also filled with gratitude for the incredible support I received from the team at Massey University and my advisory groups from the Growing Up in New Zealand study. Picture a group of dedicated individuals, each playing an important role in my academic adventure. Our postgraduate coordinators—Professor Howard Lee, Professor John O'Neill, Professor Roseanna, and the ever-supportive postgraduate administrator, Roseanne MacGillivray- were like guiding stars, helping me navigate my learning needs to be met with care and attention. A special thanks to Dr. Linda Clarke for her invaluable support in editing my thesis. I appreciate the services provided by other Massey departments, such as the student health and counselling services, the Library and the Graduate Research School, offering a range of resources that fuelled my

research. I also thank my advisory groups from the Growing Up in New Zealand team: Associate Professor Elizabeth Peterson, Professor Annette Henderson, Professor Te Kani Kingi, and Associate Professor Pania Te Maro from Massey. I am deeply appreciative of every facet of assistance that contributed to my growth. Together, they made my learning journey not just a path to knowledge but a truly enriching experience.

My gratitude extends to my family, friends, and colleagues. I want to thank my parents for not pressuring me into traditional domestic roles, such as wife and mother. Their support has allowed me to fully appreciate and embrace gender equality. This support has empowered me to pursue my diverse capabilities and aspirations as a woman across all areas of my life. Equally important are my friends and colleagues, who have been a constant source of support and solidarity, forming a close-knit community that uplifts and encourages one another. The friendship and support I was given have been invaluable as we navigate the challenges and triumphs together as allies.

Should anyone feel overlooked, I sincerely apologise. Overall, I am immensely thankful for the journey of working towards my PhD. It has been a wonderful and challenging experience filled with valuable lessons. One key insight I have gained is that this thesis represents the start of an ongoing exploration that I hope to continue. The field of exploring young children's learning and development, particularly using advanced quantitative methodologies, is evolving and challenging. So I encourage future researchers to engage in impactful work that advances research in New Zealand. And I will do so, too.

Table of Contents

ABSTRACT	1
ACKNOWLEDGEMENTS	6
TABLE OF CONTENTS	9
ABBREVIATIONS AND TERMS.....	16
LIST OF TABLES	18
LIST OF FIGURES	21
CHAPTER 1 THESIS INTRODUCTION	23
1.1 RESEARCH OVERVIEW.....	24
1.2 RESEARCHER BACKGROUND	26
1.3 RESEARCH RATIONALE	28
<i>1.3.1 Social, Emotional, Language and Executive Function Capabilities</i>	<i>28</i>
<i>1.3.2 Holistic Learning and Development</i>	<i>29</i>
<i>1.3.3 Diversity in Learning and Development.....</i>	<i>30</i>
<i>1.3.4 Research to Understand Children's Learning and Development</i>	<i>31</i>
<i>1.3.5 Policy Context for Early Learning.....</i>	<i>33</i>
1.4 INTRODUCTION SUMMARY	36
1.5 THESIS STRUCTURE	37
CHAPTER 2 LITERATURE REVIEW	40
2.1 SOCIAL COMPETENCE	40
<i>2.1.1 Importance of Social Competence</i>	<i>41</i>
<i>2.1.2 Theories of Social Competence</i>	<i>42</i>

2.1.3 Definitions.....	45
2.1.4 Domains of Social Competence.....	46
2.1.5 Assessments of Social Competence.....	48
2.1.6 Summary.....	50
2.2 EMOTIONAL COMPETENCE.....	51
2.2.1 Importance of Emotional Competence.....	51
2.2.2 Theories of Emotional Competence.....	52
2.2.3 Definitions.....	55
2.2.4 Domains of Emotional Competence.....	57
2.2.5 Assessments of Emotional Competence.....	60
2.2.6 Summary.....	61
2.3 LANGUAGE SKILLS.....	62
2.3.1 Importance of Language.....	62
2.3.2 Theories of Language.....	64
2.3.3 Definitions.....	69
2.3.4 Domains.....	74
2.3.5 Assessments of Language.....	76
2.3.6 Summary.....	78
2.4 EXECUTIVE FUNCTION.....	79
2.4.1 Importance of Executive Function.....	80
2.4.2 Theories of Executive Function.....	82
2.4.3 Definitions.....	85
2.4.4 Domains.....	86
2.4.5 Assessments of Executive Function.....	90
2.4.6 Summary.....	92

2.5 CHAPTER SUMMARY	92
CHAPTER 3 HOLISTIC VIEWS OF DEVELOPMENT AND OBSERVED SUBGROUP VARIATIONS	94
3.1 THE INTERWOVEN NATURE OF SOCIAL, EMOTIONAL, LANGUAGE AND EXECUTIVE FUNCTION CAPABILITIES	95
3.2 THEORETICAL VIEWS OF HOLISTIC DEVELOPMENT	97
3.3 PERSON-CENTRED ANALYSIS METHODS	102
3.4 VARIATION AMONG OBSERVED SUBGROUPS: GENDER AND ETHNICITY	115
<i>3.4.1 Gender and Differences in Development by Gender</i>	<i>116</i>
<i>3.4.2 Ethnicity and Differences in Development by Ethnicity</i>	<i>121</i>
3.5 IMPORTANCE OF ENSURING COMPARABILITY	125
<i>3.5.1 Comparability of Measurement Across Observed Subgroups</i>	<i>127</i>
<i>3.5.2 Comparability on Person-Centred Model Level</i>	<i>133</i>
<i>3.5.3 Chapter Summary</i>	<i>138</i>
CHAPTER 4 METHODOLOGY	140
4.1 THE GROWING UP IN NEW ZEALAND STUDY	140
4.2 THE CURRENT RESEARCH	142
4.3 HYPOTHESES	146
4.4 SAMPLE	147
4.5 ETHICAL CONSIDERATIONS	148
4.6 SELECTED VARIABLES	150
<i>4.6.1 Observed Grouping Variables</i>	<i>151</i>
<i>4.6.2 Children's Capabilities</i>	<i>154</i>
4.7 PROCEDURE	161
<i>4.7.1 Data Access</i>	<i>161</i>
<i>4.7.2 Project Consultation</i>	<i>162</i>

4.7.3 Analytical Process.....	162
4.8 SUMMARY.....	176
CHAPTER 5 PHASE 1: METHODS AND FINDINGS.....	178
5.1 INTRODUCTION.....	178
5.2 ANALYSIS.....	181
5.3 RESULTS.....	183
5.3.1 Group Enumeration.....	186
5.3.2 Optimal Solution.....	206
5.4 DISCUSSION.....	209
5.5 IMPLICATIONS.....	211
5.6 LIMITATIONS AND FUTURE DIRECTIONS.....	214
5.7 CONCLUSION.....	217
CHAPTER 6 PHASE 2: METHODS AND FINDINGS.....	218
6.1 SUMMARY OF MI IN AVAILABLE MEASURES ACROSS THE CHILD'S GENDER AND ETHNIC GROUPS.....	220
6.2 TESTING MI OF THE PERSON-CENTRED MODEL ACROSS THE CHILD'S GENDER AND ETHNIC GROUPS.....	222
6.3 ANALYSIS.....	225
6.4 RESULTS.....	229
6.4.1 Multivariate Analysis of the Optimal 3-Group Model, When Not Accounting for the DIF Effects in Model Indicators.....	229
6.4.2 Testing Gender as a Source of the DIF Effects in the Latent Profile Model.....	231
6.4.3 Testing Ethnicity as a Source of the DIF Effects in the Latent Profile Model.....	240
6.5 DISCUSSION.....	247
6.6 IMPLICATIONS.....	253
6.7 LIMITATIONS AND FUTURE DIRECTIONS.....	254

6.8 CONCLUSION	256
CHAPTER 7 GENERAL DISCUSSION	258
7.1 OVERVIEW AND SUMMARY OF SIGNIFICANT FINDINGS	259
<i>7.1.1 Profiles Observed.....</i>	<i>262</i>
<i>7.1.2 MG-CFA Approach.....</i>	<i>264</i>
<i>7.1.3 DIF Approach to the LPA Model.....</i>	<i>266</i>
<i>7.1.4 Overall Discussion.....</i>	<i>268</i>
7.2 IMPLICATIONS	277
<i>7.2.1 Methodological and Practical Implications.....</i>	<i>277</i>
<i>7.2.2 Research Implications.....</i>	<i>284</i>
<i>7.2.3 Policy Implications.....</i>	<i>285</i>
<i>7.2.4 Theoretical Implications.....</i>	<i>286</i>
7.3 LIMITATIONS OF THE RESEARCH AND FUTURE DIRECTIONS.....	288
<i>7.3.1 Summary of Limitations.....</i>	<i>289</i>
<i>7.3.2 Future Directions.....</i>	<i>293</i>
7.4 CONCLUSION	295
APPENDICES	299
8.1 APPENDIX 1. FULL ETHICS APPROVAL	299
8.2 APPENDIX 2. ETHNICITY PRIORITISATION IN CURRENT RESEARCH.....	300
<i>8.2.1 Introduction</i>	<i>300</i>
<i>8.2.2 Current Research.....</i>	<i>304</i>
8.3 APPENDIX 3. ASSESSING YOUNG CHILDREN'S SOCIAL AND EMOTIONAL CHARACTERISTICS USING THE PROSOCIAL BEHAVIOUR AND EMOTIONAL SYMPTOMS SUBSCALES OF STRENGTHS AND DIFFICULTIES QUESTIONNAIRE.....	309
<i>8.3.1 Introduction</i>	<i>309</i>

8.3.2 Validity of the Five-Factor Structure of the SDQ.....	311
8.3.3 Validation of the Parent-Reported SDQ in the GUiNZs.....	312
8.3.4 Current Research.....	313
8.3.5 Method.....	315
8.3.6 Results.....	318
8.3.7 Conclusion.....	325
8.4 APPENDIX 4. FACTOR STRUCTURE OF THE AFFECT KNOWLEDGE TEST (AKT): APPLICATION TO YOUNG CHILDREN'S EMOTION KNOWLEDGE AND EMOTIONAL EXPRESSION IN THE NEW ZEALAND CONTEXT.....	327
8.4.1 Introduction.....	327
8.4.2 Construct Validity of the AKT.....	330
8.4.3 Current Research.....	331
8.4.4 Method.....	332
8.4.5 Results.....	333
8.4.6 Conclusion.....	341
8.5 APPENDIX 5. THE PARENT RATING OF ORAL LANGUAGE (PROL): APPLICATION TO YOUNG CHILDREN'S EXPRESSIVE COMMUNICATION.....	343
8.5.1 Introduction.....	343
8.5.2 Method.....	344
8.5.3 Results.....	346
8.5.4 Conclusion.....	351
8.6 APPENDIX 6. THE ADAPTED PEABODY PICTURE VOCABULARY TEST-THIRD EDITION (PPVT-III): APPLICATION TO YOUNG CHILDREN'S RECEPTIVE VOCABULARY AND COMMUNICATION IN THE NEW ZEALAND CONTEXT.....	352
8.6.1 Introduction.....	352
8.6.2 Format and scoring of the PPVT-III.....	353
8.6.3 Validity of the PPVT-III as an Assessment of Vocabulary Among Children.....	353

8.6.4 Cultural Biases on PPVT-III.....	355
8.6.5 The Adapted PPVT-III in the GUiNZs.....	357
8.6.6 Current Research.....	360
8.7 APPENDIX 7. THE ADAPTED PENCIL TAP TASK-HAND CLAP TASK: APPLICATION TO YOUNG CHILDREN'S EXECUTIVE FUNCTION IN THE NEW ZEALAND CONTEXT	366
8.7.1 Introduction	366
8.7.2 Current Research.....	368
8.8 APPENDIX 8. PROFILE MEMBERSHIP TRANSITIONS IN CHAPTER 5.....	373
8.8.1 Profile Membership Transitions Between Profiles in Variances Fully Constrained Models.....	373
8.8.2 Profile Membership Transitions Between Profiles in Variances Freed Models.....	374
REFERENCES.....	375

Abbreviations and Terms

The following is a list of abbreviations used throughout the thesis. Each abbreviation is presented in full the first time it appears in a chapter, along with the abbreviation.

AKT	Affect Knowledge Test
CFA	Confirmatory factor analysis
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
DIF	Differential item functioning
ECE	Early childhood education
GUiNZs	Growing Up in New Zealand study
LCA	Latent class analysis
LPA	Latent profile analysis
MG-CFA	Multiple-group confirmatory factor analysis
MI	Measurement invariance
MIMIC	Multiple Indicators Multiple Causes
MNI	Measurement non-invariance
PPVT_III	Peabody Picture Vocabulary Test-III
PROL	Parent Rating of Oral Language
SDQ	Strengths and Difficulties Questionnaire

The thesis is written in English; however, it incorporates some integration of common Te Reo Māori kupu (i.e., words from the Māori language) that are commonly

used in early childhood education in New Zealand. Specifically, this includes the word *whānau*, which signifies family, characterised by a family's multi-layered, dynamic, and inclusive nature. The term *whānau* fosters a sense of group unity, extending immediate family to close friends and much wider family (Metge, 1995). Other Te Reo terms include the names of key documents, policies, or frameworks that are common in early childhood education in New Zealand. These are the proper names of such documents, and direct translation is not provided. The most commonly mentioned document is *Te Whāriki*, which is the early childhood curriculum that provides a bicultural framework for early childhood education services in New Zealand (Ministry of Education, 2017).

This study is solely based on the GUiNZs 54-month dataset and employs a cross-sectional design, capturing children's capabilities at a single time point. Therefore, the thesis does not examine longitudinal or developmental trajectories, and references to "development" are interpreted as describing children's current capabilities rather than processes of change over time.

List of Tables

Table 3-1 <i>Summary of Selected Person-Centred Studies on Developmental Characteristics in Children Aged 4 to 6.5 Years</i>	105
Table 4-1 <i>Summary of Research Aims and Questions</i>	143
Table 4-2 <i>Sociodemographic Characteristics of the Overall Sample and Reduced Sample Included in LPA and MIMIC Modelling</i>	153
Table 4-3 <i>Overview of Analytical Process</i>	165
Table 4-4 <i>Overview of Measures and Scores Used by Measures in Each Step</i>	166
Table 5-1 <i>Descriptive Statistics for Scores on the Indicators</i>	185
Table 5-2 <i>Correlations Between Scores of the Selected Indicators</i>	185
Table 5-3 <i>Fit Statistics and Group Proportions for 1- to 6-Group Models With Variances Fully-Constrained of Children's Capabilities at 54-Month Assessment (n =5707)</i>	188
Table 5-4 <i>Fit Statistics and Group Proportions for 1- to 6-Group Models With Variances Freed of Children's Capabilities at 54-Month Assessment (n =5707)</i>	200
Table 6-1 <i>Models Assessing Associations of Gender and Five Ethnic Groups With Group Membership From Multinomial Regression</i>	230
Table 6-2 <i>Model Comparison for Stepwise DIF Tests for the Child's Gender With the 3-Group Latent Profile Model of Young Children's Capabilities</i>	236
Table 6-3 <i>Model Comparison for Stepwise DIF Testing for Ethnicity With the 3-Group Latent Profile Model of Young Children's Capabilities</i>	243
Table 6-4 <i>Summary of Findings of Uniform and Nonuniform DIF Effects Using the</i>	

<i>Child's Gender and Ethnicity as Covariates.....</i>	252
Table 7-1 <i>Three Person-Centred Studies Using the GUiNZs 54-Month Dataset.....</i>	270
Appendix 2 Table 1 <i>Ethnicity Categorisation Based on Mother-Prioritised Child's Ethnicity in the Current Research.....</i>	307
Appendix 3 Table 1 <i>Items of the Prosocial and Emotional Subscales in the SDQ With the Number of Response Endorsements by Item (n = 6156).....</i>	318
Appendix 3 Table 2 <i>Unstandardised Coefficients and Items of the Five Items One- Factor Models of the SDQ Prosocial Behaviour and Emotional Symptoms Subscales.</i>	320
Appendix 3 Table 3 <i>Model Fit Indices for CFA and MG-CFA of the SDQ Prosocial Behaviour Subscale.....</i>	321
Appendix 3 Table 4 <i>Model Fit Indices for CFA and MG-CFA of the SDQ Emotional Symptoms Subscale</i>	322
Appendix 4 Table 1 <i>Items and the Number of Responses Endorsements by Item in the Included Sample (n = 5605).....</i>	334
Appendix 4 Table 2 <i>Items and Factor Loadings of the One-Factor CFA Models of the 6 Interviewer-Observed AKT Items in the GUiNZs 54-Month Dataset.....</i>	337
Appendix 4 Table 3 <i>Model Fit Indices for One-Factor Models of the Six-Item AKT Administered in the GUiNZs 54-Month Dataset.....</i>	338
Appendix 4 Table 4 <i>Model Fit Indices for the One-factor AKT (M3) with Multi-Group Invariance Comparison by Children's Gender and Ethnicity</i>	340
Appendix 5 Table 1 <i>Items and Response Options for the Parent Rating of Oral Language (PROL) With the Proportion and Number of Endorsements by Item Response</i>	

<i>Option in the Included Sample (n = 6133)</i>	346
Appendix 5 Table 2 <i>Items and Factor Loadings of the Five Items One-Factor Model of the PROL</i>	349
Appendix 5 Table 3 <i>Model Fit Indices for the One-Factor Model of the PROL Items Administered in the GUiNZs 54-Month Maternal-Report Assessment</i>	350
Appendix 6 Table 1 <i>Frequency of the PPVT-III Across Children's Gender and Mother-Prioritised Single-Combined Ethnicity</i>	362
Appendix 7 Table 1 <i>Summary Statistics for the Raw Scores of the Adapted Luria Hand Clap Task</i>	370
Appendix 8 Table 1 <i>Changes in Numbers of Participants in 3-Group Versus 4-Group Variances Fully-Constrained Models (n)</i>	373
Appendix 8 Table 2 <i>Changes in Numbers of Participants in 4-Group Versus 5-Group Variances Fully-Constrained Models (n)</i>	373
Appendix 8 Table 3 <i>Changes in Numbers of Participants in 3-Group Model Versus 4-Group Model With Variances Freed (n)</i>	374
Appendix 8 Table 4 <i>Changes in Numbers of Participants in 4-Group Model Versus 5-Group Model With Variances Freed (n)</i>	374

List of Figures

Figure 4-1 <i>Methodological Process</i>	164
Figure 4-2 <i>Visualisation of the LPA Conceptual Model in Phase 1</i>	170
Figure 5-1 <i>Visualisation of the LPA Conceptual Model in Phase 1</i>	183
Figure 5-2 <i>The Elbow Plot of Sample-Size Adjusted Bayesian Information Criterion (SABIC) From Unconditional Latent Profile Models With Variances Fully Constrained of Children's Capabilities in the GUiNZs 54-Month Dataset</i>	187
Figure 5-3 <i>Clustered Column With Standardised Scores of the 2-Group Model of Children's Capabilities Across Six Indicators</i>	190
Figure 5-4 <i>Clustered Column With Standardised Scores of the 3-Group Model Across Six Indicators</i>	192
Figure 5-5 <i>Clustered Column With Standardised Scores of the 4-Group Model Across Six Indicators</i>	194
Figure 5-6 <i>Clustered Column With Standardised Scores of the 5-Group Model Across Six Indicators</i>	196
Figure 5-7 <i>Clustered Column With Standardised Scores of the 6-Group Model Across Six Indicators</i>	198
Figure 5-8 <i>The Plot of Sample-Size Adjusted Bayesian Information Criterion (SABIC) From Unconditional Latent Profile Models With Variances Freed of Children's Capabilities in the GUiNZs 54-Month Dataset</i>	199
Figure 5-9 <i>Clustered Column With Standardised Scores of the 2-Group Model With</i>	

<i>Variances Freed Across Six Indicators</i>	202
Figure 5-10 <i>Clustered Column With Standardised Scores of the 3-Group Model With Variances Freed Across Six Indicators</i>	203
Figure 5-11 <i>Clustered Column with Standardised Scores of the 4-Group Model With Variances Freed Across Six Indicators</i>	205
Figure 5-12 <i>Clustered Column With Standardised Scores of the 3-Group Model Across Six Indicators</i>	208
Figure 6-1 <i>Model-Estimated Latent Profile Factor Distribution by Gender for Factors With DIF Effects</i>	238
Figure 6-2 <i>Model-Estimated Latent Profile Factor Distribution by Ethnicity for Factors With DIF Effects</i>	245
Appendix 3 Figure 1 <i>Measurement Models for Parent-Report Prosocial and Emotional Symptoms SDQ Items in the GUiNZs 54-Month Dataset</i>	319
Appendix 7 Figure 1 <i>Distribution of the Luria Hand Clap Task Count Scores in the GUiNZs 54-month Dataset</i>	370
Appendix 7 Figure 2 <i>Distribution of the Luria Hand Clap Task Count Scores by the Child's Gender</i>	371
Appendix 7 Figure 3 <i>Distribution of the Luria Hand Clap Task Count Scores by Mother-Prioritised the Child's Ethnicity</i>	372

Chapter 1 Thesis Introduction

Unlocking the immense potential of every child requires a robust understanding of their learning and development. Knowledge about children's learning and development is relevant for families, educators, researchers, and policy-makers, with each stakeholder group able to impact children's learning and development in various ways. The present research is targeted at researchers and policy-makers and has implications for families and educators.

In New Zealand, the early childhood curriculum *Te Whāriki* expresses the vision that children grow up “as competent and confident learners, strong in their identity, language and culture” (Ministry of Education, 2017, p. 2). Within *Te Whāriki*, the development of children's capabilities is viewed holistically and understood to be influenced by multiple factors. Within education, young children's learning and development are increasingly being addressed from a strengths-based, holistic, and culturally connected perspective. Yet, the research base to understand children's learning and development has historically focused on developmental delays or risks associated with domain-specific deficits in children's learning and development, and mostly, without accounting for or acknowledging children's identity, language, and culture. This is often paralleled in policy initiatives and funded interventions focusing on domain-specific programmes to support children's learning and development. The present research is part of the shifting tide to explore methods and approaches to better understand and account for children's learning and development in strengths-

based, holistic, and culturally connected ways.

1.1 Research Overview

The learning and characteristics explored in the present research comprised various affective and cognitive domains, including children's social, emotional, language, and executive function skills. The principal aim of this research was to identify shared patterns of young children's learning and development, focusing on the interconnection of social, emotional, language, and executive function capabilities, within a diverse population. Moreover, the research sought to derive unbiased estimates for these aforementioned capabilities, as well as the shared patterns of young children's learning and development across distinct groups, such as the child's gender and ethnicity. The research processes were designed and conducted in ways intended to mitigate the risk of misspecifications and misinterpretations of the identified patterns across groups.

The research is a cross-sectional study utilising secondary analyses of data from the longitudinal Growing Up in New Zealand study (GUiNZs). It focused on young children before school entry using data from the GUiNZs 54-month dataset (i.e., children assessed during Data Collection Wave 5 (DCW5)). GUiNZs is a birth cohort study recruited to be broadly representative of the New Zealand birth (Morton et al., 2015), comprising responses from over 6,000 preschool children and parents at DCW5. DCW5 data was collected from parents (mothers) and direct child observations. The present study used a range of mother-reported and direct observation measures related to

children's social, emotional, language, and executive function skills.

The research was conducted through a two-part analysis process in which 1) the shared patterns of children's learning and development were explored across the overall sample, and 2) the consistency of measurement properties and identified patterns were assessed across the child's gender and ethnic groups within the diverse sample.

The first part of the work utilised confirmatory factor analysis (CFA) to validate the structure of the selected measures for analysis in the present research. Following confirmation of factor structure, measures were used as indicators of children's learning and development to explore shared patterns using latent profile analysis (LPA), which is known as a person-centred approach to data analysis. A person-centred approach recognises that people may share combinations of characteristics that meaningfully cluster together. The second part of the work utilised multiple-group confirmatory factor analysis (MG-CFA) for measurement invariance (MI) tests of the selected measures to examine the consistency of measurement properties across the child's gender and ethnic groups within the diverse sample. The results of invariance tests for the selected measures were employed to support the subsequent estimation of potential biases from the covariates (i.e., child's gender, ethnicity) to the latent profile factors through Multiple Indicators Multiple Causes (MIMIC) modelling, as well as to investigate the comparability and consistency of the shared patterns across the child's gender and ethnic groups.

This approach to analysis was designed to support a holistic approach to understanding children's learning and development across social, emotional, language, and executive function skills. Interpretations of shared patterns aimed to use a strengths-based approach to describing children's emerging capabilities. The focus on consistency of measures and the identified shared patterns across the child's gender and ethnic groups was intended to support fairness in the interpretation and application of findings, thereby fostering equity and inclusion within early childhood education research in New Zealand.

1.2 Researcher Background

As I approach the destination of my PhD journey, I find myself reflecting on my identity as a researcher. Despite nearing the finish line, I still hesitate to fully embrace this role due to the perceived gaps in my knowledge. Nonetheless, this journey has profoundly enriched my personal and early academic life.

My enthusiasm for children's behaviour initially sparked my research interest. From early on, I was captivated by the intricacies of prosocial and "problem" behaviours in young children. These behaviours, rooted in social norms and morality, highlight the fundamental role of social interaction and community in human development, particularly during early childhood, a crucial period of brain development. Delving into these ideas led me to explore multiple definitions and theoretical frameworks of prosocial behaviour, notably from scholars like Nancy Eisenberg (Eisenberg et al., 2015). Their work shed light on the diverse forms of prosocial behaviour, such as helping,

sharing, and comforting, and underscoring the complexity of prosocial behaviour arising from the intricate interplay of diverse factors, including biological, psychological, social, and situational influences. However, measures for prosocial behaviours are limited in the GUINZs; for example, the prosocial behaviour subscale in the Strengths and Difficulties Questionnaire (SDQ) offers only a narrow range of prosocial behaviours.

My exploration also underscored the interconnectedness of various domains—such as language and executive function—with prosocial behaviour. Research using a variable-centred approach often reveals broad associations and trends within populations, yet it overlooks individual complexities and heterogeneous populations (such as individuals with diverse characteristics). Motivated by a desire to delve deeper, I was fascinated by the person-centred approach to data analysis. This holistic analytic method not only identifies subgroups within populations but also offers a nuanced understanding of individual behaviours and needs.

Under the guidance of supportive supervisors, I embarked on a journey to enhance my methodological and analytical skills for large-scale datasets. In my research, using large-scale datasets from the GUINZs has been pivotal and challenging. Leveraging these datasets has enabled me to explore sophisticated statistical analyses, though navigating their complexities—like data management and software proficiency—has been a steep learning curve.

Despite the challenges, the journey has been immensely rewarding. It has shaped not only my research but also my personal growth. The process has equipped me with

invaluable insights that will undoubtedly influence my future endeavours.

1.3 Research Rationale

The present research is multifaceted. The following sections outline key points about: the importance of social, emotional, language, and executive function capabilities; holistic learning and development; diversity in learning and development; current and past research to understand children's learning and development; and the policy context for early learning. Collectively, these create the rationale for the present research.

1.3.1 Social, Emotional, Language and Executive Function Capabilities

Exploration of young children's learning and development before school entry is important in the early years, as it is a critical period for brain development that lays the foundation for children to reach their full potential (Berk, 2016). Specifically, the present research focuses on young children's social, emotional, language and executive function capabilities. Each of these capabilities is complex, and the ways they are understood, described, and examined differ across the literature. For the present research, I use the term social competence, which refers to children's capabilities and effectiveness in social interactions (Rose-Krasnor, 1997). I also use the term emotional competence, which refers to capabilities in emotional knowledge, emotional expression, and emotional regulation (Denham et al., 2016; Saarni, 1999a). Language skills refer to the ability to understand, express, and communicate thoughts with others verbally and non-verbally (*APA Dictionary of Psychology*, 2007). Executive function skills refer to a

set of cognitive processes that manage and regulate ongoing, goal-directed behaviours, with a focus on effortful control (Berk, 2013; Diamond, 2013). These skills were selected for the present research because they are key aspects of children's learning and development that are associated with later life well-being (Cole et al., 2010; Diamond, 2016; McLaughlin, 2017; McLeod et al., 2019; Ministry of Education, 2019a).

1.3.2 Holistic Learning and Development

While understanding individual domains of learning (e.g., social, emotional, language, and executive function) is important for ensuring clarity and allowing researchers to assess and reflect on these constructs effectively, educators and researchers recognise that developmental changes across domains are profoundly interrelated and evolve in complex ways throughout different stages of learning (Lewis, 2000; McLachlan, 2018; Ministry of Education, 2017). Thus, changes in children's capabilities are interrelated and holistic. This necessitates an integrated, holistic view that acknowledges multifaceted growth and learning. Such holistic views recognise children as complete entities and underscore the dynamic interaction between various aspects of children's learning and development.

Theories of holistic learning and development acknowledge the significance of each domain and the interwoven nature of developmental characteristics (Johnson, 2023; Magnusson, 1995). To account for the integrated, dynamic, and multifaceted nature of children's learning and development, educators and researchers often explore

profiles of children's learning and development to understand shared patterns of similarities and distinctions within and across children (Bergman et al., 2003; Magnusson, 1995). This approach to learning and development can also be connected to an ecological perspective to account for the broader environment and context for children's learning and development (Bronfenbrenner, 1986; Bronfenbrenner & Ceci, 1994). Taken together, a holistic and ecological view helps inform the design and delivery of effective early childhood education (ECE) to develop and support children's capabilities.

1.3.3 Diversity in Learning and Development

Children's diverse identities are recognised as important within contemporary social and educational contexts. The present research identified gender and ethnicity as key aspects of children's emerging identities, which may have educational relevance in how learning and development are examined and understood. Both gender and ethnic identities have implications for supporting equitable educational outcomes (Bird et al., 2024; McLaughlin et al., 2024). Research that addresses these aspects of children's identity may help support equity and inclusive practice in diverse contexts. While essential to acknowledge and account for, it is also complex to classify gender and ethnicity in meaningful and appropriate ways for research. For the present research, mothers selected the main ethnicity for their child (i.e., mother-prioritised ethnicity for child), with options including Māori, Pacific Peoples, Asian, European, Other, and dual ethnicities. The current research also acknowledges the multidimensional nature of the

gender construct; however, given the prevalent interchangeable use of gender and biological sex in research, gender in this study was operationalised as binary categories of 'boys' and 'girls', inferred based on the child's sex assigned at birth. The advantages and limitations of these approaches to addressing gender and ethnicity will be explored in Chapters 3 and 4.

1.3.4 Research to Understand Children's Learning and Development

Research acknowledges the importance of children's learning and development. Most studies focus on examining aspects of children's capabilities through a single domain focus in which skills such as social, emotional, language or executive function are examined in isolation from each other or to understand the association between sets of skills (Denham et al., 2013; Neumann et al., 2023; Zava et al., 2024). This is sometimes described as a variable-centred approach to research, in which associations between domains of interest in a population or the relative contribution of predictors to outcomes are a key focus of the research (Bauer & Shanahan, 2007; Laursen & Hoff, 2006). In contrast, the present research builds on the emerging body of research that examines the interconnectedness of children's learning from a person-centred approach. A person-centred approach refers to the identification of subgroups of populations with shared patterns of characteristics that can then be used to explore other associations or predictions of outcomes (Bergman et al., 2003; Laursen & Hoff, 2006). The use of person-centred analyses in the current research acknowledges children's social, emotional, language and executive function capabilities from a holistic

and dynamic perspective. The implications of this type of research are intended to inform the development of targeted and appropriate services and programmes that account for the full range of children's capabilities. This may be particularly important for children with specific needs, for whom the identification of relative strengths and needs may enhance the types of resources and support provided.

Further, there is a growing trend in research to examine and account for children's diverse identities in fair and appropriate ways. This includes moving beyond a focus on validity and reliability in measurement to consider fairness and appropriate interpretation of results. In the present study, this is done through approaches that examine whether constructs represented by a measure or model are assessed in equivalent ways across different groups or time points, with such assessments referred to as MI testing. Studies conducting MI tests aim to mitigate the risks of biases from measures by confirming that the observed differences between groups are due to actual variations in the constructs measured, rather than differences in how the measures function (i.e., what they measure) across these groups (Kim et al., 2017; Millsap, 2011; Vandenberg & Lance, 2000). Such biases may arise when response options are interpreted differently or elicit varying responses across different groups (McGovern & Lowe, 2018). Biased measures can result in inaccurate assessments and misguided interpretations, potentially leading researchers to draw unwarranted or incorrect conclusions (Oort et al., 2009). These findings underscore the importance of diversity and sensitivity and guide the development of culturally appropriate and

targeted services and programmes that support young children's learning and development.

1.3.5 Policy Context for Early Learning

During the period the present research was conducted, the leading policy statement for ECE in New Zealand has been the *He taonga te tamaiti*: Every child a *taonga*: Early learning action plan 2019-2029 (Ministry of Education, 2019b). In the action plan, the development of the ECE priorities is guided by the vision that every child has innate potential, and high-quality early learning services benefit children, families and whānau (Ministry of Education, 2019b). The action plan intends that funding and regulatory policies support a well-coordinated ECE system, promoting participation and success by fostering inclusive, culturally supportive early learning environments where qualified, diverse educators help every child thrive and develop key capabilities. However, challenges in policy construction and implementation within ECE have long been impacted by changes in government and changing government priorities (Haggerty & Alcock, 2016; May, 2014). For example, the previous early learning strategic plan, *Pathways to the Future: Ngā Huarahi Arataki* (Ministry of Education, 2002) developed under a Labour government as a 10-year strategic plan, was later retracted under a National government. The fate of the early learning action plan 2019-2029 is currently unknown, with the recent changes in government.

The different approaches that governments take to ECE policy construction and implementation often stem from a varying understanding of children and early

learning, inclusive education, fairness, and appropriateness in education settings (Hardy & Woodcock, 2015), as well as the mix of competing social and economic priorities.

These issues reveal deeper concerns about values, purposes, and practices within ECE systems (May, 2014).

Nonetheless, effective policies are important for creating a robust ECE system that benefits children, families, and society by ensuring a solid foundation of lifelong, life-wide learning and development, as well as social and economic integration. As dynamic and living documents, policy formation draws from a range of data and research. In New Zealand, research such as the GUiNZs and the Dunedin Multidisciplinary Health and Development study provides valuable insights for policy decision-making. Conversely, such funded research often mirrors policymakers' priorities in addressing health, social, educational, and economic challenges in ECE. Government funding supports initiatives and programmes designed to bridge gaps and improve outcomes.

Evaluating the effectiveness of these policies that come from research findings requires addressing both the technical and normative constraints in measuring outcomes (Fullerton, 2021). A relevant example in the New Zealand context can be seen in the recent government investment in the ENGAGE programme (Healey & Halperin, 2015), highlighting social-emotional learning with a focus on self-control. This initiative was supported as part of the government's priority on social cohesion and inclusion of New Zealand's diverse communities and aligns with the Child and Youth Well-being Strategy policy (the Department of the Prime Minister and Cabinet, 2019).

Research to support the adoption of this programme has been drawn from the self-control research findings from the Dunedin Study (Moffitt et al., 2013). The justification for the programme reflects the research on the impact of self-regulation on lifelong outcomes (Healey & Healey, 2019; Healey et al., 2022). Notably, this research was conducted from a variable-centred approach focusing solely on self-regulation and does not specifically account for other aspects of children's learning, which may exaggerate the role of self-control in the broader context of children's learning and development. This is notable as the government has invested over 20 million dollars into supporting the programme in an effort to improve children's self-control. The effectiveness of these types of domain-specific programmes in ECE on lifelong outcomes remains uncertain and requires long-term validation.

It is noteworthy that current government policy remains informed by the Dunedin Study, despite the more contemporary research supported by the GUINZs data and more recent insights into children's learning and the constructs associated with self-control, more commonly referred to as self-regulation (Morton et al., 2017). Recent research in the GUINZs highlights that self-control in children is neither static nor fixed, but rather dynamic and variable, particularly during the early years (Morton et al., 2020). To advance modern policy development, it would be beneficial to incorporate and prioritise recent studies, as noted by Bird et al. (2016). However, few studies in GUINZs to date have used person-centred methods to examine children's learning and development holistically (Corkin et al., 2022; McLaughlin et al., 2024; Russell et al.,

2022). Consequently, there is a need for a person-centred approach to explore shared patterns of young children's capabilities.

Building on work from McLaughlin et al. (2024) and Russell et al. (2022), the present research endeavours to demonstrate a person-centred approach to understanding children's learning and development that considers the diverse capabilities and identities of all preschool children. This type of approach to population data may help inform New Zealand policymakers in designing initiatives that support children's learning holistically and from a strengths-based perspective. The present research also highlights the need for policies that address children's unique identity, language, and culture.

1.4 Introduction Summary

Young children's social, emotional, language, and executive function capabilities are important in children's learning and development and influence lifelong and life-wide outcomes (Moffitt et al., 2011). The current research explored these capabilities from a holistic, strengths-based view and accounted for emerging gender and ethnic identities from a person-centred approach. Specifically, secondary analyses of the GUiNZs 54-month dataset were conducted to explore shared patterns of young children's social, emotional, language and executive function capabilities using a latent profile analysis, followed by the exploration of similarities and distinctions of these shared patterns across children's diverse identities (e.g., gender, ethnicity) by examining MI on the measures and model level across children's gender and ethnic groups.

The current research endeavours to offer insights into the latent profiles of young children's capabilities in New Zealand that are consistent with the GUINZs sample. The study also aims to provide insights into how the measures used to assess children's capabilities and the latent profiles generated from the research are comparable across groups by the child's gender and ethnicity. The research may help guide policy recommendations and contribute to future research, including research that examines how these capabilities develop in a patterned way over time, and advance future research examining how these capabilities have developed over time.

1.5 Thesis Structure

This thesis is structured into seven chapters. Following Chapter 1, "Thesis Introduction", the subsequent chapters are organised as follows:

Chapter 2, "Literature Review", extensively reviews the literature on social competence, emotional competence, language, and executive function in young children and their interindividual differences. This chapter is the discussion of each domain, divided into five sections: (1) the importance of each capability, (2) relevant theoretical explanations, (3) definitions of these capabilities, (4) an in-depth review of the domains of each capability, followed by (5) how each domain is assessed more generally. Each section concludes with a summary of its contents and underscores the study's significance in contributing to a comprehensive understanding of young children's capabilities.

Chapter 3, "Holistic Views of Development and Observed Subgroup Variations",

commences by considering the interdependent and holistic nature of young children's capabilities, adopting a holistic perspective. It delves into the homogeneity and heterogeneity of interindividual capabilities, exploring the potential for shared patterns (homogeneity) and variability manifested in interindividual differences and by known subgroups (heterogeneity). This chapter then examines key methodologies for clustering young children's characteristics, emphasising a person-centred, holistic approach and explaining why the subgroups of gender and ethnicity are important considerations in early childhood.

Chapter 4, "Methodology", begins with an introduction to the GUiNZs, contextualising its relevance to child development research. It then introduces the current research, outlining the research questions, hypotheses, sample, and ethical considerations that guide the current thesis. It also details the research procedure, including data access, project consultation and the analytical process. The chapter concludes with an overview of the methods employed and their integration within each phase of the study.

Chapter 5, "Phase 1: Methods and Findings ", explores the shared patterns of young children's social, emotional, language and executive function capabilities using latent profile analysis (LPA). This phase of the research builds upon prior data due diligence by assessing the construct validity of each selected measure in Appendices 3 to 7. The objective is to examine whether distinct latent subgroups—characterised by shared patterns of children's social, emotional, language and executive function

capabilities—can be identified in the 54-month dataset in GUiNZs.

Chapter 6, “Phase 2: Methods and Findings”, first examines the measurement invariance (MI) on the constructs of the selected measures followed by assessing the invariance of the latent subgroups through a differential item functioning (DIF) approach using the MIMIC modelling technique, aiming to assesses potential biases from covariates (i.e., the child's gender, ethnicity) to the latent profile factors, and address the knowledge gap regarding gender and ethnic differences in young children's capabilities within the New Zealand context.

Chapter 7, “General Discussion”, provides a summary of the significant findings derived from the research and integrated with the literature to describe how these findings enrich the comprehension of the shared patterns observed in young children's capabilities related to social, emotional, language and executive function. This chapter also discusses the limitations and implications inherent in the research. The implications offer insights into the methodological, practical, research, policy, and theoretical implications of the research. Possible directions for future research are outlined.

Chapter 2 Literature Review

To help children reach their full potential, a comprehensive understanding of their learning and development is needed. Children's learning and development, explored in the present research, involve various cognitive and affective elements, including executive function, language skills, social competence, and emotional competence. This chapter aims to explore and describe the intricacies of these four domains of development and establish a foundation for how they are understood and approached in the current research.

This chapter unfolds in a structured manner, with each capability presented in four distinct sections followed by a chapter summary. Each section weaves in contemporary literature, shedding light on the pivotal role of foundational skill sets in shaping young children's social and emotional competence, language, and executive function capabilities. Each section begins with an introduction to the importance of these capabilities. This is followed by an exploration of relevant child development theories to support the theoretical understanding of each capability. Definitions and relevant skill sets within each capability are described. This leads to an exploration of key issues related to the assessment of each capability. The chapter begins with reviews of social and emotional competence, followed by an exploration of language skills and executive function.

2.1 Social Competence

Social competence refers to capabilities and effectiveness in social interactions

(Rose-Krasnor, 1997). This section outlines evidence regarding the importance of social competence and key theories about social development in young children. The section then discusses definitions of social competence and key subset skills. The section ends with the assessment of social competence, highlighting the strengths of prosocial behaviours as an important domain of assessing social competence among younger children.

2.1.1 Importance of Social Competence

Social competence is crucial to young children's social interactions, self-esteem, transition to school, school readiness and academic success, and lifelong outcomes from the very early years of life (Cacioppo, 2002; Denham, 2006; Denham et al., 2013; Hamre & Pianta, 2001; Ladd et al., 1999; McLaughlin, 2017; Ministry of Education, 2019a; Phillips & Shonkoff, 2000). Social competence supports young children in building up and maintaining relationships and communicating with others. By ages 5-6, children's social interactions typically progress with extended collaborative social play and acquired social skills (Bierman & Erath, 2006). Socially competent children have higher levels of school readiness. For example, when entering school, socially competent children experience more positive development in relationships with peers and teachers, and adjustment to school (Birch & Ladd, 1998; Thompson & Raikes, 2007; Zava et al., 2024). In addition, children's social competence, accepted peer status and social popularity are linked to their early academic participation and better academic achievement (Denham et al., 2013; DeRosier et al., 1994). Previous longitudinal studies

and extensive literature show that young children with better social skills are more likely to contribute positively to their communities, have better mental health and well-being in adulthood, and become lifelong learners (Goodman et al., 2015).

Lack of social competence has also been associated with social difficulties, externalising problems, and poor academic outcomes (Hanish et al., 2005; Johnson et al., 2000; Perren & Alsaker, 2006). Children with poor social skills may face peer rejection and develop conflicted child-teacher relationships (Bellanti & Bierman, 2000; Longobardi et al., 2021; Pianta & Stuhlman, 2004). Difficulties with peers, teachers, and family may adversely/bi-directionally reinforce children's externalising problems and impact children's attitudes towards school, which further results in school avoidance and decreased academic achievement (Gagnon et al., 1995; Kochenderfer & Ladd, 1996; Longobardi et al., 2021; Nurmi, 2012; Pianta & Stuhlman, 2004). In addition, lack of social skills and problems with peers may affect young children's mental health. For instance, cross-sectional studies indicate that depressive symptoms and internalising disorders in children can be predicted by self-oriented social skills involving social participation and assertiveness (Henricsson & Rydell, 2006; Perren & Alsaker, 2009).

2.1.2 Theories of Social Competence

In this section, I explore three key theoretical perspectives on children's social development. The key selected theories— attachment theory, social learning theory, and the Prism Model of Social Competence by Rose-Krasnor (1997)— provide conceptual insights into social competence. These perspectives delve into the interplay

between children's social interaction, social competence, and environmental influences in shaping social development. Each theory is derived from distinct theoretical traditions but collectively enriches the understanding of the mechanisms behind social development.

Attachment theory. Attachment is a deep emotional bond that emerges during the first year of life through interactions between children and their caregivers, and is dependent on caregivers' attunement and their capacity to respond in a sensitive and nurturing way (Allen, 2023; Keller, 2018). Four attachment categories identified by Bowlby, Mary Salter Ainsworth and Mary Main are secure, insecure-avoidant, insecure ambivalent and disorganised attachment (Ainsworth, 1979; Bowlby, 1982; Main & Solomon, 1986). Attachment is important because it is a child's first interpersonal relationship with their caregiver, providing a secure base to explore the social world and a safe haven for seeking comfort in times of uncertainty or distress, with a long-lasting impact on their emotional life and future relationships (Ainsworth, 1979). Attachment theory explains how early experiences with the primary caregiver shape children's social development, and early attachment experiences shape their understanding of themselves, others, and interactions, thereby influencing their responses to caregivers, peers, and other adults in various settings (Allen, 2023; Bowlby, 1982). Attachment theory also highlights that early social interaction with caregivers significantly shapes a child's social development trajectory. Specifically, a secure and

close attachment increases the chances of a satisfying relationship, and an insecure attachment may increase the risk of problems throughout life (Matias et al., 2014).

Social learning theory. As conceptualised by Albert Bandura, social learning theory provides a framework for understanding the processes by which individuals acquire new behaviours through observations and imitation of others (Bandura, 1978; Bandura, 1986; Ladd, 2007). Social learning theory highlights the importance of social contexts that influence behaviour development, especially at the formative stages of development. Social learning theory also underscores the significance of peers as powerful agents in the socialisation process. As one of the important roles that impact young children's social learning and formation of identity, peers are often viewed as influential models that a child may observe and imitate (Wen Xia et al., 2023). Interaction with peers may allow for collaborative learning and contribute to establishing social norms (Hartup, 1996; Hartup & Moore, 1990). Recognising the role of peer influence within social learning theory is important for educators, parents, and relevant stakeholders, as well as informing strategies that promote adaptive social learning (Dong Hwa & Juhu, 2003; Hartup, 1996).

In addition to the above larger theories of development associated with social development, the *Prism Model of Social Competence* by Rose-Krasnor (1997) is often used to understand the structure of social competence more specifically. In this model, there are three layers related to social interaction that represent both interpersonal and intrapersonal elements. These layers are: 1) a theoretical level that examines the overall

effectiveness of social interactions, 2) an index level that considers the effectiveness of social interactions from individual perspectives and 3) a skills level that focuses on interpersonal connectedness and the use of skills (Rose-Krasnor, 1997).

2.1.3 Definitions

While the importance of social competence and the necessity to support it in young children is well-established, greater consensus on the definition and description of social competence in the research literature would support the progress of this work. This section explores variations in definitions and descriptions of social competence and highlights critical social competence areas (described as subset skills) relevant to the present study. Implications of these definitions for construct measurement are also considered.

Definitions of social competence reflect three orientations: interpersonal relations, intrapersonal skills, and a combination of both. Researchers have defined social competence in interpersonal relations as the ability to select and carry out their interpersonal goals successfully and appropriately, engage effectively in complex interpersonal interactions, and understand people effectively (Guralnick, 1999). In contrast, definitions reflecting intrapersonal skills highlight the ways that distinctive individual characteristics help achieve personal goals and maximise the probability of producing, maintaining, or enhancing positive effects in social interaction, such as problem-solving, social engagement, and empathy (Rubin & Rose-Krasnor, 1992). Additionally, definitions incorporating interpersonal and intrapersonal skills focus on

interactions with others, integrating oneself into society, and internalising society's norms and values (Rubin et al., 2006).

Variations in these conceptual definitions reveal the complex multidimensional nature of social competence. These conceptualisations also highlight the critical role of the effectiveness of a child in social interaction with peers and adults, which has implications for the measurement of social competence (Dodge, 1985; Fabes et al., 2006; Rubin & Rose-Krasnor, 1992). Before exploring measurement issues, the next section examines the range of key subset skills.

2.1.4 Domains of Social Competence

The complex nature of social competence builds on integrated skills that are either directly observable behaviours or internal competencies, whilst higher-level components of social competence call for operationalised explanations of observable skills (Domitrovich et al., 2007; Han & Thomas, 2010). Observable socially competent behaviours involve prosocial behaviours central to maintaining peer relations and social initiatives (Dodge et al., 1986; Rose-Krasnor, 1997; Rydell et al., 1997; Sarason, 1981). Prosocial behaviours support other-oriented perspectives to satisfy the needs of others, and social initiatives reflect self-oriented competence to fulfil one's needs (Perren et al., 2012). Peer relations and prosocial behaviours are described in more detail in the sections that follow.

As a significant component of social initiatives, children's peer relations have been well described as children's relations/friendships with peers (Bukowski, 2003; Hartup &

Moore, 1990). Peer relations are considered an important factor in social development, providing children with extrafamilial opportunities to understand and explore the effects of their behaviours (Malti et al., 2012; Rubin et al., 2006; Rubin et al., 2011). The engagement of social play and the selected peer affiliations enable preschoolers to form early friendships and social networks with peers and form a substantial socialisation function in developing social competence (Daniel et al., 2015; Strayer & Santos, 1996). Such social interactions include taking turns, managing conflicts, and ensuring one's needs are met in a socially adaptive manner, in a range of contexts such as early childhood settings, homes, and playgrounds (Engelmann et al., 2012; Green & Rechis, 2006; Melis et al., 2016). Young children's peer interactions may be necessary for developing empathy and prosocial behaviours (Strayer & Roberts, 2004). More specifically, interaction with prosocial friends concurrently links to more prosocial behaviours among preschoolers (Chung-Hall & Chen, 2010; Eivers et al., 2012), whereas the emergence of problems with peers is linked to early conduct problems. For example, particularly aggressive children may seek out other aggressive children as friends who may see aggressive and antisocial behaviours as acceptable, which possibly results in continued shaping and amplification of aggressive behaviours and individual maladaptive development (Hanish et al., 2005; Snyder et al., 1997).

Thus, prosocial behaviours are critical to peer relations and are an essential indicator of young children's social competence. Eisenberg et al. (2015) defined prosocial behaviour as 'voluntary actions intended to benefit others'. Children with

prosocial behaviours are positively responsive to others' needs and welfare and exhibit behaviours such as helping, sharing, and engaging in efforts to comfort someone who seems distressed (Brownell et al., 2009; Eisenberg et al., 2015). Prosocial behaviours are also considered an important antecedent of peer acceptance, academic achievement, and a precursor of children's social adjustment (Collie et al., 2018). Specifically, young children engaging in prosocial behaviours have more academic success and enjoy more positive relationships with peers and teachers (Skalická et al., 2015). Additionally, young children's prosocial behaviours are positively associated with their future employment, educational and health outcomes (Jones et al., 2015).

Peer relations and prosocial behaviours both represent types of observable interpersonal skills, but researchers view them as distinct concepts that interact with and influence each other. Specifically, positive peer relations represent a key aspect of social competence, while prosocial behaviour is seen as a measure for assessing the quality of peer relationships. The interplay between peer relations and prosocial behaviours influences outcomes in both areas (Berndt, 2002). Children's prosocial behaviours can predict positive peer relationships (Sebanc, 2003; Vitaro et al., 1990).

2.1.5 Assessments of Social Competence

As noted above, a focus on the *effectiveness* of a child in social interaction with peers and adults, presents challenges for the measurement of social competence (Dodge, 1985; Fabes et al., 2006; Rubin & Rose-Krasnor, 1992). The conceptualisation of effectiveness relates to behavioural outcomes oriented by both long- and short-

term developmental needs (Waters & Sroufe, 1983). Such definitions lack specific implications for measurement, as it is challenging to determine which behaviours contribute to the effectiveness of social interaction. Different selection criteria may conversely lead to disagreements on conceptualisations. Using limited specific skills to reflect the flexible, multidimensional construct of social competence may be misleading, as these skills act as a portion of the whole construct. Tapping into highly situational and age-specific social competence subdomains (subset skills in the current work) may help better understand social competence and address measurement issues (Anderson & Messick, 1974; Waters & Sroufe, 1983).

Assessment in this way often focuses on the selection of social competence subset skills that are comparatively easy to alter or that are identified as malleable. The multifaceted nature of skills has resulted in a wide range of often adult-reported assessments of children's social competence. Researchers have also noted problems related to focusing assessment on the social competence subset skills. First, there are challenges in accurately and reliably measuring young children's social competence, including finding relevant measurement tools for this task. Second, different skills within the same subdomain may not consistently relate to each other, both within measurement tools and across different measures (Wigelsworth et al., 2010; Zimmer, 2017). Furthermore, measurement tools may focus on a deficit, rather than a strengths-based approach to assessment. In their review of social competence assessments, Zimmer (2017) noted that "the majority of these tools are not strengths-based:

developed in the psychopathological field for diagnostic purposes, they tend to identify problem behaviours and mental health disorders, rather than measuring social and emotional competency level” (p. 6).

While there is a wide range of measures that might be used to assess young children’s social competence, the most common measure in New Zealand has been the Strengths and Difficulties Questionnaire (SDQ) by Goodman (1997). Specifically, the parent-reported SDQ has been used as a measure in the Before School Check (B4SC). The total and subscale scores of the SDQ predict behavioural and psychosocial outcomes of general and clinical cohorts in New Zealand contexts (Clark et al., 2014; Cornforth et al., 2012; D’Souza et al., 2019b; Lambie & Krynén, 2017; Thompson et al., 2021). More information about the SDQ is available in the methods chapter.

2.1.6 Summary

To summarise, social competence is vital to young children’s learning and development. However, the complexities of social competence are reflected in the variety of definitions and associated operationalisations of this skill base. Critical concepts of social competence emphasise interpersonal and intrapersonal skills or a combination of these skills within a person and effective social interactions between persons and the environment. Specific skills reflect effectiveness in young children’s social interaction through observable behaviours (peer relations and prosocial behaviours) and internalised competencies (problem-solving and perspective-taking). Given the challenges with direct measurement of these skills, assessments often focus

on adult reports of key skills to assess and describe children's emerging social competence.

2.2 Emotional Competence

Emotional competence refers to a person's ability to express emotions, understand the emotions of self and others, and regulate emotional expressiveness appropriately for the social context (Denham et al., 2016; Saarni, 1999a). This section provides an outline of evidence regarding the importance of emotional competence, key theories about emotions and emotional development and discusses definitions of emotional competence and key subset skills. The section ends with key considerations related to the assessment of emotional competence.

2.2.1 Importance of Emotional Competence

As a crucial component of psychological well-being, young children's emotional competence supports their physical health, social competence, school readiness and performance, enjoyment of learning, overall well-being and later outcomes (Aviles et al., 2006; Denham & Bassett, 2018; Denham et al., 2016; Durlak et al., 2011; Eisenberg et al., 2006; Gallagher & Vella-Brodrick, 2008; Guerra & Bradshaw, 2008; Ministry of Education, 2019a; Morkel & McLaughlin, 2015; Parke, 1994; Rimm-Kaufman & Pianta, 2000; Schutte et al., 2007; Ulloa, 2011). Research highlights the long-term effects of preschoolers' emotional recognition and interpretation on their social behaviour and academic competence (Izard et al., 2001). Longitudinally, emotionally competent children demonstrate improved learning outcomes and higher levels of self-efficacy

and motivation (Payton et al., 2008; Zins et al., 2007), and preschool children's emotional regulation is bidirectionally associated with school adjustment and school readiness (Denham, 2006; Fung et al., 2020; Miller et al., 2006; Shields et al., 2001). Children with greater emotional competence and emotional knowledge are more likely to be liked by peers and prosocially behaved (Denham, 2001; Garner & Waajid, 2008; Trentacosta et al., 2006).

Conversely, poor skills in one or more components of emotional competence during preschool years are associated with risks to young children's physical and mental health, lower school performance, problematic social interaction, poor peer relationships, behavioural difficulties including behavioural problems, aggression, and withdrawal (Denham & Couchoud, 1990b; Denham et al., 1990; Hill et al., 2006; Rubin & Clark, 1983; Rydell et al., 2003; Ulloa, 2011). Research shows that young children's difficulties in understanding and regulating emotions are significantly related to disruptive behaviour problems (Duncombe et al., 2013). Lack of emotional knowledge in young children has also been associated with the exhibition of internalising and externalising problems (Hannesdottir & Ollendick, 2007; Trentacosta & Fine, 2010).

2.2.2 Theories of Emotional Competence

Emotions influence every aspect of our lives. Emotions in preschool years are complex and varied due to children's rapid biological changes and a more integrated dynamic environmental system in which children are involved (Denham, 2018; Plutchik, 1980, 2003; Saarni, 1990; Saarni, 1999a). In addition to biological responses, emotions

are viewed as contextually anchored responses in social contexts that convey social transactions and relationships; taken together, this reflects the dynamic nature of the emotion (Saarni, 1999a). Key theories such as evolutionary psychology, functionalist theories, cognitive neuroscience, and attachment theory provide conceptual insights into the function of emotions.

Evolutionary psychological perspectives view emotions as biological adaptations and cross-culturally universal features of facial expressions, with emotions having been identified (Al-Shawaf & Lewis, 2020; Ekman, 1971). From this perspective, emotions are evolved from ancestors and are genetically related to the emotions of other species. Adaptations to the environment would be difficult without emotions. Among evolutionary psychological approaches, *coordinating mechanisms* highlight the evolved function of emotions in coordinating mind and body to solve specific adaptive problems (Al-Shawaf & Lewis, 2020; Cosmides & Tooby, 2000; Tooby & Cosmides, 1990). Within the coordinating mechanisms, emotions are perceived as regulatory mechanisms to coordinate the mind and body to ensure harmonious coactivation. For example, fear coordinates mechanisms for preventing or escaping danger (Tooby & Cosmides, 1990). While psycho-evolutionary theories of emotions are seen as universal and biologically hardwired, with each primary emotion serving a specific evolutionary purpose (Izard, 1984; Izard, 1992), functionalism theories of emotion suggest that emotions are adaptive responses that help individuals meet specific goals within their environment (Campos et al., 1994; Lazarus, 1991; Saarni, 1990; Saarni, 1999a; Saarni et

al., 2006; Saarni et al., 2008).

Functionalists' perspectives consider emotional competence as an essential component of child development and emphasise the intra- and inter-personal functions of emotions, viewing emotions as context-dependent responses that help regulate social behaviour and interpersonal goals. For example, Lazarus (1991) proposed in the adaptational model of emotion that the generation of emotion occurs while the individual elicits a readiness to act for significant impacts. The model hypothesised a multidimensional appraisal theory of emotion that can be broken down into cognitive appraisal, physiological response, and action sequentially. Moreover, Saarni et al. (2006) highlighted the involvement of appraisal of ongoing person-environment interaction and readiness to act in particular ways. From these perspectives, each emotion functions uniquely to serve a person's social and cognitive goals and well-being.

Cognitive neuroscience of emotions focuses on brain development and the role of cognition in emotions (Deak, 2011; Lane & Nadel, 2000). Cognitive neuroscientists emphasise the significant role of the brain in the experience and identification of emotions. For example, the amygdala ties emotional knowledge to memories, decision-making and experience. The right cerebral hemisphere is indicated for emotion recognition, control of emotional expression and negative emotions, whereas the left hemisphere of the brain specialises in positive emotions (Deak, 2011; Harmon-Jones & Allen, 1998; Lane & Nadel, 2000; Silberman & Weingartner, 1986). Cognitive

neuroscience has provided insights into the psychological and biological processes of emotions, whereas social cognitive theories focus on the unique ways people react to various events due to the inseparability between an individual's affective responses and cognitive processes.

Given the impact of children's brain development and life experiences on their emotions, it is important to consider the attachment between children and their caregivers. Attachment theory underlines interpersonal roots of emotionality and individual differences in cognitive and affective processes relevant to emotional competence (Bowlby, 1980; Connors, 2011; Kafetsios, 2004). It also highlights the importance of a nurturing, responsive primary caregiver(s) in secure attachment with the child, such as facilitating comfort when the child feels distressed or threatened and encouraging the child to explore the environment. Research evidence supports the consequences of attachment on emotional competence, suggesting that the accuracy of emotion identification is positively associated with secure attachment. Secure individuals are relatively accurate in identifying facial expressions of negative emotions (Kafetsios, 2004; Magai et al., 1995).

The range of theories that relate to the development of children's emotional competence highlights the biological, functional, and social foundations for emotions and emotional competence.

2.2.3 Definitions

Unlike other areas of learning and development, there is relative consistency in the

definition or description of emotional competence, although variation in the overarching terms used to describe emotional competence exists in the research literature. For example, terms such as emotional literacy or emotional intelligence are sometimes used (Mayer & Salovey, 1997; Sharp, 2001). For this work, the term emotional competence is used for consistency with its use in New Zealand Ministry of Education initiatives for social and emotional competence. Below is an overview of major concepts related to the definitions of emotional competence.

Emotional competence has been defined as a broad construct that encompasses the understanding and regulation of emotions, the capacity for self-efficacy, and the adaptive emotional responses that help the child reach goals, cope with challenges and distressing emotions, and engage effectively in social interactions (Saarni, 1990; Saarni, 1999a; Saarni, 2007; Saarni et al., 2008). Emotionally competent children can identify emotions of self and others and respond emotionally, apply their emotional knowledge and expression simultaneously and strategically to relationships with others, and negotiate interpersonal exchanges and emotional regulation (Saarni, 1999b). Building on Saarni's definition of emotional competence, Denham et al. (2016) defined emotional competence as the ability to express a variety of emotions, understand the emotions of self and others, and regulate emotional expressiveness and experiences, when necessary, purposefully, and fully. Children can acknowledge the appropriateness of emotions and regulate their behaviours in culturally responsive ways through interaction with the physical and social world (Denham, 2018). Similarly, Zins et al.

(2007) defined emotional competence as a person's ability to understand, manage and express the social and emotional aspects of one's life in ways that enable the successful management of life tasks such as learning, forming relationships, solving everyday problems, and adapting to the complex demands of growth and development.

Overall, these definitions view emotional competence as learned abilities and knowledge-attributed attitudes relevant to cognitive, affective, and social aspects of emotional capabilities through organisational relationships. Hence, it is essential to understand emotional competence in organisational contexts as it implies an individual's awareness as a part of the social process of adjustment. Central to the definitions, three core aspects of emotional competence are generated: 1) understanding and knowledge of one's own and others' emotions, 2) emotional expressiveness and 3) the regulation of emotions. Each of these aspects of emotional competence is described in the following sections as a subset of skills of emotional competence. It is important to note that the skills of emotional competence are not acquired sequentially. Each skill influences the development of other skills, creating bidirectional associations that influence overall development.

2.2.4 Domains of Emotional Competence

2.2.4.1 Emotional Knowledge.

Children's emotional knowledge encompasses their abilities to identify and label emotions and build links between certain situations and particular emotions (Denham, 1998; Harris, 1989; Morgan et al., 2010; Pons et al., 2004). Normatively, awareness of

the self's emotions appears at 18 months with individual variation expected (Berk, 2013). During preschool years, due to new environment demands and the formation of relationships besides their primary caregivers, children's emotional identification and their abilities to identify and evaluate the associations between external causes, situational cues and specific emotions develop rapidly (Pons et al., 2003; Shoshani et al., 2021). By age 5, most preschoolers can focus on multiple cues of emotions and make more complex emotional assessments accurately through referencing causes, consequences and behavioural signs of emotion, such as frowning means unhappiness (Bierman, 1988). Higher emotion knowledge in young children may be associated with the exhibition of prosocial behaviour, better effortful control, effective communication, and higher social and academic competence (Denham et al., 2003; Fabes et al., 2003; Izard et al., 2001; Martinsone et al., 2022; Morgan et al., 2010; Mostow et al., 2002).

2.2.4.2 Emotional Expression.

As a central aspect of emotional competence, emotional expressions are defined as the ability to display emotions and behaviours perceived as expressive of emotion (Michela et al., 2014; Smith & Abell, 2016). Emotional expressions vary with the person's developing capacities, goals, and context. Multiple interacting cues, including vocal, facial, and gestural cues, eliciting different emotions are influenced by social expectations in sociocultural contexts and differ across situations (Eisenberg et al., 2007; Lewis, 2007). With appropriate emotional learning, children's emotional expression tends to be more well-organised and specific, which reflects their internal states more

precisely (Berk, 2013, 2016). Through communication with primary caregivers, parents selectively mirroring emotional behaviours help infants construct emotional expressions that closely resemble adults (Gergely & Watson, 1999). Positive emotional expressiveness in young children supports the formation of friendships and enhances peer likability (Denham et al., 2003). Conversely, children experiencing and expressing more negative emotions may show aggression with peers (Denham et al., 2003). Additionally, children's expressions of pleasant and unpleasant emotions may relate to their use of emotional regulation strategies (Prosen & Smrtnik Vitulić, 2018; Saarni, 1999a).

2.2.4.3 Emotional Regulation.

Emotion regulation involves an attempt to influence emotions or the strategies of adjusting emotional state to a comfortable level of intensity to accomplish goals (Berk, 2013; Naragon-Gainey et al., 2017). Infants have limited capacities for emotional regulation and rely on co-regulation with trusted caregivers, whereas preschoolers' emotional regulation capacities are strengthened to adapt to increasingly complex emotions and the demands of a range of social interactions (Berk, 2016). Preschoolers verbalise a range of emotional self-regulation strategies with the development of language and communication, and their use of emotional regulation strategies is associated with fewer emotional outbursts over the early childhood stage (Cole et al., 2010; Thompson & Goodvin, 2007).

Emotion regulation supports emotion recognition and understanding during

toddlerhood and encourages emotional perspective-taking and critical thinking skills during preschool (England-Mason et al., 2023; Saarni, 2011). However, emotional regulations require several cognitive capacities, mainly executive function consisting of effortful control, attention focusing and shifting, inhibition and planning (Eisenberg & Spinrad, 2004). The section on executive function will explore the definitions and domains of executive function and its associations with emotional regulation.

2.2.5 Assessments of Emotional Competence

As noted above, the subset skills of emotional competence are not acquired sequentially, and there is a bi-directional influence among skills. This necessitates a clear understanding of each construct and has implications for the assessment of these emerging capabilities in young children.

Similar to social competence, there is a range of adult-reported assessments of children's emotional competence, each with a different focus on the specificity and range of subset skills (Denham et al., 2016). These tools often include adult ratings for subset skills for children's social and emotional capabilities and are prone to the same concerns and issues as noted above (Zimmer, 2017). A range of these tools may also include aspects of children's mental health or well-being (Humphrey et al., 2011).

Moving beyond adult-reported measures, one common assessment of children's emotional knowledge and expression is through facial expression recognition and emotion naming. While these approaches do not capture the complexity of emotional regulation, the ability to measure children's emotional knowledge and expression is

used as an appropriate indicator of emotional competence in young children (Denham, 1986).

Among various measurements of emotional competence, the *Affect Knowledge Test (AKT)* was designed by Denham (1986) to measure preschoolers' developmentally appropriate emotional knowledge and expressions by using puppets with detachable faces to depict facial expressions, such as happy, angry, and sad. AKT has been well-recognised in measuring preschoolers' emotional expression and knowledge cross-sectionally in the New Zealand context (Ahmad et al., 2019; Ahmad et al., 2021; Russell et al., 2022; Walsh et al., 2019).

2.2.6 Summary

To summarise, emotional competence is fundamental to young children's development and well-being, impacting their social relationships, school readiness, and long-term outcomes. Emotional competence encompasses skills such as expressing emotions, understanding one's own and others' emotions, and regulating emotional responses appropriately within various social contexts. Theories, including evolutionary psychology, functionalist theories, cognitive neuroscience, and attachment theory, highlight the biological, cognitive, and social foundations of emotional competence, emphasising its role in adaptive and interpersonal functions. Key skills of emotional competence include emotional knowledge (recognising and labelling emotions), emotional expression (displaying emotions in socially appropriate ways), and emotional regulation (managing emotional responses to maintain well-being and social harmony).

Assessments of emotional competence often rely on adult reports and can be complemented by tests that assess facial expression recognition and emotion naming skills, which can evaluate emotional knowledge and expression.

2.3 Language Skills

Language skills refer to the ability to understand, express, and communicate thoughts with others verbally and nonverbally (*APA Dictionary of Psychology*, 2007). This section of the thesis describes distinctions between language and communication for background and clarity. However, language and communication are generally presented as a unified concept in the remainder of the thesis. This section provides an outline of evidence regarding the importance of language skills, and key theories about language development. Definitions of language and communication, and key subset skills are discussed. The section ends with key considerations related to the assessment of children's language skills.

2.3.1 Importance of Language

Young children's language skills have a direct impact on their social-emotional development, cognitive development, and later academic achievement (Cole et al., 2010; Grantham-McGregor et al., 2007; Hart & Risley, 1995). Language provides opportunities to improve social competence, and children's early verbal capabilities facilitate the acquisition of interpersonal functioning and social problem-solving skills (Barnett et al., 2012; Bruce & Bell, 2022; Hay et al., 2004). For instance, cross-sectional studies indicate a positive association between language and emotion regulation in

young children, providing support for theoretical frameworks in which competence in expressive language supports young children to communicate in socially acceptable ways and regulate their behaviours (Cole et al., 2010; Vallotton & Ayoub, 2011).

Moreover, longitudinal research demonstrates that children at 24 months of age possessing larger vocabularies not only attain higher levels of literacy and numeracy but also exhibit enhanced self-regulation as they transition to preschool (Morgan et al., 2015).

In contrast, children lacking language and communication skills may experience poorer social, emotional, and behavioural functioning and academic performance (Cross, 1998, 2011; McCabe & Meller, 2004). Longitudinal and cross-sectional studies show expressive language to be significantly negatively associated with physical aggression from infancy to preschool, with preschool-aged children with language delays less likely to engage prosocially and/or to behave aggressively (Dionne et al., 2003; Qi & Kaiser, 2004; Sanson et al., 1993). Children who have difficulty with language skills may display higher levels of early conduct problems and internalising problems than typically developing peers, with these behaviours perhaps acting as a replacement for verbal expression (Fujiki et al., 2001; Redmond & Rice, 1998). Young children with language impairment may be less accurate in emotion labelling and share lower emotional knowledge than other typically developing children (Delaunay-El Allam et al., 2011; McCabe & Meller, 2004). Longitudinal studies of preschoolers identified with speech and language concerns in early childhood show that these children displayed

underdeveloped reading skills and sustained lower performance in literacy compared to typically developing children when they started school (Foster & Miller, 2007; McLeod et al., 2019). Together, such research supports the important role of language skills and effective communication in children's early development.

2.3.2 Theories of Language

In this section, I outline four key theoretical perspectives on language and communication and provide an overview of key processes associated with language and communication development. The four selected theories include: Chomsky's nativist theory on language, Hymes's communicative competence theory, social interactionists' perspectives on language, and theories related to language and brain development. These perspectives delve into the interplay between innate abilities and environmental influences in shaping language development. Each theory enriches the understanding of the mechanisms behind language and communication development.

Noam Chomsky's nativist theory, including his proposition of a Language Acquisition Device (LAD) and conceptualisation of linguistic competence, has contributed to understanding language acquisition and structure (Chomsky, 1976, 2002; Otto, 2014). Nativist theory proposes that language is a unique human accomplishment connected to structures of the brain, which process language differently from other stimuli (Berk, 2013). Central to LAD is an innate language-specific system that enables a child to combine words into grammatically coherent sentences, comprehend the meaning of the sentences, and actively interpret language input based

on their current knowledge, such as the vocabulary they have acquired (Chomsky, 1976). Chomsky makes a distinction between language *competence* and language *performance* to further delineate aspects of the linguistic system, specifically grammar, in generating well-formed sentences while accounting for native speakers' overall language performance. Chomsky's perspectives on linguistic development posit linguistic *competence* as an innate, unconscious knowledge of grammatical structure that native speakers intuitively develop during language acquisition.

Chomsky's nativist theory has been criticised due to its narrow focus on the grammatical component of language, overlooking the social interaction in which the speakers are involved (Canale, 1980; Hall, 1977). One criticism is that Chomsky's approach to linguistic structure, based on the rigidly defined LAD system, tends to overlook the dynamic and evolving nature of language, failing to account for its inherent flexibility (Hall, 1977). In Chomsky's emphasis on grammatical knowledge, other aspects of linguistic communication, such as gestures, and non-verbal cues that contribute to effective communication, are ignored. Nevertheless, Chomsky's perspectives on language place emphasis on the role of language structure and the incorporation of acquired knowledge.

Beyond Chomsky's emphasis on individuals' linguistic knowledge and the formal structure of language as the basis of linguistic competence, Dell Hymes's theory of communicative competence places importance on the patterns of language use within social contexts. Hymes takes a social perspective on language acquisition and defines

communicative competence as primarily reflecting the ability to communicate appropriately (Hymes, 1972). According to Hymes, communicative competence encompasses not only knowledge of language forms but also an understanding of the relationships between language form and function as they relate to social interaction.

Hymes's perspectives on communicative competence are embodied in four distinct qualities: (1) grammatic possibility, which refers to the ability to construct language that incorporates the appropriate grammatical components and which reflects an understanding of how grammar structure influences meaning; (2) implementation feasibility, which considers the psychological and linguistic elements involved in producing and understanding language and acknowledges the cognitive processes that play a role in effective communication; (3) contextual appropriateness, which involves adherence to social norms in language use in ways that align with cultural and societal expectations; (4) performability, which relates to the capability of using language effectively within specific contexts. This final quality highlights the practicality of communication, emphasising the need to use language effectively to convey ideas, emotions and achieve particular goals. Overall, these four aspects of communicative competence broaden the understanding of language and communication, calling for a reconsideration of Chomskian linguistic competence and a shift toward viewing language as a social behaviour.

Building on language as a social behaviour, the perspectives of social interactionists emphasise the crucial role of sociocultural interaction in children's language

development (Bruner, 1985, 1990; Vygotsky, 1978). Lev Vygotsky's initial research underscored the impact of social interaction on language development, considering language as the cornerstone for various advanced cognitive processes such as planning, categorisation, problem-solving, self-reflection, controlled attention, and deliberate memorisation (Berk, 2013). According to Vygotsky, the rapid advancement in language acquisition originates from socially constructed and culturally transmitted higher mental functions (Vygotsky, 1978). Vygotsky's theories also suggest that verbal language establishes a framework for organising and developing culturally influenced thoughts and behaviour patterns. Additionally, Bruner (1985) proposed that the environment supports children's language development through a Language Acquisition Support System (LASS), which involves a network of more knowledgeable individuals interacting with and supporting the child's language development. Thus, through observing social interactions, children develop an understanding of communicative functions before expressing themselves linguistically (Halliday, 2007; Tomasello, 2000). The perspectives of social interactionists provide insights into the mechanisms and significance of social interactions for children's language development.

Recent advances in neuroscience have afforded the ability to gather evidence on the bidirectional relationship between language acquisition and the development of associated neurological structures. The interconnections between the brain and language development highlight the significance of a sensitive period for language

acquisition and the maturation of the brain in understanding and producing language (Berk, 2016; Friedrich & Friederici, 2010). The first 5 years of brain development, coinciding with the preschool years, are characterised by notable and rapid changes in children's language development. Through exposure to sets of familiar and novel objects, Childers and Tomasello (2002) estimated that children acquire approximately 10,000 words from birth to 5 years old, highlighting the linguistic growth experienced in early childhood.

Studies of the brain have identified Broca's area and Wernicke's area, two specific regions in the left hemisphere of the cerebral cortex, as critical for language understanding and production. Broca's area is considered to facilitate grammatical processing and language production, while Wernicke's area is responsible for word comprehension (Berk, 2013). Moreover, research indicates that the rate and volume of children's language acquisition also influence the development and specialisation of these structures (Mills & Conboy, 2005; Mills et al., 1993). Further, higher vocabulary producers exhibit greater accuracy and efficiency in word identification, suggesting that vocabulary development plays a significant role in the organisation of neural systems related to phonetic details within the context of word comprehension (Mills et al., 2004; Torkildsen et al., 2009). Such evidence, supporting a critical period and exposure-based development of key neurological structures involved in language processing, production, and comprehension, underscores the importance of acquiring communicative competence for young children to thrive in their culture and society.

2.3.3 Definitions

The discussed theories acknowledge the intricacies of the development of language and communication competence. Simply put, language serves as the basis of how people perceive things, communicate, and interact with others in daily life. Individuals' abilities to connect with others and make sense of the world rely heavily on the functional use of language that conveys and receives information. It is important to note that language and communication do not necessarily occur concurrently. In the early phases of language development, receptive language (language comprehension) encompasses a wider array of concepts and information that extend beyond the use of verbal language, which includes elements like symbolism, art, and music, as well as non-verbal social cues such as facial expressions and body language, among others (Cheung et al., 2022). Indeed, this necessitates an understanding of these as separate, though often inter-dependant systems. For example, a child may possess age-appropriate language skills but display limited motivations to communicate, indicating difficulties in communication rather than language, which may or may not hinder later language development (McCormack et al., 2011). Clarifying the concepts of language, communication, and communicative competence enables developmental scientists to make useful distinctions and thus explore relationships between these concepts. In this section, definitions of language, communication, and communicative competence are provided. From there, the discussion covers specific domains of language development and associated approaches to measurement in early childhood.

2.3.3.1 Language.

Different contexts and disciplines share divergent meanings of language.

Informally, language is understood as a culturally specific communication system, such as English (Hauser et al., 2002). Linguistically, language refers to an internal component of the mind or brain, viewed as a unique human cognitive capacity that involves the use of arbitrary symbols and structured rules to convey meaning (Friederici et al., 2017).

Early linguistic definitions by Bloom and Lahey (1978) defined language as a code whereby ideas about the world are represented through a conventional system of arbitrary signals for communication. The Taxonomy of Language framework proposed by Bloom and Lahey (1978) integrates three significant components of language: form, content, and use. The form of language includes syntax, morphology, and phonology, which are structural aspects of language. The content involves semantic components of language, such as vocabulary knowledge and understanding of objects and events. The use of language relates to pragmatics, which involves the ability to use language in social contexts. These components of language contribute to linguistic communication through various aspects, including phonological, semantic, syntactic, morphemic, and pragmatic knowledge.

In psychology, language is defined as a specific communicative system that enables people to express or communicate thoughts, feelings and ideas through verbal and non-verbal structures, including speech sounds and written symbols used by a particular group of speakers with distinctive vocabulary, grammar, and phonological

systems, and through alternative communication methods, such as sign language (*APA Dictionary of Psychology*, 2007). Owens (2020) defines language as a socially shared code or system of symbols, such as signs and words, for thinking and communication with others. Overall, definitions of language emphasise the communicative nature of language, the symbolic systems, grammatical structure, and cultural contexts.

2.3.3.2 Communication.

Definitions of communication share several similarities, conceptualising communication as a process that is both symbolic and transactional. Communication is described as a continuous and complex process that unfolds over time. For instance, one definition views communication as the process of the communicator transmitting stimuli, typically through verbal means, to modify the behaviour of audiences (Hovland et al., 1953). The definition by Hovland et al. focuses narrowly on verbal signals, and it overlooks other essential aspects of communication, including non-verbal messages such as facial expressions and body language. In contrast with the definition of Hovland et al. (1953), the definition by Kuhl (2004) acknowledges the importance of both non-verbal and verbal use of language in interacting with others and considers interactional use of verbal and non-verbal language a crucial element for language development. Further, Kuhl's definition aligns with perspectives of communication as fundamentally symbolic in nature. Communication relies on signs and symbols (either verbal or non-verbal) that represent shared meaning that does not necessarily have an inherent or natural connection with the signs or symbols. The symbols are simply agreed upon,

with interpretation learnt through social connections and day-to-day communicative transactions. The transactional view of communication emphasises the role of feedback from the receiver. This can be observed in the definition by West and Turner (2017), which defines communication as a dynamic and transactional process where individuals create, share, and interpret messages to convey meaning and establish connections. In summary, these definitions collectively agree upon the multi-dimensional nature of communication as a dynamic process invoking various components, such as symbols and contexts.

The variations in how communication is conceptualised pertain to its social nature and intentionality. The social nature of communication encompasses interpersonal and intrapersonal communication and its impact on individuals in our surroundings. The social nature is exemplified by a definition that characterises communication as a social exchange (Goyer, 1970). Debates surrounding communication as intentional or unintentional behaviour relate to the complex nature of communication. This includes distinctions between spoken, intentional communication and body language that may unintentionally convey actual feelings or reactions; and symbolic or metaphorical communication that often conveys nuanced and culturally situated ideas (Miller, 2005). These distinctions align with the argument by Kellermann (1992) that communication is receiver-directed and intentional, sometimes unconscious and often guided by cultural scripts and ingrained habits. Overall, the variations in definitions of communication highlight the social nature of communication and ongoing discussions regarding its

intentionality. The various definitions shed light on the complexities of communication, including shared experiences, intentional nature and various behaviours involved.

2.3.3.3 Communicative Competence.

Building on the understanding of communication, communicative competence refers to the combined knowledge of language and underlying skills that enable effective communication. Initially defined by Dell Hymes (1972), communicative competence encompasses individuals' knowledge of forms and structures (grammar) of language and their ability to apply them in social situations (i.e., how well they perform). However, Canale (1980) later refined Hymes's definition by removing the performance aspect and proposing four components of communicative competence: (1) grammatical competence, which involves knowledge of sounds, letters, syntax and other elements; (2) sociolinguistic competence which encompasses awareness of sociocultural rules that determine language acceptability; (3) strategic competence, which entails utilising communication strategies to deal with interactional breakdowns and optimise learning strategies; and (4) discourse competence, which focuses on creating cohesive and coherent textual structures. By differentiating skills from the performance of communicative tasks, this definition provides a useful framework, such that communicative competence can be viewed as encompassing the comprehensive understanding of language and the associated skills that enable speakers to communicate effectively and appropriately in social contexts (Canale, 1980; Cazden, 2011; Shugar, 1986). Competence, in this context, applies to speakers' knowledge of

constructing and using language, rather than their performance in specific situations (Foster-Cohen, 2013).

2.3.4 Domains

The definitions of language, communication, and communicative competence highlight the intricacies and complexities of skills and considerations within the development of the communicative system. For this research, the following sections will focus on receptive and expressive language capacities, as key domains of interest in children's learning and development and aligned with the available information in the GUINZs.

Receptive language refers to the ability to understand and interpret the language used by others, also viewed as language comprehension. It involves actively receiving and comprehending language through activities like listening or reading. In the early stages of language development, a child's abilities with language are primarily receptive (Cheung et al., 2022; Stremel, 1994). As young children progress in their language development, they gradually gain the ability to receive and interpret language input, comprehend words, and understand sentences from their surrounding environment (Berk, 2013, 2016). Through language exposure, young children continually develop their receptive language skills to comprehend increasingly complex language structures and nuances. Receptive language is important for young children as it forms the foundation of their overall language development.

Expressive language encompasses the capacity of individuals to express and

communicate information to others, also seen as language production. It involves using language skills to explicitly generate labels and activate related concepts internally (Feldman, 2019). In other words, expressive language is not only about making sounds but also intentionally and accurately matching the sounds and internal thoughts with an object or representation. This is complex and includes thinking processes of being able to recognise an object or concept, knowing the name or label for that object or concept, selecting and being able to express the name or label. Alongside conceptual understanding and vocabulary selection, language skills are used to produce a message, which includes speaking and writing (Landa, 2005; Smith et al., 2014; Stremel, 1994). From the production of first words, young children gradually learn new vocabulary, phrases, and grammatical structures and create sentences (Berk, 2013). Children's oral language develops earlier than written forms of language due to their early social interaction and the complexities of learning written language. As expressive oral language develops, children start to communicate with greater sophistication in a range of social contexts, and eventually facilitate effective communication.

Definitions of receptive and expressive language emphasise that these capacities are distinct from each other, prompting discussions of whether language comprehension and production should be treated as separate, diametric questions for investigation, or viewed as interconnected aspects of language processing (Hendriks, 2013; Pickering & Garrod, 2013). As indicated, receptive language (comprehension) and expressive language (production) are considered to involve separate cognitive

processes and abilities (Smolak, 1982). However, alternative perspectives suggest that language production and comprehension strongly overlap in neural circuits (Menenti et al., 2011; Skipper et al., 2005).

2.3.5 Assessments of Language

The relationships between receptive language (language comprehension) and expressive language (language production) have implications for measurement approaches to obtain valid profiles of children's linguistic skills (Conti-Ramsden & Durkin, 2012). For example, the potential asymmetry in language comprehension and production development, such that the development of expressive language skills is dependent on receptive language skills, might imply that both dimensions should be measured. Research indicates that preschoolers frequently exhibit greater language comprehension capabilities than production, likely because language comprehension is more accessible than production in the early stages of development (Ibbotson, 2022). Specifically, language expression requires children to activate the appropriate words to convey their ideas, which can be more challenging than language comprehension. A child may grasp specific language characteristics but struggle to produce language that effectively represents knowledge.

Previous research has employed various measures for assessing young children's language abilities, ranging from instruments designed for general language abilities to those targeting specific dimensions of language, according to research purposes. Language assessments typically evaluate the general language abilities of children aged

over 3 years and commonly encompass both expressive and receptive language modalities, providing a comprehensive understanding of their overall language abilities. For example, the standardised Australian and New Zealand version of Clinical Evaluation of Language Fundamentals–Preschool Second Edition by Wiig et al. (2006) evaluates multiple aspects of language, including receptive language, expressive language, core language, language content and language structure. On the other hand, some measurements focus on a single modality of language, such as the Peabody Picture Vocabulary Test, which assesses the receptive vocabulary knowledge of children aged 2 years and above (Dunn, 2021; Rothman, 2005), or the second edition of Expressive Vocabulary Test (EVT-2; Williams, 2007) which assesses expressive vocabulary and word retrieval without involving reading or writing tasks. The diversity of language assessments reflects the distinct domains underlying children’s language development and highlights the importance of selecting appropriate tools that align with specific research or practical purposes (Denman et al., 2017). Researchers and practitioners can choose from a range of measurements tailored to capture the relevant dimensions of language skills.

Two primary concerns are pertinent to the instruments used in New Zealand for assessing young children’s language development. First, most language assessments used in early learning have been developed in Western English-speaking contexts, with some instruments translated and adapted for use with non-English speakers. This means there are limited measures appropriate for dual or multi-lingual children for

whom English is not a primary language at home (Education Review Office, 2024).

Multilingual children may develop language skills at a different rate compared to monolingual children, as they may mix grammar or vocabulary from different languages they are learning, leading to a longer processing time.

Second, some instruments were designed to screen language problems, but they often lack diagnostic accuracy (Denman et al., 2017). This limitation is also evident in previous systematic reviews of screening for early language delay, which showed that while screening measurements may be adequate, they tend to have lower sensitivity than specificity in identifying language difficulties in children (Law et al., 1998; Law et al., 2000). The use of parent report measures of children's language and communication capabilities has been promoted as one way to overcome the shortcomings of direct assessments (Ebert, 2017).

2.3.6 Summary

To summarise, language skills are essential for young children's development, influencing their social-emotional growth, cognitive abilities, and academic success. Language skills encompass a range of abilities, including understanding and expressing thoughts through verbal and nonverbal communication. Theories of language development—such as Chomsky's nativist theory, Hymes's communicative competence theory, social interactionist perspectives, and neuroscience findings on brain-language interconnections—highlight the complex interplay of social and cognitive factors in language acquisition. Key language components include receptive language

(understanding and interpreting language) and expressive language (conveying information through spoken or written words), both crucial for social interaction and learning. Assessments of language development often include tools measuring expressive and receptive capabilities, with measures tailored to specific skills like vocabulary or comprehension. However, these assessments present challenges, particularly regarding cultural and linguistic diversity, indicating a need for more inclusive and diagnostically accurate tools as well as supplementing tools with parent report measures to better understand and support children's diverse language development trajectories.

2.4 Executive Function

Executive function refers to a set of cognitive processes that manage and regulate ongoing, goal-directed behaviours, with a focus on effortful control (Berk, 2013; Diamond, 2013). During infancy and preschool years, executive function develops rapidly, laying a crucial foundation for the subsequent maturation of advanced cognitive processes throughout adulthood (Garon et al., 2008). It may be useful to note that the use of the term executive function (EF) is relatively recent within early childhood education. While EF has been studied since the 1970s, advances in neuroscience have resulted in greater awareness of attention to EF in research. EF can be viewed as a subset of cognitive development that arose from work in cognitive psychology in the 1950s and later work on human memory systems, which proposed control processes in sensory register and short-term and long-term memory stores

(Atkinson & Shiffrin, 1968). The term control process was later adapted into 'executive control' by Belmont and Butterfield (1971) as cognitive models were applied to learning, particularly among individuals with disabilities. Karl Pribram is often credited as the first researcher to use the term executive function (Pribram, 1973).

This section presents an overview of the evidence highlighting the significance of EF, explores key theories on the development of EF in young children, and defines EF and its essential components. The section concludes by discussing the assessment of EF in young children.

2.4.1 Importance of Executive Function

As a complex cognitive process involving regulatory mechanisms, EF enables the growth of adaptive, goal-oriented behaviours (Barkley, 1997, 2001; Carlson et al., 2013; Hughes & Ensor, 2011; Hughes & Graham, 2002). Early EF is associated with gross motor skills and physical activity and significantly contributes to the development of young children's social and emotional competence and cognitive abilities including self-regulation and language, and later-life outcomes such as school readiness, and academic success (Blair & Razza, 2007; Cook et al., 2019; Diamond, 2016; Espy et al., 2004; McClelland et al., 2007; Veldman et al., 2023; Welsh et al., 2010). EF equips young children with the ability to exercise top-down attentional control over instinctive or impulsive thoughts and responses, enabling them to manage their thoughts, emotions, and behaviours in social interactions, develop secure relationships and explore the surrounding world (Berk, 2013; Diamond, 2016; Gooch et al., 2016). Furthermore,

research has explored relationships between children's fundamental motor development or physical activity and EF to find a range of associations that suggest EF and motor skills influence each other in important ways (Cook et al., 2019). Additionally, higher levels of EF skills are linked to improved literacy, numeracy, and mathematics skills upon entering school (Bull et al., 2008; Bull & Lee, 2014; Prager et al., 2023; Welsh et al., 2010). In short, EF is important in facilitating effective social interactions, associated with motor development and physical activity, and supports academic learning.

Deficits in early EF are associated with difficulties in social functioning, academic underachievement in later school years and poor outcomes in individuals with ADHD, as shown in studies involving clinical and community-based participants (Biederman et al., 2004; Castellanos et al., 2006; Clark et al., 2010; Fan & Wang, 2023; Morgan et al., 2019; Willcutt et al., 2005). Clark et al. (2002) demonstrated that lower levels of EF carry social consequences, while Biederman et al. (2004) found a high risk of significant academic impairments in individuals who lack EF skills. Moreover, research has indicated that deficits in EF, especially in inhibitory control in young children, are associated with increased difficulties in language and literacy (Gooch et al., 2016; Pauls & Archibald, 2016). Additionally, children who struggle with manipulating information in working memory, applying cognitive strategies flexibly, and inhibiting inappropriate strategies or information are more likely to encounter conceptual challenges in various aspects of mathematics (Clark et al., 2010). Overall, various areas of young children's

lives are significantly influenced by EF.

2.4.2 Theories of Executive Function

Theoretical perspectives regarding the development of EF have provided a framework for the conceptualisation and evaluation of the construct. In this section, I outline three fundamental theoretical perspectives on EF to provide an overview of key processes linked to its development. The selected theories encompass Luria's developmental perspective and cognitive and neuroscience perspectives, including Miyake and Friedman's Model of Executive Functions. These brain-focused theories are briefly considered in the context of education-based perspectives.

Expanding on Vygotsky's cultural and historical theory related to language and thought, the development of neurological structures responsible for advanced cognitive abilities, including memory and attention, was postulated by Luria's neurodevelopmental model (Luria, 1963, 1966, 1973, 1980), and stages of higher cortical maturation related to specific developmental stages. Luria's model delineated five distinct stages for brain and EF development that include development of stem structures (0-1 year), activation of the primary sensory areas (1-2 years), development of single modalities in the secondary association areas of the brain (2 to preschool years), activation of tertiary areas of the parietal lobes (commencing as the child enters school at 7-8 years), and activation of the brain (from 8 years to adulthood). Luria's conceptualisation of cognitive functions involved three separate yet interconnected functional units, including (1) attention and arousal regulation; (2) processing and

storing information; and (3) planning and regulating activity. Notably, the third functional unit, associated with the prefrontal areas of the frontal lobes, aligns most closely with what we now understand as executive function. Luria's insights into frontal functions were insightful and laid the foundation for subsequent research, including the 'supervisory attentional system' (SAS) model that described processes for controlling and regulating attention when situations offered novel experiences, or the need to make new decisions (Shallice, 1988). However, research contributions from this time had key limitations. These limitations encompass biases stemming from clinical observation-based reports, which often lack quantitative rigour, and the early neuropsychological research constraints, including imprecise observations due to the absence of CT or MRI techniques.

Fortunately, recent developments in cognitive and neuroscience research, with the assistance of rapidly advancing technology, have allowed for a more comprehensive exploration of the brain, its organisation, functioning, and the underlying physiology that includes neurons, synapses, and neural networks (Fischer & Daley, 2006). These fields emphasise the importance of studies focusing on specific and experimentally approachable hypotheses regarding how the brain responds to simple stimuli (Turk et al., 2002). Building on earlier research on the frontal cortex and EF (Luria, 1966; Shallice, 1988; Stuss et al., 1995), Baddeley (1986) introduced the concept of the central executive as a scientific construct, hypothetically linked to the frontal cortex. Baddeley's conceptualisation of the central executive aligns with the 'supervisory attentional

system' (SAS) model developed by Norman and Shallice (1986), which distinguishes between actions performed without awareness and situations requiring deliberate attentional resources. Initially, EF was considered a unitary construct composed of interacting subprocesses operating within an overarching processing system to achieve a common overall function (Baddeley, 1986; Norman & Shallice, 1986; Shallice, 1988; Stuss et al., 1995). Another way to describe this view of EF is as a single integrated system or a central management system, in which different mental processes work together, managing complex cognitive tasks. More modern views of EF now posit that it is a multi-dimensional construct with distinct but interrelated components, such as inhibition, working memory, cognitive flexibility, and planning (Ackerman & Friedman-Krauss, 2017; Miyake et al., 2000).

Specifically, the Model of Executive Functions by Miyake et al. (2000) identified three core components of executive function: shifting (i.e., flexibly switching between tasks); updating (i.e., keeping information in working memory); and inhibition (i.e., the ability to suppress automatic responses). These researchers went on to propose a Unity and Diversity Theory that emphasised that while the three components of EF share common cognitive processes, they also support specialised functions and individuals may vary in their performance across these capabilities (Friedman & Miyake, 2017).

Connecting to educational perspectives, the insights into EF have gained traction in education in recent years and are increasingly being integrated into educational practices to inform teaching and learning, and are reshaping the landscape of

educational and learning contexts. There is controversy in the interpretation of EF concepts, including aspects of working memory, and how they are perceived differently by professionals from diverse backgrounds, such as those rooted in brain-based and school-based backgrounds (Grammer & Ahmed, 2023). For instance, measuring EF may be limited to experimental environments rather than classroom settings. The extent to which the assessments can reflect children's EF skills in the classrooms is not clear (Cameron et al., 2012). While controversies continue in numerous theories and models in explaining the development of the EF process (Meltzer et al., 2018), variations also manifest in the definitions and the specific skills included in EF models.

2.4.3 Definitions

Although research on EF development in early childhood has seen significant advancements since the 1990s (Espy, 1997; Gerstadt et al., 1994; Hughes, 1998), there remains an ongoing lack of consensus regarding its definitions and the underlying constructs of measurement. This lack of agreement can be attributed to two key factors, noted by Fletcher (1996). First, EF encompasses a wide spectrum of cognitive functions, such as working memory, inhibitory control, and cognitive flexibility, resulting in a diversity of its definitions. Second, EF's latent constructs are inseparable, with measurement tasks often encompassing multiple facets of attention and memory that are often designed to target specific components of EF, such as working memory or inhibitory control, and can be scored separately (Miyake et al., 2000). Despite advancements, it remains imperative to exercise prudence and diligence when

operationalising and investigating this pivotal cognitive domain.

While considerable research has been conducted on EF, there remains a limited consensus regarding a universally accepted operational definition of EF (Blair & Razza, 2007). Diverse definitions include various skills and abilities derived from distinct neuropsychological instruments identified in current literature (Barkley, 2021; Goldstein et al., 2014; Wasserman & Wasserman, 2013). The definition of EF by Welsh and Pennington (1988) is “the ability to maintain an appropriate problem-solving set for the attainment of a future goal” (pp. 201-202). This definition is commonly cited by many authors. A more modern definition provided by Lezak et al. (2012) conceptualised EF as encompassing capacities that enable a person to engage successfully in independent, purposive, self-serving behaviour. Common features in various definitions of EF are goal-setting and planning, organisation of behaviours over time, flexibility, attention, working memory and self-regulatory processes, such as self-monitoring (Meltzer et al., 2018). A good example is the definition from Blair (2002) that EF is a cognitive construct that unites working memory, attention, and inhibitory control for planning and executing goal-directed activity. Berk (2013) views EF as the set of operations and strategies necessary for self-initiated, purposeful behaviour in challenging situations. Overall, EF can be understood as a general term used to identify the complex cognitive processes that serve ongoing, goal-directed behaviours.

2.4.4 Domains

Due to the absence of a universally agreed-upon definition of EF, there is no

consensus regarding which components should be encompassed under the construct, what specific skills should be called, and how they should be measured (Wasserman & Wasserman, 2013). In child development research, the understanding of EF has broadened by employing three- or four-factor models to examine its components in various studies. For instance, three factors commonly include working memory, inhibitory control, and cognitive flexibility (Caporaso et al., 2019; Diamond, 2013, 2016; Miyake et al., 2000), and four factors often encompass working memory (mentally representing information), inhibition, intentionality (goal-directedness) and planning (Anderson, 2002; Lezak et al., 2012). Diamond (2016) asserts that EF comprises: inhibitory control, involving complex psychological processes in cognitive control to stay focused and resist temptations; cognitive flexibility, allowing individuals to change course as needed; and working memory, allowing individuals to retain and process information in the mind, while using that information for cognitive tasks. Anderson (2002) conceptualised the executive system with four components: attentional control, allowing individuals to focus and shift attention as needed; information processing, allowing individuals to receive, organise and interpret information; cognitive flexibility, allowing individuals to adapt thinking, consider other perspectives and adjust strategies and actions as needed; and goal setting, allowing individuals to plan, prioritise, and pursue goals.

Regardless of whether a three- or four-factor perspective is taken, key components of EF are viewed as distinct yet interlinked. Contemporary research on EF underscores

its multidimensional nature, highlighting the dissociable processes and the distinction of cognitive processes such as working memory and inhibitory control, which can be differently affected by a range of factors and which can be studied or measured separately (Baggetta & Alexander, 2016; Barkley, 2001, 2012; Carlson et al., 2004). Components of EF, such as working memory and inhibition, have been shown to possess dissociable characteristics, including varying developmental trajectories and timing (Carlson, 2005; Diamond, 2002, 2006, 2013). The present research specifically centres on three frequently employed core components of EF: inhibitory control, working memory, and cognitive flexibility.

Inhibitory control involves consciously suppressing one's focus or subsequent reactions (Diamond, 2013). This entails the ability to manage attention, behaviour, thoughts, and/or emotions, overriding strong internal predispositions or external temptations to engage in more appropriate or necessary actions, such as intentionally disregarding distractions or refraining from making impulsive statements. The development of inhibitory control occurs rapidly during the preschool years (Diamond, 2006). Inhibitory control empowers individuals to actively choose and alter their reactions and behaviours rather than merely acting as instinctive creatures of habit. Although inhibitory control is notably challenging for young children, it holds a vital role in regulating behaviour and attention span (Diamond, 2013). Early-life inhibitory control may serve as a predictor of lifelong outcomes (Moffitt et al., 2011).

Working memory involves the process of retaining information within one's

immediate awareness and mentally manipulating it in multiple ways (Baddeley & Hitch, 1994; Smith & Jonides, 1999). For instance, working memory may involve simultaneously holding two numbers in mind while performing subtraction operations. The development of working memory commences early in life, with infants and young children demonstrating the ability to retain one or two things in mind for extended periods (Nelson et al., 2016). Working memory is important in understanding connections between elements from an integrated perspective, such as creative endeavours that involve disassembling and recombining elements in novel ways. Moreover, working memory enables the utilisation of conceptual knowledge to guide the plans and decisions we make (Diamond, 2006, 2013).

Cognitive flexibility refers to the capacity to flexibly shift perspectives and consider a single stimulus from diverse perspectives (Miyake et al., 2000). An illustrative instance involves the ability to shift attention from a self-centred viewpoint to consider another person's standpoint in a given situation. Substantial enhancements in cognitive flexibility are evident during the preschool years (Diamond, 2006), but this ability tends to diminish with age (Magnusson & Brim, 2014). During early childhood, children demonstrate the ability to adhere to a single rule (e.g., sorting by colour), then sequentially switch between two rules (e.g., sorting by colour and then by shape), and eventually transition between sets of higher-order rules (Zelazo et al., 2003).

In general, EF undergoes significant development during early childhood, with notable improvement observed in inhibitory control and cognitive flexibility. The above

three-factor structure is commonly explored in EF research, which offers a nuanced understanding of an individual's cognitive abilities. Incorporating these three domains into research allows researchers and practitioners to evaluate distinct facets of EF abilities during early childhood years when EF experiences substantial improvement. The following section will present an overview of the development and challenges encountered in assessing EF among young children, along with a synopsis of four distinct types of measures designed for evaluating EF in young children.

2.4.5 Assessments of Executive Function

The advancement of research on EF in young children has prompted the development of new assessment tasks and enhancements to existing EF tests (Diamond & Taylor, 1996; Gerstadt et al., 1994). Earlier attempts to measure children's EF involved adapting simplified versions of tasks designed for examining EF in adults, primarily due to challenges in designing effective tests for data collection and the absence of comprehensive measurement tools during early childhood (Espy, 1997; Whitters, 2019). For instance, tasks originally designed for adults and school-aged children, such as the Tower of Hanoi, are later employed among preschoolers to evaluate diverse cognitive skills, including working memory and rule application (Simon, 1975; Welsh et al., 1991). These tests may be of little relevance to young children and lack normative information concerning developmental expectations (Anderson, 2001). Additionally, due to its synonym with frontal lobe function, early research on EF in children focused on assessing executive dysfunction within atypical participant populations, such as those

with traumatic head injury, attention deficit hyperactivity disorder, autism, and premature birth (Barkley, 1997; Espy et al., 2002; Levin et al., 1996; Zelazo et al., 2002). Tasks applied to these populations may not apply to typically developing children (Barkley, 1997; Espy et al., 2002; Levin et al., 1996; Zelazo et al., 2002). Tasks like the Wisconsin Card Sorting Test (WCST) were originally designed to measure frontal lobe dysfunction in individuals with acquired brain injury and are less commonly used in typically developing children.

To address the challenges previously described, later studies have explored four categories of measures within the extensive array of instruments and tasks designed to assess EF and its components (Silva et al., 2022), which include self-report questionnaires, informant-report questionnaires, laboratory-based assessments, and neuropsychological tests. Direct cognitive tasks in the preschool period include measures such as the Dimensional Change Card Sort task (Zelazo et al., 1995), Luria's tapping and hand tasks (Luria, 1980), Go/No-Go (Livesey & Morgan, 1991), and the Day-Night Stroop-like task (Gerstadt et al., 1994). While laboratory-based measures and neuropsychological tests exhibit high construct validity, they may lack ecological validity (Anderson, 2002; Uddin, 2021). These assessment processes require well-structured and relatively quiet environments with minimal distractions, making them seldom accessible in real-life settings (Sbordone, 1996). Some of these measures are more universally accepted, underscoring the variation among individual researchers in determining the most effective tests for assessing executive function (Carlson, 2005).

Caution is needed when interpreting the performance of EF measures in young children (Fletcher & Taylor, 1984).

2.4.6 Summary

To summarise, executive function (EF) is a critical set of cognitive processes that manage goal-oriented behaviours, particularly those requiring self-control and attentional regulation, and these processes are foundational for children's cognitive, social, emotional, and physical development as well as academic achievement. EF encompasses multiple components, including inhibitory control, working memory, and cognitive flexibility, which are developed rapidly during early childhood and are influenced by both brain development and educational experiences. Various theoretical models, including work by Luria, Miyake, and Friedman, set the foundation for a modern understanding of EF. Yet, the absence of consensus on EF definitions complicates assessments, highlighting the importance of a holistic understanding of EF beyond individual components. Ongoing challenges persist in assessing EF in children due to a lack of methodological agreement on EF, and caution is advised when interpreting EF measures in young children.

2.5 Chapter Summary

This chapter explored the four domains of development that are the focus of this study: social competence, emotional competence, language, and executive function. These areas were selected because of my interest in understanding key affective and cognitive capabilities that are highly associated with children's prosocial behaviour. As

identified in the importance section of each capability, there are strong associations between the selected capabilities as well as a range of other outcomes for children, such as school readiness, academic achievement, and overall well-being.

Theories, definitions, and common approaches to the assessment of each area were discussed to set the context for the present research. However, the present research does not seek to look at these constructs individually through variable-centred approaches, as is the case for a range of research that has been presented in this chapter. The next chapter, therefore, looks at social competence, emotional competence, language, and executive function through holistic and person-centred perspectives. This includes how child-specific identities, such as gender or ethnicity, may relate to each area of children's capabilities, which has implications for considering and examining fairness in the assessment of these constructs. The next chapter specifically considers and presents the methodological approaches used to research children's learning and development holistically and fairly.

Chapter 3 Holistic Views of Development and Observed Subgroup Variations

Chapter 2 described each key area of interest for the present study to provide background on the theoretical foundations, common definitions and descriptions, and issues of measurement for each construct. Describing each area separately was not intended to imply that they function as independent characteristics. A key premise of the present research is that capabilities operate interactively within the cognitive and affective or mental and behavioural systems, as has been highlighted in the importance section for each area in the preceding chapter.

The present chapter is a further review of the literature focused on development as a holistic, integrated concept and considers young children's interdependent and interconnected capabilities, adopting a holistic perspective. This chapter examines critical methodologies for clustering young children's capabilities (social, emotional, language, and executive function), emphasising a person-centred, holistic approach. This includes an overview of research that has used person-centred methods to examine children's learning and development. To better understand children's capabilities, the role of gender and ethnicity (as observed grouping variables) is explored in terms of how these may impact the measurement of constructs. The chapter ends by overviewing methodological approaches to address the comparability of measures across observed subgroups (gender and ethnicity) and results of the person-centred model (unobserved subgroups, which will be explained below). Key measurement invariance methods highlighted in this chapter include multiple-group

confirmatory factor analysis (MG-CFA) and differential item functioning (DIF). Taken together, the literature presented in this chapter sets the stage for the present study.

Please note that the acronym for measurement invariance (MI) and other key methodological approaches are presented in full at key points throughout the chapter to aid the reader in clarity for each relevant section.

3.1 The Interwoven Nature of Social, Emotional, Language and Executive Function

Capabilities

A guiding principle of holistic views of a child's learning and development is the interconnectedness of various dimensions such that capabilities influence other capabilities as children learn and grow (Johnson, 2023; Magnusson, 1995). For example, higher social-emotional competence is linked with reduced behavioural and emotional issues cross-sectionally (Zava et al., 2024) and longitudinally (Goodman et al., 2015), while improved behaviour management skills positively correlate with increased engagement in preschool activities among young children (Martinsone et al., 2022; Ornaghi et al., 2017). Similarly, deficiencies in social-emotional competence can lead to problematic behaviour, emotional difficulties, and peer rejections (Fantuzzo et al., 2003). Peer relations and prosocial behaviours that contribute to the development of social competence are associated with other key variables, such as executive function (O'Toole et al., 2017). For instance, young children's prosocial behaviours, such as sharing, require inhibitory control and cognitive flexibility, which are central to executive function. Preschool children demonstrating greater prosocial behaviour and peer

acceptance may develop higher executive function abilities through social interaction competence, including emotional and behavioural regulation and empathy (Hughes & Ensor, 2011; Masten et al., 2012; Moriguchi et al., 2020; Paulus et al., 2015). Moreover, the development of executive function in early childhood significantly predicts achievement in social-emotional skills over time (Diamond, 2006, 2016; Diamond & Taylor, 1996) and is cross-sectionally associated with the development of social-emotional competence and academic success (Brock et al., 2009; Horowitz-Kraus et al., 2016; Kim et al., 2013). Poor executive function among young children may manifest as internalising or externalising behaviour problems (Cumming et al., 2022; Flouri et al., 2017; Thorell et al., 2004). Furthermore, preschoolers' language abilities contribute to developing emotional understanding (Denham & Grout, 1993; Denham et al., 1994). What becomes clear from these examples is that social and emotional capabilities, executive function, and language abilities are intricately intertwined, developing concurrently with patterns, connections, and interconnections. This interwoven nature underscores the significance of viewing children's overall well-being through a holistic lens, considering multiple aspects of development.

Understandings of holistic development are typically explored theoretically and have often been applied in early learning theories and educational approaches. Holistic development can also be explored methodologically to better understand the interrelationship and dynamic influences of children's capabilities, including how patterns of development occur within and across children. The two subsequent sections

explore each of these aspects of holistic development in turn (i.e., theories and methods). Importantly, understanding the extent to which holistic development is viewed as uniform, linear, fixed, predictable, and universal, or whether holistic development is viewed as dynamic, non-linear, differential, contextual (cultural), and with multiple trajectories, depends on the theories, methods, and perspectives taken.

3.2 Theoretical Views of Holistic Development

The interconnectedness of children's learning and development necessitates an integrated, holistic view that acknowledges the multifaceted growth and learning. Such holistic views recognise children as complete entities and underscore the dynamic interaction between various aspects of children's learning and development alongside influences that drive learning and development. Holistic views also provide valuable insights for theoretical frameworks and empirical studies, revealing the integrated, dynamic, and multifaceted nature of child learning and development, while discerningly addressing both similarities and distinctions of children's learning and development in their application to early childhood education and teaching practices. Key theoretical perspectives reflecting holistic development include dynamic systems and ecological perspectives, as well as theories that focus on individual child development. These perspectives, in turn, have informed holistic teaching and learning approaches in early childhood education. Within New Zealand, Western knowledge about children's learning and development is complemented by Indigenous views of child development. Each of these is described to further set the context for the present research.

Holistic learning theories acknowledge the interconnectedness of all facets of children's development and learning and emphasise the holistic development of individuals through fostering connections between learners, teachers, the community, and the broader world while integrating real-life experiences into learning (Johnson, 2023; Marjolein & Maddie, 2023). Dynamic systems theory posits that a child's mind, body, and physical and social surroundings form an integrated system that guides their learning and development (Fischer & Bidell, 2006; Fischer & Daley, 2006). This theory posits that children's physical and social environments influence their universal development. Any alteration within this system disrupts the existing relationship between the organism and its environment (Fischer & Bidell, 2006; Thelen & Smith, 2006). Bronfenbrenner's bio-ecological model (Bronfenbrenner, 1979, 1986; Bronfenbrenner & Ceci, 1994) focuses on the bi-directional interaction between individuals and various levels of their environment, highlighting the environment's role as a complex system influencing individual differences and the influence of the individual on the system. Bronfenbrenner also illustrated the nested system as constantly evolving, reflecting dynamic system perspectives in child development. Similarly, dynamic perspectives envision children's development as evolving patterns with stability existing at any given point, as illustrated by Fischer and his colleagues, who depict development as a network of interconnected strands across physical, cognitive, and social-emotional domains (Fischer & Bidell, 1998; Fischer & Bidell, 2006; Fischer & Daley, 2006). These perspectives offer insights into understanding individual

differences and the complexities of child development and highlight the roles of the dynamic development process and interactions with the environment.

Alternative developmental perspectives hold implications for understanding the distinctions between typical and atypical individual functioning from holistic viewpoints, acknowledging children's learning and development as intricate, dynamic processes.

Cairns (1979) suggested that individual differences originate from genetic and biological factors as the primary influencers of functioning. Piaget's cognitive theory stands out as a significant framework for conceptualising individual development, presenting it as either stagewise or involving continuous change (Piaget, 1948; Piaget, 1983). For Piaget, the integration of different areas altered with each successive stage of development; however, development was generally viewed as following a logical or uniform order.

Possibly one of the most influential theories for holistic development in early childhood education is Vygotsky's sociocultural view of development (Vygotsky, 1978). This theory highlights that the dynamic interplay of cognitive, social, emotional, and cultural factors cannot be separated. Notably, Vygotsky emphasises language as a particularly important area of development that mediates across other domains such as social, emotional, and other cognitive functions (Wertsch, 1985).

Current early childhood educational approaches, such as the Reggio Emilia and Montessori approaches, exemplify the holistic learning and education perspectives. Reggio Emilia's approach provides diverse resources and multiple ways of learning,

fostering collaborative research and dialogue among children, their teachers, and parents (Brandao & Theodotou, 2020; Edwards et al., 2011). Children's learning in this framework is shaped by their interactions with various resources and environments, encouraging real-life experiences and exploration through multiple ways and environments (Adams et al., 2016; Marjolein & Maddie, 2023). Conversely, the Montessori educational method prioritises child-centred learning, focusing on the development of a whole child and utilising various materials to facilitate optimal learning (Montessori, 1995). These contemporary educational philosophies underscore the interconnectedness of all dimensions of children's learning and development and promote a holistic approach aimed at nurturing individual capacities and talents, supporting learners' growth and shaping teaching practices in early childhood education settings (Marjolein & Maddie, 2023).

Indigenous views of holistic development in New Zealand are embodied through frameworks like the early childhood curriculum—*Te Whāriki*, Kaupapa Māori approach and the Māori health and well-being model—*Te Whare Tapa Whā* (Gibbs, 2023; Marjolein & Maddie, 2023; Ministry of Education, 2017; Rochford, 2004). *Te Whāriki* identifies various aspects of children's social, emotional, cognitive, physical, cultural, and spiritual learning and development and acknowledges the interconnectedness of these domains from a holistic lens. This holistic view guides a comprehensive early childhood curriculum and helps shape a rich learning environment that nurtures children's capabilities across all domains (McLachlan, 2018; Ministry of Education, 2017).

Moreover, the Indigenous view, such as the Māori worldview (Te Ao Māori), stresses the connections between the wider family and ancestry (whānau and whakapapa) and with the natural world (Rameka et al., 2021). These connections are viewed holistically and inform Māori health and well-being (Ministry of Education, 2009). The Māori health model—*Te Whare Tapa Whā*, encapsulates four aspects of health and well-being: *Taha Tinana* (physical), *Taha Wairua* (spiritual), *Taha Hinengaro* (psychological), and *Taha Whānau* (relational) (Durie, 1994; Rochford, 2004). These indigenous perspectives reflect an understanding of holistic approaches that emphasise the interconnectedness and interdependence of multiple aspects of learning and development, health, and well-being.

To sum up, theories acknowledge the significance of each capability and the interwoven nature of these capabilities. Holistic development theories support and reflect an increased understanding of the integrated way in which children develop competencies, and can help inform the design and delivery of effective early childhood education (Konold & Pianta, 2005). A range of perspectives highlight the interconnectedness of a child's learning and development from a holistic lens, which prioritises dynamic, complex processes of individual functioning and advocates for integrating factors within a theoretical framework.

Research on holistic views of children has often focused on the formulation of issues and outcomes across distinct aspects of learning and development, which are then interpreted within a unified theoretical framework, without necessitating the

simultaneous study of the entire individual system. Such focused research endeavours make an important contribution to understanding the parts that make up the whole. However, methodological advancements support more comprehensive investigations of children's interconnected capabilities of children. These approaches are explored further in the following section.

3.3 Person-Centred Analysis Methods

“Behaviour, whether social or non-social, is appropriately viewed in terms of an organised system, and its explanation requires a ‘holistic analysis’ ”(p. 325).

- Cairns (1979)

The holistic view of development offers a theoretical framework for research design and employing analytical techniques. Correspondingly, David Magnusson's discussion (1995) on individual development through an integrated, holistic interactional lens advocated for a person-centred approach to capturing the complexity of individual development, recognising that people may share combinations of characteristics that meaningfully cluster together. Magnusson emphasised the importance of studying patterns and configurations within and across individuals (Magnusson, 1995; Magnusson & Törestad, 1993). Building on Magnusson's recommendations for research on children's learning and development, previous studies have revealed the presence of identifiable patterns of children's capabilities (Abenavoli et al., 2017; Christensen et al., 2020; Collie et al., 2019; McLaughlin et al., 2018; Russell et al., 2022). Accessing methodological tools that align with the complexities of individual functioning and

development processes facilitates advancements in research. Adopting a person-centred approach to research reflects a holistic understanding of theoretical concepts. Moreover, investigating the composition and distinctiveness of various groups can yield insights valuable for early intervention.

In contrast to the variable-centred approach, which addresses the associations between variables, the person-centred approach offers a holistic and dynamic perspective that complements and extends variable-centred research. The person-centred approach views individuals as an integrated whole that comprises multiple attributes and considers intraindividual variation in different components. The person-centred approach examines shared attributes or relations among attributes rather than as a summation of variables (Bergman & Magnusson, 1997; Magnusson & Allen, 1983, p. 17). This approach delineates individual differences by examining the interaction between variables, specifically by identifying groups of individuals who share similar functioning at the organismal level while differing from others within the same level (Larsen & Hoff, 2006; Magnusson, 2003). Through person-centred analysis methods, individuals are grouped together based on patterns of the capabilities across relevant variables, offering insights into the homogeneity and heterogeneity in individuals' capabilities. Examples of person-centred analytical methods include cluster analysis (i.e., k-means cluster), latent class analysis (LCA), and latent profile analysis (LPA). These methods have been employed across a range of disciplines, including early learning, to explore similarities and differences in young children's social-emotional and cognitive

capabilities, academic learning and health status (Collie et al., 2019; Denham et al., 2012; Russell et al., 2022; Sandilos et al., 2019; Thomson et al., 2017; Thomson et al., 2019).

Table 3-1 presents an overview of five key studies employing a person-centred approach to categorising social-emotional learning, school readiness, and developmental and health difficulties among preschool-aged children. Notably, studies included in Table 3-1 were not selected through a systematic literature review process. Instead, these studies were selected and described, because they offer insights into the nature and types of shared patterns in children's learning and development using person-centred methodologies from quantitative perspectives, with particular relevance for the capabilities studied in the present research, such as the inclusion of variables related to social, emotional, language and executive function capabilities and/or similar samples. The existing literature contains additional person-centred studies of child development that are beyond the scope of this review. The chosen studies are illustrative of the types of findings from person-centred approaches and will also be used to contextualise findings from the present research. Key characteristics and findings of each selected study are presented in the table, followed by a narrative description of each study.

Table 3-1

Summary of Selected Person-Centred Studies on Characteristics in Children Aged 4 to 6.5 Years

Citation	Overview of the study	Country	Sample size	Capabilities	Analysis method	Measures of profile indicators	Domains	Number of profiles	Label of profiles	Entropy
Denham et al. (2012b)	The study identified distinct groups of young children's emotional knowledge, self-regulation, and problem-solving patterns and observed social and emotional behaviour. Sequentially, the study also revealed that group members differed in demographic variables of gender and centre types and in later school success indices.	United States of America	275	Social-emotional learning, including emotional knowledge, emotional and social behaviours, social problem-solving, and self-regulation	K-means clustering	(1) Affect Knowledge Test (2) Preschool Self-Regulation Assessment (3) Social problem-solving: Challenging Situations Task (4) Observed affect and behaviour: Minnesota Preschool Affect Checklist Revised and Shortened	(1) Preschoolers' emotional knowledge, recognition of emotional expression and understanding of emotion-eliciting situations; (2) Strengths and weaknesses in preschoolers' self-regulation; (3) Children's choices of their behaviours and emotions in response to three problematic peer provocation situations; (4) Emotional expression, emotion regulation, and social behaviour	3	(1) SEL Risk; (2) SEL Competent-Social/ Expressive; (3) SEL Competent-Restrained	-
Thomson et al. (2017)	The study examined distinct patterns of young children's social and emotional health and then analysed associations between the patterns and sociodemographic characteristics.	British Columbia, Canada	35,818	Social-emotional learning	Latent profile analysis	The Early Development Instrument	(1) Overall social competence (2) Responsibility and respect (3) Approaches to learning (4) Readiness to explore (5) Prosocial behaviour (6) Anxiety and fear (7) Aggression (8) Hyperactivity and inattentiveness	6	(1) Overall high social-emotional functioning, (2) Inhibited adaptive (3) Uninhibited-adaptive, (4) Inhibited-disengaged, (5) Uninhibited-aggressive/hyperactive (6) Overall low social-emotional functioning.	0.95

Citation	Overview of the study	Country	Sample size	Capabilities	Analysis method	Measures of profile indicators	Domains	Number of profiles	Label of profiles	Entropy
Sandilos et al. (2019)	The study identified latent subgroups of children's school readiness skills at the beginning of the school year and then examined the transition analysis of these latent subgroups over the school year.	United States of America	899	School readiness	Latent profile analysis and latent transition analysis	(1) Early Assessment in Mathematics (2) Assessment of Number Sense and Place Value developed by the author; (3) The LENS- preschool and kindergarten science standards; (4) The Academic Rating Scale- Mathematics (5) Executive function direct assessments—Backwards digit span, the Head Toes Knees Shoulders, the Pencil Tap test); (6) Teachers reported on each child's behaviours and approaches to learning using the Teacher-Child Rating Scale	(1) Math (2) Science (3) Executive function (4) Behaviour (5) Approaches to learning	3	(1) Low-range academics/ EF, mid-range behaviours/learning approaches; (2) Mid-range academics/ EF, low-range behaviours /learning approaches; (3) High-range academics, EF behaviours and learning approaches	0.863
Collie et al. (2019)	The study understood profiles of children's social and emotional behaviours, followed by the exploration of the links to education outcomes	Australia	100,776	Social and emotional behavioural	Latent profile analysis	Australian version of the Early Development Instrument	Behaviours (1) Cooperative (2) Socially responsible (3) Helpful (4) Anxious (5) Aggressive-disruptive	4	(1) SE-Prosocial ; (2) SE-Anxious ; (3) SE-Aggressive ; (4) SE-Vulnerable	0.933
Russell et al. (2022)	The study explored the profiles of health outcomes of New Zealand children, followed by analysing the associations between socio-environmental predictors and profiles of health status	New Zealand	6109	Health and difficulties in development	Latent profile analysis	(1) The Strengths and Difficulties Questionnaire (2) The Affect Knowledge Test (3) The Hand Clap Task (4) The Dynamic Indicators of Basic Early Literacy Skills (5) A scale of gross and motor skills (6) The number of chronic medical conditions and acute health problems	(1) Physical health (2) Motor development (3) Socioemotional and behavioural development (4) Communication (5) Learning	6	(1) Healthy; (2) Early learning difficulties; (3) Early social skills flourishing; (4) Physical health difficulties; (5) Early learning skills flourishing; (6) Difficulties in development	0.83

Note. Entropy is a commonly reported model fit statistic in studies that use latent approaches. The entropies reported across studies suggest that the models and identified profiles provided a good fit to the data. SEL represents social emotional learning, EF represents executive function, and SE means social-emotional.

As can be seen in Table 3-1, Denham et al. (2012) adopted a person-centred approach to generate profiles of social-emotional learning (SEL) among a sample of 275 young children in the United States, with a focus on areas of emotional knowledge, social and emotional behaviours, social problem-solving, and self-regulation. Measures included the Affect Knowledge Test (AKT; Denham, 1986), the Preschool Self-Regulation Assessment (PSRA; Smith-Donald et al., 2007), the Social Problem-Solving: Challenging Situations Task (Denham et al., 1994), and the Adapted Affect and Behaviour: Minnesota Preschool Affect Checklist and Revised and Shortened (MPAC-R/S; Denham & Burton, 1996; Denham et al., 1991). Children were categorised into three groups using a non-latent analytic k-means clustering technique, where entropy is not reported: SEL Risk (43%), SEL Competent-Social/Expressive (29%), and SEL Competent-Restrained (28%) groups. Children in the SEL Risk group exhibited limited emotional understanding, greater challenges with behavioural control and compliance, and a preference for angry-aggressive social problem-solving strategies. They also displayed moderate levels of emotional positivity, productive play, and high levels of negative emotion and aggression, along with deficiencies in emotion regulation and prosocial peer interaction. Children in the SEL Competent-Social/Expressive and SEL Competent-Restrained groups exhibit notable strengths in emotional knowledge and self-regulation compared to those in the SEL Risk group. However, these two groups demonstrated differences in patterns of social problem-solving. Specifically, children in the SEL Competent-Restrained group engaged in less emotional interaction with peers

and fewer prosocial behaviours alongside higher aggressive responses relative to children in the SEL Competent-Social/Expressive group. In contrast, children in the SEL Competent-Social/Expressive group displayed higher levels of emotional expressiveness than the other two groups, making up the emotionality with social problem-solving, emotional regulation, and positive social behaviour.

Thomson et al. (2017) adopted latent profile analysis and delineated six distinct profiles of social-emotional health among 35,818 young children in British Columbia in Canada, as they entered school based on eight social-emotional subscales on the Early Development Instrument (EDI; Janus & Offord, 2007). The eight subscales include overall social competence, responsibility and respect, approaches to learning, readiness to explore, prosocial behaviour, anxiety and fear, aggression, hyperactivity, and inattentiveness. Six profiles encompassed overall high social-emotional functioning, inhibited-adaptive, uninhibited-adaptive, inhibited-disengaged, uninhibited-aggressive/hyperactive and overall low social-emotional function. Fifty-eight per cent (58%) of the children demonstrated overall high social-emotional competence, while around 40% exhibited vulnerabilities in certain areas. These groups were referred to as: Group 2 Inhibited-Adaptive (8%); Group 3 Uninhibited-Adaptive (16%); Group 4 Inhibited-Disengaged (6%); Group 5 Uninhibited-Aggressive/Hyperactive (8%); and Group 6 Overall Low Social-Emotional Functioning (3%). Notably, the uninhibited-aggressive/hyperactive (group 5) and overall low social-emotional function (group 6) groups displayed considerably higher levels of aggression and hyperactivity compared

to others, consistent with previous research (Egger & Angold, 2006). However, while group 5 also showed considerable readiness for exploration, the authors reported that group 6 may represent individuals more susceptible to enduring mental health challenges, potentially benefiting from personalised and sustained interventions throughout childhood (Reef et al., 2011).

Sandilos et al. (2019) conducted a study identifying three-group latent profile solutions to discern patterns of school readiness among 827 young children in the United States. Latent profile analysis was employed to identify these patterns, and profile changes over time were examined using latent transition analyses. The research incorporated children's math and science abilities, executive function, internalising and externalising behaviours, and adaptive behaviours linked with approaches to learning as critical measures. Latent subgroups exhibited similar scores on math, science, and executive function indicators but differed in scores related to behaviour and learning approaches. Group 1 (45%) displayed low scores on academic and executive function measures but mid-range scores on behaviour and learning approaches than other groups. In contrast, Group 2 (16%) exhibited mid-range scores on academics and executive function but low scores on behaviours and learning approach. Conversely, Group 3 (39%) showed high scores across academic, executive function, behaviours and learning approaches. The stability of these groups was moderate over the school year, with transitions between groups associated with the child's background information, such as the child's gender, first language, parents' education, annual income, and the

quality of the teacher-child relationship.

The study by Collie et al. (2019) identified patterns in the social and emotional behaviour (SEB) of over 100,000 Australian young children from a 2012 cohort. Using a latent profile analysis, five areas of social-emotional behaviours, including cooperative, socially responsible, helpful, anxious, and aggressive-disruptive subdomains, were examined using the Australian version of the Early Development Instrument (AvEDI; Janus et al., 2011). This study revealed four distinct latent subgroups among preschool children: SE-prosocial (70%), SE-anxious (12%), SE-aggressive (13%), and SE-vulnerable (5%). The first and largest latent subgroup, comprising 70% of the participants, demonstrated higher-than-average levels of cooperative, socially responsible, helpful behaviour, along with lower-than-average levels of anxious and aggressive-disruptive behaviours. In contrast, the other three latent subgroups exhibited distinct patterns of capabilities. For instance, children in the third latent subgroup presented low levels of cooperative, socially responsible and helpful behaviours but above average levels of anxiousness and aggressive-disruptive behaviour compared to the first group.

Russell et al. (2022) examined the health status of 6,109 young children in New Zealand, using data from *the Growing Up in New Zealand* dataset. Their research covered acute and chronic health problems, motor skills, behavioural difficulties, emotional knowledge, executive function, and early literacy skills. The six latent subgroups identified were healthy, early learning difficulties, early social skills flourishing, physical health difficulties, early learning skills flourishing, and difficulties in

development. More than half of the children (52.6%) were classified as healthy. Notably, the early learning difficulties group (19.5%), identified as the second largest latent subgroup, displayed the lowest and second lowest scores on executive function and early literacy skills, respectively. Children in the early social skill flourishing group (14.5%) demonstrated the lowest scores on behavioural challenges but the highest on emotional knowledge. Conversely, the fourth group (5.6%) scored highest on acute and chronic health problems. Those in the early learning skills flourishing group (4%) exhibited the highest executive function and early literacy skills. Finally, the last group (3.6%) displayed the lowest levels of health across all variables. This research contributes to a holistic understanding of health status and carries implications for healthcare service delivery.

Person-centred studies contribute to a holistic understanding of children's social-emotional and cognitive characteristics, executive function, early learning, and health status. These analytical methods allow for identifying meaningful patterns in children's capabilities and associations with external factors. Despite variations in critical areas of interest among the studies described, a common thread is led by a holistic lens to uncover patterns in children's capabilities, which collectively advocate for exploring these capabilities through a person-centred approach, using non-latent and latent analytic techniques.

The critical contrast between latent class and latent profile analysis, and other cluster analysis methods, lies in the clustering approaches and their underlying

assumptions. The latent methods classify objects or individuals using a probabilistic, mixture modelling approach, which describes data distribution from a model-based clustering approach and allows objects to be members of multiple clusters. Other cluster analysis methods, including k-means clusters, find classification with arbitrarily chosen distances and produce non-empty, exhaustive and mutually exclusive clusters (Brusco et al., 2017). In recent years, the use of latent (model-based) approaches has been viewed as favourable due to improved estimation and classification of distinct patterns. To this end, latent class and profile analysis offer alternative means for assessing model fit and better capturing uncertainty within the classification.

In their review of best practices with latent profile analysis, Spurk and colleagues (2020) note that research exploring shared patterns of capabilities using latent methods can be conducted for different purposes, ranging from purely exploratory to fully confirmatory in nature. Researchers adopting a purely exploratory approach will not make prior assumptions about the number, size, or shape of the profiles, or potential predictors. Others take a confirmatory stance, guided by clear hypotheses regarding these aspects. These variations suggest that, in some cases, researchers may not be able to rely on solid theoretical assumptions to predict the specific number or nature of the profile that should emerge.

Regardless of the purpose and the extent to which prior theory guides the research, the selection of models from latent methods is a methodological and conceptual process. It is a methodological process in which model fit can be evaluated

against a range of statistical criteria to indicate how well the proposed model aligns with the observed data, including measures such as goodness-of-fit indices, comparative fit indices, and root mean square error of approximation (RMSEA), which provide insights into the adequacy of the model in representing underlying structures (Hu & Bentler, 1999). In addition to assessing model fit indices, researchers may review classification statistics, such as the average posterior probability. The average posterior probability refers to the average probability of the class model accurately predicting class membership for individuals (i.e., the likelihood of individual assignment to each latent subgroup) (Muthén & Muthén, 2000). As noted above, a statistic referred to as entropy is used to measure class separation in the model or the degree of classification uncertainty (Shannon, 1948). High entropy indicates that individuals are classified into one latent class with minimal ambiguity, while low entropy suggests that the model struggles to distinguish between latent subgroups. The selection of models from latent methods is also conceptual in that the selection and interpretation of the model must align with the theoretical framework and research questions, ensuring that the latent variables and their relationships reflect meaningful constructs. This requires the researcher to critically evaluate whether the model captures the underlying phenomena of interest and makes sense within the broader context of existing research.

Spurk and colleagues (2020) suggest that when possible, good practice will give some indication of hypotheses regarding the potential outcomes of clustering techniques because even exploratory models should be informed by theory and past

research to determine reasonable models or the number of classes, rather than relying solely on fit statistics, which may indicate several possible models as reasonable appropriate. Observations may suggest the potential absence of distinct sub-populations in the shared patterns, indicating population homogeneity. As noted in previous research (Collie et al., 2019; Russell et al., 2022), large proportions of samples may be homogeneous in their profile (e.g., that aspects of learning and development are related in a common way for a majority of children), while only small proportions of the sample display distinctive patterns of association on indicators of development. For instance, in the 4-group model identified by Collie et al. (2019), 70% of the sample displayed population homogeneity in the social-emotional prosocial profile. However, a small proportion of the sample displayed distinct profiles, with 12% grouped under the social-emotional anxious profile, 13% under the social-emotional aggressive profile, and 5% under the social-emotional vulnerable profile. The findings across the studies described in Table 3-1 contribute to the existing literature on young children's social-emotional behaviours and early learning and offer insights pertinent to informing the present research.

3.4 Variation Among Observed Subgroups: Gender and Ethnicity

Empirically exploring observed subgroups with shared characteristics within a population may provide valuable insights into understanding similarities and differences among individuals. This approach is beneficial for exploring variations by observed subgroups such as gender, ethnicity, or socio-economic status. Research into the patterns associated with observed subgroups helps address health, social and educational equities, bridging gaps related to these observed subgroup populations. Findings from such studies also aid in planning targeted services and programmes tailored to these groups' specific needs, ensuring the delivery of culturally appropriate services.

However, understanding these observed subgroups requires a more profound comprehension of their complexities within the contemporary social and educational contexts in New Zealand. In the present research, gender and ethnicity have been selected as key known observed subgroups due to their impact on children's emerging identities and educational relevance. Both identities are emphasised in the current early childhood curriculum—*Te Whāriki*. The inclusive nature of *Te Whāriki* embraces gender and ethnicity and advocates for children to experience an environment with equitable learning opportunities regardless of gender, ability, age, ethnicity, or background. The Education Review Office (2023) highlights the recognition of ethnic diversity and identity, which can also be recognised with equitable educational outcomes, and the importance of culture and identity for educators.

Research on these identities is important for understanding inclusive practice in early childhood, promoting gender and ethnic equity, and supporting health and well-being across various domains. This includes research that ensures measurement equivalence and comparability of findings across gender and ethnic groups in order to inform equitable practices. Effective and equitable practices consider each child's strengths, context, and needs, providing all children with opportunities to reach their potential.

In the following sections, gender and ethnicity as observed groups are first explored by describing or defining the constructs. Both gender and ethnicity are complex and are treated as social constructs shaped by culturally context-based values, beliefs, and identities. These definitions and descriptions are followed by individual descriptions of differences in social-emotional characteristics, language skills and executive function across gender groups (sex assigned at birth) and ethnic groups. Lastly, the importance and complexity of using a person-centred approach in research investigating the consistency of distinctive patterns of capabilities across gender and ethnic groups are then discussed.

3.4.1 Gender and Differences in Development by Gender

3.4.1.1 Gender.

Research acknowledges the ambiguity of conceptual definitions and measurements between gender and sex (Glasser & Smith, 2008; King et al., 2021). Definitively, gender is considered to be socially and culturally constructed values, beliefs, and identities

deemed suitable for individuals based on cultural contexts, while sex refers to biological mechanisms and physical characteristics (Acker, 1990; Unger, 1979; Unger & Crawford, 1993). Although gender as a social construct addresses conventions, roles, and behaviours that are culturally bound (Krieger, 2003), researchers argue that biological and sociocultural factors are typically inseparable, suggesting that the distinction between these terms should be abandoned (van Anders, 2015; Yoder, 1999). Additionally, the measurements used in research often reflect these intertwined definitions, such as asking participants if they identify as male or female and their perspectives on male-female behavioural disparities (Muehlenhard & Peterson, 2011). Nevertheless, the use of 'male' and 'female' in these inquiries appears to conflate the two constructs, as the former addresses biological sex, whereas the latter pertains to gender differences, leading to a blurred differentiation between these two terms. Consequently, researchers increasingly employ these constructs interchangeably, especially in discussions concerning social issues (Glasser & Smith, 2008; Muehlenhard & Peterson, 2011; Ruble et al., 2006). In the present research, I address the concept of gender in young children using the binary categories of 'boys' and 'girls'; the rationale, issues and limitations of this approach are described in subsequent chapters.

3.4.1.2 Differences in Development by Gender.

Chapter 2 has outlined research and current understanding of four components of child development (social and emotional competence, language, and executive function) in young children. However, there remain gaps in our understanding of how

and why these components may vary across gender groups. Previous studies have examined gender-specific characteristics and variations in young children (Chaplin et al., 2005; Fabes et al., 2001; Rubin, 2003). Starting from preschool, girls score higher than boys in identifying others' emotional states and underlying causes, comprehending more intricate, self-conscious emotions, and demonstrating advantages in kindness and consideration (Bosacki & Moore, 2004; Chaplin et al., 2005). Toddler girls slightly outpace boys in early vocabulary growth (Van Hulle et al., 2004), and as children enter schools, a plethora of research shows that girls consistently outperform boys in reading and writing achievement (Reilly et al., 2019). Extending from emotional knowledge, understanding, and expressions to social competence, preschool girls typically exhibit higher social competence and display more prosocial behaviours than boys, and they also handle conflicts with peers in socially acceptable approaches more often than boys (Garner et al., 1997). By the age of 4, boys tend to display more externalising behaviours than girls, such as higher levels of inattentive, hyperactivity and aggressive behaviours (Keenan & Shaw, 1997; Pianta & Caldwell, 1990; Rubin, 2003; Wichstrøm et al., 2012). Girls, in contrast to boys, demonstrate greater effortful control, including the ability to inhibit impulses and redirect attention away from irrelevant or emotionally arousing stimuli-traits that contribute to girls' better school performance and lower incidence of behavioural problems (Berk, 2013). Nevertheless, conclusive differential associations between a child's gender and executive function in early childhood remain elusive (Grissom & Reyes, 2019).

Extensive literature has illustrated differentiation in characteristics across gender groups, yet controversies persist regarding the extent of differences in development by gender (Barbu et al., 2015; Fenson, 1994). Primarily, given that biological factors operate within social, cultural, and environmental contexts, it remains controversial whether gender, viewed as a social construct, serves as a meaningful source of variation in developmental differences (Berk, 2013; Hyde, 2005). For instance, studies indicate that a child's gender influences language development, with boys often lagging behind girls (Barbu et al., 2015; Rowe, 2012). These differences tend to be attributed to biological factors, such as the functional organisation of the brain, which may favour females in language acquisition, which contributes to gendered-based disparities in language performance (Adani & Ceganec, 2019). This view contrasts with the perspective that regards gender as a social-cultural construct. Second, it remains unclear whether gender influences characteristics in a domain-specific or domain-general manner when considering these characteristics from person-centred approaches. Bando et al. (2024) investigated gender disparities in language, socioemotional and motor skills development among children aged 7 to 48 months across nine countries. Their findings indicated that girls consistently outperformed boys in language and socioemotional development, with no gender differences observed in motor development. These results suggest that gender effects on development are domain-specific, neglecting the interconnected nature of multiple areas of development. While the study aligns with previous research outcomes (Adani &

Cepanec, 2019; Barbu et al., 2015), the variable-centred approach adopted in the research provided limited insights into the interplay between these areas of development. Therefore, to better understand gender-based developmental disparities, it may be useful to explore person-centred approaches that emphasise the holistic nature of children's learning and development.

The person-centred approach may help identify young children's characteristics associated with gender groups to indicate early disparities that may contribute to positive and negative outcomes among unobserved distinct groups of children. Few studies have examined the differences in multiple characteristics between gender groups, although existing person-centred studies have suggested the emergence of distinct social-emotional, language, and executive functions in early childhood (Collie et al., 2019; Russell et al., 2022). Further, exploring differences in development by gender using a person-centred approach allows for the generalisability of the optimal latent group solution across gender groups, facilitating the examination of developmental differences overall and within each profile. Moreover, it is necessary to understand the importance of distinguishing real differences from measures in contrast to differences resulting from measurement biases before making comparisons. This requires the examination of measurement invariance (MI) of the latent group solution across gender groups to ensure their comparability. MI refers to the principle that indicators of a measure or model remain consistent across different conditions (e.g., different time points, observed subgroups) that are not directly related to the attribute being

measured (Millsap, 2011). Examining MI is important, as it is a central assumption of all group comparisons and summary statistics, even if rarely tested. When MI is not achieved, also referred to as measurement non-invariance, or not tested, the inferences drawn from the findings may be misrepresented or inaccurate. A detailed discussion is presented in the section titled 'Comparability of measurement across observed groups'.

3.4.2 Ethnicity and Differences in Development by Ethnicity

3.4.2.1 Ethnicity.

Ethnicity refers to groups sharing a common identity-based ancestry, language, or culture, often influenced by religion, beliefs, customs, and historical experiences such as migration or colonisation, while race is viewed as both a biological characteristic and a social category (Cornell & Hartmann, 1998; Institute of Medicine (US) Committee, 1999; Quintana, 2007). Additionally, ethnicity is often defined as 'cultural practices and outlooks of a given community of people that set them apart from others' (Giddens, 1997, p. 210), making it a measure of cultural affiliation rather than race, ancestry, nationality or citizenship (Statistics New Zealand, 2005). Despite this distinction, ethnicity and race are frequently used interchangeably (Cokley, 2007). Ethnicity is increasingly viewed as a social construct, particularly in analysing health status and other social outcomes in New Zealand (Callister et al., 2007; Morning, 2008). Though complex to identify or measure ethnicity (Callister et al., 2007), ethnicity data is used to report the differences in outcomes, deliver culturally appropriate services and plan services that meet the unique needs of ethnic groups (Statistics New Zealand, 2004).

Policies and social investment initiatives may focus on specific ethnic groups as a way to address historic and systemic disparities (Durie, 2005).

Within New Zealand, for example, the Treaty of Waitangi guarantees rights and protections for Māori (and non-Māori). Despite the foundational legal framework of rights and protections, significant disparities exist in outcomes, including varying health, social and educational outcomes among different ethnic groups, which drives the interest in measuring these disparities to reduce gaps (Callister et al., 2007). However, the presence of multiple ethnicities complicates the comparison of outcomes in New Zealand. Migration, ethnic intermarriage, and the option to record multiple ethnicities in surveys present significant methodological challenges for researchers in identifying, measuring, and categorising ethnic groups for research analysis purposes (Callister et al., 2007).

In the current research, children are classified into single or dual ethnicities based on the Census ethnicity categorisation by Statistics New Zealand (2005). Rationale, issues, and limitations of this approach are described in **Appendix 2 Ethnicity**

Prioritisation.

[3.4.2.2 Differences in Development by Ethnicity.](#)

Children learn the socially constructed meaning of identities (i.e., ethnicity, gender, etc.) early in life, and their characteristics (emotional expression, social behaviour, executive function, and language) are expected to vary by context and culture (Berk, 2013; Lewis et al., 2009; Lewis, 2000). Young children become more aware of cultural

expectations, understanding that different cultures have different expectations and that what is appropriate in one context may not be appropriate in another (Ministry of Education, 2017). Conversely, differences in social, behavioural, and cultural norms may influence child interaction styles (Fabes et al., 2006), where different cultural groups use different repertoires of conflict resolution strategies, and children alter their behaviours to adapt to the norms (Berk, 2013). For instance, children must master their culture's rules for when and how to convey emotion as part of this increasing emotional self-regulation (Berk, 2013). The expression of emotions and situation-based emotions may also vary across cultures, and cultures may modify parental influences on children's emotional competence (Morgan et al., 2010; Pala & Lewis, 2021). Cross-cultural research suggests variabilities in individual emotional display and related skills (Ekman, 2003), whereas similarities are presented in associating photographs of different facial expressions of emotions (Ekman & Friesen, 1971). The study by Girouard et al. (2011) demonstrated that the ethnicity of playmates influences the quality of social interaction among preschoolers, revealing a preference for same-ethnic partners to engage more interactively and more solitarily with a cross-ethnic partner. Specifically, the study showed that Asian Canadians initiated more cooperative (helping or assisting a playmate or the peer largely nonverbally) and conflictual (physical aggression, competition conflict, dissociative verbalisation) interactions, while French Canadians used more affiliative verbalisations (relate a story, ask, or give for information, brief communication, invite, etc.).

In New Zealand, ethnic differences in young children's social, emotional, executive function and language are evident in the GUiNZs. For instance, research by Russell et al. (2022), using a person-centred approach, indicated that Māori and Pacific children are more likely to experience early learning difficulties, physical health and difficulties profiles, which are associated with higher levels of socioeconomic disadvantage. Corkin et al. (2021) found that Asian children exhibit higher levels of cool executive functions compared to European children, aligning with previous research by Lewis et al. (2009) and Sabbagh et al. (2006). Cool executive functions (EF) are the emotionally neutral aspects of EF and primarily focus on cognitive demands associated with academic achievement and intellectual development (Zelazo et al., 2003). These findings underscore the importance of cultural sensitivity through young children's ethnicities when evaluating their characteristics, as these characteristics can be context-independent.

However, the variable-centred approach used in research to examine developmental differences in young children across ethnicities assumes population homogeneity, which may lead to over-generalisation of the findings (Laursen & Hoff, 2006). A variable-centred approach analyses characteristics from a single lens, neglecting the complex interplay of diverse individual characteristics within the populations. Additionally, there is limited research in understanding developmental differences in young children across ethnicities from a person-centred approach, explicitly using latent class or profile analysis. Therefore, exploring children's

characteristics by ethnicity through a person-centred approach is valuable. This method allows for a systematic and quantitative assessment of how much a latent class membership generalises across different ethnic groups (Morin et al., 2016). However, issues may arise when analysing data across different ethnic groups with small subgroup sizes, impacting the choice of analytical methods (Kalkan & Kelecioğlu, 2016).

3.5 Importance of Ensuring Comparability

This chapter ends with a section to explore measurement equivalence and comparability across observed subgroups as key aspects of ensuring fairness in the assessment and interpretation of findings without bias. This includes ensuring the comparability of measures and latent profiles across observed subgroups. The following section further highlights the importance of measurement comparability and fairness before technical aspects of measurement invariance (MI) are described for measures and person-centred models.

Ensuring the comparability of test scores provides implications for fairness, equity, and inclusion in early childhood education (Soland et al., 2022). When comparisons are inevitable to inform policy and decision-making, they should be approached in ways that are appropriate and useful (National Academy of Education et al., 2020, p. 23). Proper implementation of MI enhances comparability by ensuring fairness in interpreting and using results from measures to support conclusions. Fairness issues arise when individual test scores are used without establishing measurement invariance (Kline, 2013). In other words, by ensuring measures perform equally well and hold the

same meaning across different groups, the implementation of MI endeavours to minimise biases in the measures and promote fairness to ensure the measurements assess the intended construct consistently for all children.

Furthermore, accurate and consistent comparisons of observed subgroups' characteristics are critical for sustaining a meaningful equity agenda. In New Zealand social and education contexts, comparisons of characteristics may be a key component of equity initiatives, enabling equitable interpretations of similarities and differences across multiple subgroups. Programmes supporting educational equity must consider young children's overall characteristics across observed subgroups, such as gender, ethnic and socio-economic groups. For instance, funding for early learning services is designed to promote educational equity by addressing disparities among various groups, including children from lower socio-economic communities, those with special needs who come from non-English-speaking backgrounds, and those from languages and cultures other than English. However, differing sizes and compositions of an observed subgroup may present challenges in evaluating and maintaining such comparability. Understanding differences across observed subgroups is essential for ensuring the effectiveness and equity of educational initiatives for all children. On the other hand, equity concerns arise when cut-off scores classify individuals without accounting for differences in a holistic lens. For example, using cut-off scores of the SDQ total difficulties scores, a substantial proportion of preschoolers were disproportionately categorised as falling within the borderline range (scores of 14-16)

(Downs et al., 2012; Koglen et al., 2007). This classification approach risks overlooking the nuanced domain of children's specific strengths and difficulties. Hence, MI can effectively address equity concerns by establishing equivalent constructs and meeting reliability requirements, allowing for accurate comparisons of test scores on the same measure.

Additionally, testing measurement invariances enables the accessibility and inclusivity of diverse subgroups of individuals' characteristics (Rodriguez et al., 2023). Including observed subgroups in testing MI ensures accommodations for diverse ethnicities, languages, and cultures. Education policies and initiatives, such as *Te Whāriki*, consider all children's needs and avoid excluding specific subgroups from evaluations and interpretations. *Te Whāriki* aims to foster a more inclusive environment for young children with diverse characteristics in general education settings. Conversely, including all children necessitates rethinking the application and administration of the educational assessments. The rationale for inclusion may introduce measurement challenges in establishing score validity. Testing MI with greater inclusion of children in diverse ethnic groups minimise biases in the assessed constructs across ethnic groups when used in ethnically heterogeneous samples. Hence, before comparing patterns of young children's characteristics, it is critical to interpret them equally and include all children to address meaningful group differences and provide substantive implications.

3.5.1 Comparability of Measurement Across Observed Subgroups

Following the understanding of observed subgroup variations, it is essential to

examine the comparability of measurement across gender and ethnic subgroups to evaluate the potential challenges in the measurement of young children's characteristics. The diversity of the early childhood population and access to both cross-sectional and longitudinal studies enable comparisons of characteristics across various groups, such as gender, ethnicity and socioeconomic status, by examining how these factors impact children's learning and development (Rutkowski & Svetina, 2014). For instance, the study by Neumann et al. (2021a) in the Growing Up in New Zealand study has explored the validity, generalisability, and comparability of measures of cognitive functioning in early childhood across diverse populations, including children's gender, socio-economic backgrounds and mother-prioritised children's ethnicity. These efforts reflect a growing understanding of the need to explicitly address challenges in measurement when making comparisons across groups (Kim et al., 2017). However, some researchers may overlook or fail to consider potential issues in directly comparing results from diverse populations or different timelines, which may jeopardise the comparability of the test scores and undermine comparisons of characteristics (Mandell & Sackett, 2009; Soland et al., 2022). Hence, it is essential to test comparability across observed groups to support the generalisability of constructs.

A fundamental approach to address methodological challenges in measurements in cross-sectional and longitudinal studies is to test measurement invariance (MI). Introduced by Mellenbergh (1989) and later defined by researchers, MI implies that the measurement properties of a psychometric test are identical or equivalent across

different groups or over time (Millsap, 2011; Millsap & Kwok, 2004, p. 93). MI is essential for assessing hypotheses regarding group differences, whether these involve simple mean differences between groups or complex evaluations of theoretical models across group studies (Kim et al., 2017; Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). In confirmatory factor analysis, MI may be reflected in the factor structure, loadings, intercepts, and error terms associated with indicators of a latent factor across observed groups.

Levels of invariance include full, non-invariance and partial invariance (Putnick & Bornstein, 2016). Full MI is established when a test consistently measures a construct across different groups, indicated by an identical number of factors and all factor loadings and intercepts being held equal across groups (Collins & Lanza, 2010). Specific criteria for evaluating MI are discussed more thoroughly later in this thesis. To put it differently, full MI is established when individuals from different populations interpret and respond to the construct being measured in the same way, such that observed differences are due to the differences in the construct itself rather than differences in the measurement. Conversely, measurement non-invariance occurs when individuals from different groups respond differently to a test item despite having the same level of the construct being measured. The goal of testing and reporting MI is to ensure group differences reflect performance or perspective without bias, or through the use of techniques to minimise bias, to enhance fair interpretation.

For example, Putnick and Bornstein (2016) suggested that configural non-

invariance is a situation where the pattern of item loadings on latent factors differs in two cultures (e.g., at least one item may exhibit variations in the pattern of item loadings in one culture). This suggests the presence of measurement biases, which may lead to erroneous conclusions for group-based comparison purposes (Morin et al., 2016). Alternatively, partial invariance, referring to the practice that accepts some violations of full MI (i.e., when some but not all measurement parameters are equal across groups) (Byrne et al., 1989), can also support the comparison of measures across observed subgroups (Cieciuch et al., 2019; Van De Schoot et al., 2013). Steinmetz (2013) suggested that under partial invariance, the difference in latent means between the groups is considered unbiased when at least two-factor loadings and intercepts are set to be equal across groups or over time. According to Asparouhov and Muthén (2014a), it is acceptable when no more than 20% of parameters are non-invariant across groups for partial invariance.

The steps of invariance typically assessed in MI tests are configural invariance, metric (weak) invariance, threshold (strong) invariance, and strict (residual/error) invariance, which is hardly achieved in reality (Schmitt & Kuljanin, 2008). Configural invariance involves the same number of factor/factors and patterns of factor loadings explaining the variance-covariance matrices associated with the groups' responses. Metric (weak) invariance involves the values of the factor loadings of each variable on each factor being the same across groups. Threshold (strong) invariance requires the factor loadings and item intercepts of the observed variables regressing on the latent

factors to be the same across groups. Threshold (strong) invariance is the minimal level to ensure a meaningful interpretation of latent group means comparisons (Kline, 2016; Robitzsch & Lüdtke, 2023). However, Muthén and Asparouhov (2018) pointed out the rarity of achieving full threshold invariance, especially when dealing with numerous groups. Research often proceeds to group means comparisons when threshold invariance or partial threshold invariance is achieved (Lubke & Muthén, 2005; Meredith, 1993). In terms of questions addressed in MI tests, Vandenberg and Lance (2000) provided examples such as: (1) Are there gender or ethnic differences that affect how individuals respond to instruments in similar ways? (2) Do respondents from different cultures interpret a given measure conceptually similarly?

Research has explored MI in young children's characteristics across various observed subgroups using multi-group confirmatory factor analysis (MG-CFA). This method involves the assessment of the relative fit of increasingly restrictive nested models when factor structure, loadings, thresholds, and error terms are constrained to be equal across groups. Riglin et al. (2024) found evidence of metric MI for latent profiles, as measured by the parent-rated Strengths and Difficulties Questionnaire (SDQ) across age groups from the Millennium Cohort Study (ages 3–17 years) and the Avon Longitudinal Study of Parents and Children (4–25 years). McKown (2019) identified partial threshold invariance in social-emotional comprehension across sex and ethnic groups among 4,419 American children in kindergarten through third-grade using the Social Emotional Comprehension measure (SEL Web). Willoughby et al.

(2012) suggested strong longitudinal MI of six individual executive function tasks (e.g., Working Memory Span, Silly Sounds Stroop, Spatial Conflict and Animal GO/NO-GO) among young children aged 3 to 5 years. These findings show invariance across domains such as emotional symptoms, social-emotional comprehension, and executive function by observed factors such as age, sex, ethnicity, and over time. Importantly, these studies highlight the emerging research that is seeking to better understand the comparability of measures across observed groups using MG-CFA.

Testing MI in young children's characteristics across groups using MG-CFA involves four primary challenges. First, inadequate demographic reporting often lacks sufficient detail in grouping variables, leading to inaccuracy in invariance test results (Marsh et al., 2009). Second, inconsistencies in model fit criteria and limited evaluators for models using estimators optimised for non-normally distributed data complicate invariance assessment. These inconsistencies arise from reliance on varied criteria, such as the significance of changes in chi-squared χ^2 (Byrne et al., 1989), changes in comparative fit indices (CFI) for nested models (Rutkowski & Svetina, 2014) and alternative indices like changes in root mean squared error of approximation (Δ RMSEA) and changes in standardised root mean square residual (Δ SRMR) (Chen, 2007; Meade et al., 2008), compounded by shifting standards and evaluators, and the inadequacy of a single criterion for all scenarios. Third, data sparseness, particularly in mixture modelling, impacts parameter estimates when subgroup sizes are small, often as little as 5% of the overall sample (Collins & Lanza, 2010). This may reflect overall or group sample size

and low endorsement of rating scale options, which may be exacerbated where MI is assessed within LPAs, where unobserved subgroup profiles are established within the population. Fourth, MG-CFA methods may be overly sensitive to trivial levels of non-invariance, detecting minor or insignificant differences that do not meaningfully impact the overall model or its interpretation, necessitating strong justification for model acceptance (Khademi et al., 2023). These limitations highlight the need for methodological precision and thoughtful application to ensure reliable MI assessments in the current research.

3.5.2 Comparability on Person-Centred Model Level

Testing measurement invariance (MI) is also applicable at the latent class or profile model level to assess the comparability of similarities and differences across model solutions (Morin et al., 2016; Olivera-Aguilar & Rikoon, 2018). Collins and Lanza (2010) and Masyn (2017) extended the definition of MI to a latent class framework. In LCA, MI is achieved when individuals belonging to the same latent class have the same probability of providing any given observed response pattern; that is, they respond in a statistically similar way, regardless of which group they come from (Collins & Lanza, 2010). Full MI in LCA is established when the number of latent classes and all item-response probabilities are identical across groups (Collins & Lanza, 2010, p. 118). This allows for consistent interpretations across all groups and offers a simpler model than one where measurement varies across groups. Comparisons at the model level aim to identify the potential effects of MI on model conclusions and enable invariance

assessment when indicators cannot be examined at the construct level.

Testing MI in LCA/LPA models can be approached in various ways, such as multiple-group LCA and the implementation of covariate analysis. Collins and Lanza (2010) discussed using multiple-group LCA to assess group similarities and differences at the person-centred model level. The simplest method involves treating groups as separate data sets and fitting models to each group individually. This process includes exploring general latent structure differences to determine whether the identical number of latent classes holds across groups, then establishing the invariance of item-response probabilities by comparing model fits. If both steps show a good fit, MI is established across groups, and the more parsimonious model is selected. Similarly, Morin et al. (2016) outlined a three-step process for comparing group similarities and differences in LPA: (1) ensuring the same numbers of latent profiles in each group; (2) testing the similarity of continuous profile indicators' means and variances in a specific profile, and (3) confirming the consistency of latent profiles in each group. Additionally, establishing relationships with covariates is crucial for validating the construct of extracted profiles (Marsh et al., 2009; Muthén, 2003). Masyn (2017) developed a stepwise procedure to detect the direct effects from latent class variable predictors to latent class indicators (measurement non-invariance or DIF) in mixture models using the Multiple Indicators Multiple Causes (MIMIC) modelling technique. This involves (1) enumerating classes in an unconditional model using observed latent class indicators without incorporating covariates; (2) classifying individuals and estimating classification

errors based on the model from step 1; (3) estimating associations between predictors and latent class membership based on information from step 2. Other methods for testing MI in person-centred models exist but are outside of the scope of this discussion.

Limited research has assessed MI at person-centred model levels for young children's development characteristics across observed groups using a multiple-group approach. For example, Garraza et al. (2011) explored gender (boys and girls) differences in patterns of emotional risk factors among 18,437 children aged 5 to 22-year-old with severe emotional disturbance, using multiple-group latent class analysis (MG-LCA) across gender groups and phases (1993-2000) of the Children's Mental Health Initiative (CMHI) in the United States (Clogg & Goodman, 1984; Collins & Lanza, 2010; McCutcheon, 2002). Specifically, by extending research by Walrath et al. (2004), they identified a 4-class model of emotional risk factor patterns in Phase One: high-risk (highest scores in risk factors of physical and sexual abuse, runaway, suicide attempt, drug/alcohol use, sexually abusive), abuse (second highest in risk factors of physical and sexual abuse, low in other risk factors), status-offence (second highest in risk factors of runaway and drug/alcohol use, low in other risk factors), and low-risk (low scores in all risk factors). The comparison on the model level shows that girls were more likely to be in the high-risk groups than boys. Other studies using MG-LCA/MG-LPA include research on drinking behaviours among U.S. adolescents (Chung & Anthony, 2013), substance use disorder (SUD) across U.S. rural and urban adults aged 18 years and

older (Brooks et al., 2017) and racial differences on mental health and cognitive symptomatology among U.S. older adults (Min et al., 2023). Most studies testing MI in person-centred models focus on adult participants in U.S. contexts and are deficit-based. In New Zealand, research has explored MI on measures of children's characteristics across observed groups, such as gender, ethnicity, or countries (Neumann et al., 2021b; Rubie-Davies et al., 2016; Simsek et al., 2021; Wilson et al., 2023). However, studies testing MI at the person-centred model level are scarce. Therefore, exploring MI at the person-centred model level within the New Zealand context represents an important next step in the New Zealand research literature.

Limitations of the multiple-group approach include the sample size and the interpretations of MI when full MI does not hold across groups. The nature of multiple-group latent class or profile analysis compares latent class or profile characteristics across groups or time points, starting with the assumption that population membership is known and drawn from two or more subpopulations or time points. LCA and LPA are considered 'large sample' analytical methods, with models and fit statistics performing well with sample sizes greater than 500, often showing high classification accuracy (Finch & Bronk, 2011). However, models may not be adequately identified with sample sizes smaller than 300 (Henson et al., 2007). Additionally, the latent classes/profiles require sufficient statistical power to detect differences in the outcomes and ensure the differences are meaningfully interpreted. Where full MI at the model level does not hold, partial invariance can be implemented by constraining some item response

probabilities to be equal across groups in LCA or mean and/or variances in LPA. Fit indices may not always support the assumption of MI, so considerations of parsimony, model identification, and interpretations must be handled cautiously.

An alternative method of testing MI in person-centred models is differential item functioning (DIF) analysis (Asparouhov & Muthén, 2014a; Bakk, 2024; Bettencourt et al., 2022; Masyn, 2017). Among the methods for testing DIF effects, the Multiple Indicators Multiple Causes (MIMIC) modelling is the most widely used technique. This method treats group variables (gender, ethnicity, countries, etc.) as covariates of the latent group-membership variables and their factors or indicators, allowing for the assessment of the direct effects of the covariates on the indicators of LCA or factors of the LPA in the current research. Non-significant direct effects suggest MI of the latent profile factors held at the model level. In contrast, significant direct effects suggest the presence of measurement non-invariance or DIF of the latent profile factors. When DIF is detected, researchers should carefully consider the comparability of the optimal solutions across observed subgroups. Cautions are needed when considering person-centred models with covariates. As Lubke and Muthén (2005) warn, “It is important to realise that the assigned class membership of a participant is model dependent and not an innate quality of the participant ” (p. 37). The limitation of this approach is that if qualitative changes occur in the nature of the groups as the inclusion of covariates in the LCA/LPA, then the assumption that covariate effects only the latent probabilities is violated, which may distort the interpretation of groups (Marsh et al., 2009). The

judgment of whether MI holds sufficiently is not only based on the statistical criteria but also on conceptual importance. Additional details and discussion about approaches to MI used in the present study are presented in Chapters 4 and 6.

3.5.3 Chapter Summary

The present chapter discusses a holistic view of young children's characteristics and considers young children's interdependent and interconnected capabilities holistically. This chapter emphasises a person-centred, holistic approach as a critical methodology for clustering young children's capabilities, including an overview of research that has used methods to examine children's learning and development in this way. To better understand children's development, the role of gender and ethnicity (as a proxy variable) is explored in terms of how variations may present in young children's capabilities and impact the comparability of profiles at the person-centred model level. The chapter ends with a brief overview of methodological approaches to address measurement comparability across observed subgroups.

Building upon theories and prior research, the current thesis explores young children's characteristics through a person-centred, integrative, strengths-based approach. It identifies and characterises groups of young children with shared patterns of capabilities through LPA. This is important as it posits the existence of a limited number of unique patterns in associations between variables and emphasises individual differences by examining the meaningful patterns within a sample. To extend these analyses, the current thesis further examines whether the covariates, such as the child's

gender and ethnicity, introduce DIF to the latent profile factors. Addressing these potential biases is important for ensuring fairness in the interpretation of children's capabilities among diverse populations, thereby supporting equitable educational outcomes. Detailed methodologies are presented in Chapter 4.

Chapter 4 Methodology

The purpose of the thesis was to explore the shared patterns of young children's capabilities in New Zealand and examine the comparability of these patterns across gender and ethnic groups. This chapter briefly introduces the Growing Up in New Zealand study (GUiNZs) as the most suitable dataset to be drawn upon for this investigation. It then outlines the research aims, questions and hypotheses, sample, ethical considerations, and selected measures. The chapter concludes by detailing the analytical procedures and statistical analysis techniques employed in the current research.

4.1 The Growing Up in New Zealand study

The GUiNZs sample reflects the diversity of the New Zealand children in terms of key socio-demographic and ethnic characteristics. Participants (mothers/caregivers and their newborn children) were recruited in 2009 and 2010 in the Auckland, Counties Manukau, and Waikato District Health Board regions. The cohort sample ensured sufficient statistical power for analysing capabilities within subgroups by ethnicity and household deprivation. Written informed consent was obtained from all participating parents, who also gave their consent on behalf of their children. A total of 6,822 mothers consented to participate in antenatal interviews between 2009 and 2010, with 6,156 children completing the 54-month follow-up interviews. A detailed account of the baseline study design and recruitment procedures has been provided elsewhere (Morton et al., 2013). Today, the GUiNZs sample comprises multiple Data Collection

Waves (DCWs), spanning from the Antenatal phase (DCW0) to the Extreme Weather Events phase (DCW13EW). A balanced set of age-appropriate information across the interconnected areas of children's development was gathered at each wave. This information was provided by various informants (such as the child's mother) and gathered through direct observations of the child by the researchers and developmental and anthropometric assessments. Biological samples were additionally taken from the child in 2013-2014. Insights acquired from GUiNZs regarding the factors influencing children's development and well-being have enabled health and education researchers to construct a detailed and nuanced understanding of children's characteristics and have the potential to inform policy for the benefit of New Zealanders (Morton, 2019).

The GUiNZs datasets gathered in the 54-month assessment wave were selected for the current work as they offered a range of variables that provided good conceptual age-appropriate coverage of young children's learning and development capabilities related to social, emotional, language, and executive function skills prior to school entry. These initially included nine specific variables, representing young children's prosocial behaviours, peer problems, conduct problems, hyperactivity-inattention, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function. Due to an issue with the accessibility of data for the three subscales of one of the measures, and after a thorough discussion with supervisors, I dropped three variables representing children's social and emotional difficulties: peer

problems, conduct problems, and hyperactivity-inattention. Additional information about this decision was described in 4.6.2 'Children's Capabilities', explaining the Strengths and Difficulties Questionnaire (SDQ). This decision meant that, based on the reviewed literature and available data, the current research addressed children's capabilities across the following six domains: prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function from observations and assessments administered at the GUiNZs 54-month data collection wave.

4.2 The Current Research

The current research sought to identify shared patterns of capabilities among young children before school entry in New Zealand, as well as to assess the comparability of these patterns across children's gender and ethnic groups. The research was exploratory, cross-sectional, and correlational in nature. These secondary analyses were conducted in two sequential phases, which are detailed in this section. To further clarify the focus and objectives of each research phase, Table 4-1, '*Summary of Research Aims and Questions*', outlines the specific purposes and corresponding research questions addressed in each phase.

Table 4-1

Summary of Research Aims and Questions

	Phase 1	Phase 2
Research aims	<ul style="list-style-type: none"> · Explore whether groups displaying distinct shared patterns of capabilities (i.e., prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function) are observed among young children prior to school entry in New Zealand. 	<ul style="list-style-type: none"> · Explore the comparability of the shared patterns of children's capabilities (i.e., prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function) across child gender and ethnic groups.
Research questions	<ul style="list-style-type: none"> · To what extent do the hypothesised factor structures of measures of capabilities (i.e., prosocial behaviours, emotional symptoms, emotional knowledge and expressive language) fit the observed data¹ from parents and children participating in the GUiNZs 54-month data collection wave? · Do children's social, emotional, language and executive function capabilities display distinct latent subgroups with shared patterns of association in the population? · What are the dominant patterns of social, emotional, language, and executive function capabilities displayed by each latent subgroup? 	<ul style="list-style-type: none"> · Do measures of capabilities (i.e., prosocial behaviours, emotional symptoms, emotional knowledge and expressive language) display Measurement Invariance (MI) across child gender and ethnic groups? · Are the covariates (i.e., the child's gender and ethnic groups) potential sources of differential item functioning (DIF) or measurement non-invariance (MNI) to the latent profile factors? · Is the optimal solution that best describes the population as the final model generated in Phase 1 comparable across the child's gender and ethnic groups?

Note. ¹ The observed data represent available responses for at least one item in the measures of prosocial behaviours, emotional symptoms, emotional knowledge and expressive language.

Phase 1 aimed to identify distinct latent subgroups of children displaying shared patterns of young children's capabilities in the 54-month GUiNZs dataset, including prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function. To achieve this aim, the research was guided by three research questions. Each of these questions offered implications for the analytical procedures briefly mentioned below and described in further detail later in 4.7.3 *Analytical Process*. The research questions are as follows:

As a necessary step in the measurement process, the first research question in Phase 1 addressed the construct validity of the available measures of capabilities using confirmatory factor analysis (CFA). The second question sought to reveal whether groups of children displaying distinct shared patterns in these capabilities could be identified. This question was addressed using latent profile analysis (LPA) to generate latent group models exploring up to six subgroups of possible profiles informed by prior research on children's capabilities using a person-centred approach. An optimal number of distinct groups will be identified based on 1) a priori statistical criteria, 2) a theoretical understanding of the interrelatedness of the targeted capability/skill domains, and 3) prior research. Once the optimal number of groups of children sharing distinct patterns of capabilities is identified, the third question sought to describe and interpret these dominant patterns of their capabilities (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function) within each latent subgroup.

In Phase 2, the current work adopts two techniques to confirm the comparability of measurement across child gender and ethnic groups at the construct and model levels. Firstly, at the construct level, to confirm the comparability of constructs assessed for validity in Phase 1, measurement invariance (MI) of the selected measures of capabilities (prosocial behaviours, emotional symptoms, emotional knowledge, and expressive language) will be assessed across gender and ethnic groups through multiple-group confirmatory factor analysis (MG-CFA). This aimed to ensure the constructs of the available measures were understood equivalently across the child's gender and ethnic groups. Secondly, to confirm the comparability of the LPA profiles generated in Phase 1, the Differential Item Functioning (DIF) approach using Multiple Indicators Multiple Causes (MIMIC) modelling will be used to assess whether the covariates (i.e., gender, ethnicity) serve as potential sources of DIF for the latent profile indicators (i.e., indicators of prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function) in the LPA model. Where appropriate, in a final model, subsequent analyses will be conducted to compare patterns of young children's capabilities across gender (i.e., boys and girls) and ethnicity (i.e., Māori, Pacific, Asian, European, and Other ethnicities). The investigation of the impact of DIF effects by covariates interacting with latent profiles was examined to understand young children's capabilities among diverse populations, which further supported fairness in assessment.

Together, these research questions and associated analysis techniques were

intended to uncover the distinctive profiles and variations among children prior to school entry in New Zealand, allowing for a more nuanced understanding of how these capabilities may be understood, and their generalisability by child gender and ethnicity.

4.3 Hypotheses

Although the current work was exploratory in nature, the current research drew from previous research and child developmental theories to form the research hypotheses related to the number and nature of profiles that may be expected and the relationship with observed grouping variables. An important aspect of drawing hypotheses included careful consideration of the nature of data available from the GUiNZs 54-month dataset that represents six selected domains of children's capabilities (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function) and how these might be grouped or associated under the larger areas of children's social, emotional, language and executive function development.

Theoretically, the latent profile analysis (LPA) examined in Phase 1 will support the identification of distinct subgroups representing different patterns in the associations between children's capabilities. However, as illustrated in Table 3-1, *'Summary of Selected Person-Centred Studies' Characteristics,* and discussed in Chapter 3, previous research on capabilities in young children has identified between three and six latent profile groups. Distinct patterns emerged in only small groups, characterised by differences in social, emotional, behavioural, academic, or other health difficulties.

Associations between capabilities remained largely homogenous, reflecting low, medium, or high scores in capabilities across the remaining domains. Given the nature of the sample and range of capabilities assessed, the current research hypothesised that subgroups would be largely characterised by overall differences in children's capabilities across the domains of development (i.e., above, below, or close to the average in capabilities), rather than a large number of profiles with unique or distinct patterns across the domains. (e.g., subgroups specified by specific capabilities).

Hypotheses in Phase 2 related to potential differences in the optimal solution by observed grouping variables (i.e., gender and ethnicity). Hypotheses regarding MI for indicators of capabilities (i.e., prosocial behaviours, emotional symptoms, emotional knowledge, and expressive language) were not specified in this paragraph but were addressed in individual appendices. Specifically, hypotheses in Phase 2 are: (1) the child's gender (boys and girls) will have no direct effects on the latent profile indicators (no DIF, meaning that as a form of measurement non-invariance, DIF is not observed); in other words, MI of the latent profile factors in the person-centred model across the child's gender groups were hypothesised, and (2) the mother-prioritised child's ethnicity (Māori, Pacific, Asian, European, and Other ethnicities) will have no direct effect on the latent profile indicators; that is to say, MI of the latent profile factors in the person-centred model across the child's ethnic groups were also hypothesised.

4.4 Sample

The current research drew on data from the preschool phase (DCW5) of the

GUINZs cohort study. Data collection of this wave took place in 2013 during the child's 54-month follow-up interviews and included 6,156 children and their parents. The determination of inclusion and exclusion criteria of the sample for the current study was guided by the research aims and analytic requirements of each phase.

Initially, participants with no data on a given indicator were excluded from CFA analyses and construction of factor scores from the corresponding indicator in Phase 1, and participants with dual ethnic identities were additionally excluded from MG-CFA in Phase 2 (details about this procedural decision are available in **4.6.1.2** and **Appendix 2**). Data were available for 6,156 children on any of the selected measures and were included in the analysis to validate the factor structure and descriptive analysis of the measures. For the LPA in Phase 1 and MIMIC modelling in Phase 2, participants were included if they provided data from at least one mother interview task and one child observation task (i.e., in the 54-month dataset to mitigate potential biases associated with non-attendance at mother interviews or children observations). Notably, participants may have been excluded due to missing gender information or dual ethnic identities in the MIMIC modelling in Phase 2. Overall, measures for 5,707 children were included in the LPA and MIMIC modelling analyses.

4.5 Ethical Considerations

Ethical approvals were obtained for both conducting the GUINZs and for the current research. Approval from the GUINZs was initially granted by the Ministry of Health Northern Y Regional Ethics Committee (NTY/08/06/055), and mothers provided

written informed consent for their own participation as well as their children. For the thesis, an additional full-risk Massey University ethics application was submitted and was approved (NOR 22/57) by the Massey University Human Ethics Committee, as attached in **Appendix 1**. Although the current research represents a secondary data analysis in nature, the full-risk application was appropriate due to the sensitive nature of the data, which included information about young children, and a focus on outcomes and equity in measurement for younger Māori, the indigenous peoples of New Zealand. Ethical considerations guided all aspects of the research. The key ethical considerations for the secondary data analysis in the current work are related to data security, avoidance of harm and ensuring social justice, particularly in relation to the analysis and interpretation of findings. Each of these is described in more detail.

Data security was ensured through the University of Auckland and Massey University systems. Research data management plans about data reuse, storage, and analysis were guided through the GUiNZs workshops and Massey University library training sessions. Data and information were additionally used, shared, and stored with care, ensuring adherence to the GUiNZs Data Access Agreement (V1.0 FM505, October 2014) and the principles¹ of *Te Ara Tika* (Hudson et al., 2019).

Avoidance of harm and ensuring social justice were addressed through data due

¹ The principles of *Te Ara Tika*—whakapapa (purpose), tika (research design), manaakitanga (cultural and social responsibility), and mana (justice and equity)—serve as Māori ethical guidelines. These principles emphasise respectful and active engagement, cultural integrity, and the well-being of individuals and communities, guiding researchers and ethics committees to conduct research that upholds Māori values and promotes positive outcomes for Māori communities.

diligence, appropriate partnership, and prior consultation for the research design. For instance, to determine whether the equitable findings of LPA applied to Māori and ensure appropriate results interpretations for Māori, tests of MI assessed the equivalence of latent profile constructs across observed subgroups. Checking MI by ethnicity required a strategy to classify ethnicity, including classifying children with multiple ethnicities. The strategy selected clearly identified limitations and appropriate cautions to ensure interpretation accuracy. Cultural consultation for analytic procedures and interpretations of findings occurred throughout the research.

Furthermore, working as an international researcher required appropriate consultation and support for understanding the New Zealand context and data interpretation, including consideration of terminology, Indigenous views of development, and alternate interpretations of findings to minimise harm, support the benefits of the research and address social justice issues. For example, focusing on strengths-based interpretations and preventive implications consistent with the developmental period for young children over deficit-based interpretations and implications can help mitigate risks of harm, increase benefits, and address social justice issues.

4.6 Selected Variables

Data for the present study can be classified as observed grouping variables, including the child's gender and ethnicity and data about young children's capabilities

(prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function).

4.6.1 Observed Grouping Variables

The child's gender and ethnicity were chosen as observed grouping variables for the Phase 2 study due to their substantial impact on young children's emerging identities and their relevance to education as emphasised in *Te Whāriki*, the current early childhood curriculum.

4.6.1.1 Gender.

In the current work, gender was constructed based on the children's biological sex at birth using the binary categories of 'boys' and 'girls'. Data for the child's gender was derived from the maternal responses during the 2009/2010 GUiNZs data collection wave, where mothers were asked to indicate whether they had given birth to a boy or a girl. These were the only two response options. The use of binary gender is a practical approach that allows for the investigation of differences in development by gender in young children (Hyde et al., 2019). A major limitation of this approach is that it does not align with contemporary understandings of gender diversity and fluidity and fails to encompass commonly recognised options, such as intersex and non-binary. As children grow in educational settings, the range of gender identities they identify with may expand further (Martin & Ruble, 2004). For example, recent research from GUiNZs to explore gender diversity at age 12 years showed self-reported gender identity and expression differed from the designations provided at birth for some of the cohort

(Neumann et al., 2023). Thus, acknowledging and discussing these limitations is essential for advancing understanding and exploring potential approaches in this field.

4.6.1.2 Ethnicity.

In the current work, a child's main ethnicity or ethnicities was determined by the mother's report of the child's ethnicity, which allowed the mother to nominate up to two options for the child's main ethnic identity. In the current work, mothers' responses were then classified into mutually exclusive single/combined ethnicity groups. Where mothers identified a single main ethnicity for the child, single-ethnicity groups were reported as described following the administrative prioritisation in Level 1 categories by Statistics New Zealand (2004): Māori, Pacific peoples, Asian, Other ethnicities (excluding European), and European. Where mothers identified the child as having combined/dual ethnicity groups (i.e., nominated two main ethnicities that fell within different Level 1 categories), these children were excluded from analyses due to low sample sizes where ethnicity was to be assessed as a grouping variable (i.e., tests of measurement invariance), and complexities of interpretation. A detailed report on ethnicity prioritisation is attached in '8.2 Appendix 2. Ethnicity Prioritisation'.

It is important to note that the breadth of the GUiNZs data regarding ethnicity categorisation offers multiple ways to classify, summarise and report ethnicity to address complexity (Yao et al., 2021). However, the quantitative analyses employed in the current research necessitated each case to be assigned to one mutually exclusive response option, leading to the use of the children's single/combined ethnicity, as

prioritised by their mothers (Ministry of Health, 2017; Statistics New Zealand, 2005). This approach has limitations, as it may not fully capture an individual's ethnic identity, which may comprise multiple ethnicities (i.e., more than two or non-‘main’ ethnicities), and the dynamic nature of ethnicity identification. In the current sample, this approach to categorisation additionally resulted in insufficient sample size within and between groups for children with combined/dual ethnic identities to be included in key Phase 1 and Phase 2 analyses, including LPA, and associated MIMIC models. Ongoing recognition and discussion of issues related to adequate sample sizes are necessary to advance the knowledge and potential methods.

4.6.1.3 Present Sample.

Demographics (gender and ethnicity) of cases included the full sample, and the reduced sample included in LPA and MIMIC modelling were presented in Table 4-2.

Table 4-2

Sociodemographic Characteristics of the Overall Sample and Reduced Sample Included in LPA and MIMIC Modelling (n)

Observed group	Categories	Overall sample (n = 6156)	Reduced sample included in LPA and MIMIC modelling(n =5707)	
Child Gender (6W)	Girl	2986	2791	
	Boy	3164	2916	
	Missing	<10	<10	
Mother-prioritised child ethnicity (54M) Single/combination*	Dual ethnicities	466	445	
	Single-ethnicity	Māori only	518	474
		Pacific Peoples only	729	673
		Asian only	670	594
		Other ethnicities only	661	559
		European only	3008	2876
Missing	104	91		

Note. Other ethnicities include MELAA, Other and New Zealanders. * represents missing values in the overall 6156 sample that include 104 system missing, user missing values and residual categories, which include 'Don's know' and 'References' options. 'n<10' signifies cell suppression, indicating the removal of a cell's value deemed sensitive, according to the GUINZs data output guidelines.

It is acknowledged that the approaches to classifying the child's gender and ethnic groups did not fully reflect the diverse gender identities or ethnically diverse populations living in contemporary New Zealand society. These limitations were discussed with associated analyses presented in Chapters 5 and 6.

4.6.2 Children's Capabilities

Key domains of young children's capabilities for the thesis were represented by age-appropriate indicators of prosocial behaviours, emotional symptoms, emotional knowledge, expressive and receptive language, and executive function. Indicators of the child's social, emotional, and language capabilities were completed by the child's mother. Specifically, mothers completed the Strengths and Difficulties Questionnaire (SDQ) and the five-item Parent Rating of Oral Language (PROL) measure. In addition, GUINZs' researchers conducted face-to-face assessments to observe and evaluate children's emotional knowledge, receptive language, and executive function skills. The face-to-face assessment included the Affect Knowledge Test (AKT), Peabody Picture Vocabulary Test-III (PPVT-III), and the adapted Luria Hand Clap task.

The use of mother-reported measures and direct child observation meant that information was collected from multiple perspectives to support a holistic and nuanced understanding of the child's social, emotional, language and executive function capabilities. Using multiple data sources—specifically, mother-reported measures and

direct child observations can help validate findings by cross-checking information, thereby reducing bias in the representation of children's capabilities, ensuring the inclusion of valuable data, and maximising the sample size and statistical power. Each of the measures and the scoring used in the present study is described below.

4.6.2.1 The Strengths and Difficulties Questionnaire (SDQ).

The Strengths and Difficulties Questionnaire (SDQ) has been widely used as a behavioural screening tool among children 4-17 years of age (Goodman & Goodman, 2009; Stone et al., 2010). Versions of the SDQ vary across age groups assessed, including parent/teacher-reported forms for children aged 4-17 and self-reported forms for children aged 11-17 (Goodman, 1997; Goodman et al., 1998). In the GUiNZs, the parent-reported standard version of SDQ was used in the 54-month data collection wave. The 25-item SDQ consists of five subscales, four assessing behavioural difficulties (i.e., conduct problems, hyperactivity, emotional symptoms, and peer problems) and one assessing behavioural strengths (prosocial behaviour), which relate to aspects of social and emotional capabilities as outlined by Denham (2006). Each SDQ item is endorsed by the respondent on a 3-point ordinal scale ('not true', 'somewhat true', 'certainly true'), scored 0-2, with five positively-worded items among the Difficulties subscales reverse-scored such that higher scores indicate more significant behavioural difficulties. Items of each of the five subscales are summed, with their total scores ranging from 0-10.

Due to an issue with timing² related to verifying reverse-scored items for the dataset provided of the three selected subscales of the SDQ (peer problems, conduct problems, and hyperactivity-inattention), the current study only included the 5-item prosocial behaviour and 5-item emotional symptoms subscales of the SDQ as indicators of children's social and emotional capabilities. The decision to proceed with these two subscales also aligned with the stated purpose of focusing on children's strengths and capabilities rather than deficits or difficulties where possible. The prosocial behaviours subscale focused on children's prosocial behaviours such as kindness, helping behaviours and being considerate of other people's feelings. While the emotional symptoms subscale focused on emotional difficulties, such as being worried, anxious, having many fears, or being easily scared, it was determined that this aspect of children's social-emotional functioning, as reported by mothers, should be included.

Prior to Phase 1 LPA analyses, the construct validity of the prosocial behaviour and emotional symptoms subscales was assessed using Confirmatory Factor Analysis (CFA) to confirm their factor structure, with measurement equivalence across observed gender and ethnic groups assessed using Multi-Group CFA. Results are reported in

² The five SDQ subscales were initially included in the study, and confirmatory factor analyses (CFA) were conducted to assess their validity. Following the 8Y Data Users Guide, four items in the peer problems and hyperactivity-inattention subscales were reverse-coded, and one conduct problems item was noted missing. Examination of the dataset indicated that these four reverse-coded items all appeared to have values, raising concerns that missing data in the dataset provided may have been recoded as actual values prior to the current analyses. Clarification from the GUIiNZs team could not be obtained between April and November 2023. After consultation with supervisors, to ensure analytical reliability, analyses excluded the three subscales containing reverse-scored items and proceeded with the two subscales that could be independently and reliably interpreted.

Chapter 5, and a complete report on the SDQ and assessment of construct validity in the current sample is presented in 8.3 Appendix 3.

4.6.2.2 The Affect Knowledge Test (AKT).

The Affect Knowledge Test (AKT) was designed to assess preschoolers' expressive and receptive emotional knowledge (Denham, 1986) important for young children's social and emotional capabilities. The AKT has been widely used as an age-appropriate tool to assess 2.5- to 5-year-old children's recognition of emotions (i.e., happy, sad, angry, and afraid) and understanding of situational emotion knowledge (Bassett et al., 2012; Sette et al., 2015). Versions of the AKT include the full (AKT) and shortened (AKT-S) versions (Denham, 1986; Denham et al., 2015). The AKT-S comprises a 4-item expressive/verbal emotion identification task in which children are asked to recognise and name facial expressions depicted on four puppets (Denham, 2006). This version was further adapted to the 6-item AKT in the GUiNZs at the 54-month data collection wave (Morton et al., 2017).

The current research used data from the 6-item adapted AKT administered by the GUiNZs to assess young children's emotional understanding and expression capabilities. The GUiNZs adopted the shortened version of the AKT expressive emotion identification task due to time constraints. Cards with simple cartoon faces were used in place of puppets showing different emotions, and eyebrows were removed from the faces to make them more gender-neutral. Faces expressing the emotions of disgust and surprise were added to the original four emotions (happy, sad, angry, scared) in

the AKT expressive identification task to avoid the previously observed ceiling effect among children aged 54 months (Denham, 2006). Response options included: “0 = incorrect emotion with the opposite emotional valence; 1 = incorrect emotion within the same emotional valence; 2 = correct emotion or acceptable synonym”. The adapted AKT was assessed using a confirmatory factor analysis (CFA) to confirm its factor structure, and multiple-group confirmatory factor analysis (MG-CFA) to assess measurement invariance across the child’s gender and ethnic groups. Results are reported in Chapter 5, and a complete report on the AKT and assessment of construct validity in the current sample is presented in 8.4 Appendix 4.

4.6.2.3 The Parent Rating of Oral Language (PROL).

The current work used the Parent Rating of Oral Language (PROL) to indicate children’s expressive language skills. The GUiNZs modified the PROL from the Teacher Rating of Oral Language and Literacy (TROLL) (Dickinson & McCabe, 2001; Dickinson et al., 2003). The five questions of the PROL asked parents to indicate the child’s pattern of asking questions, the frequency of trying out new words, the child’s ability to communicate personal experiences clearly and logically, the frequency of being understood when speaking to other adults, and the child’s ability to communicate when the child is not first understood. Each item was rated on a four-point ordinal scale, with higher scores reflecting greater expressive language capabilities.

Prior to Phase 1 LPA analyses, the construct validity of the PROL was assessed using confirmatory factor analysis (CFA) to confirm its factor structure, and multiple-group

confirmatory factor analysis (MG-CFA) to assess measurement invariance across child gender and ethnic groups. Results are reported in Chapter 5, and a complete report on the PROL and assessment of construct validity in the current sample is presented in 8.5 Appendix 5.

4.6.2.4 The Shortened Peabody Picture Vocabulary Test-III (PPVT-III).

The revised, shortened version of the Peabody Picture Vocabulary Test III (PPVT-III) aims to test receptive vocabulary, verbal communication, and comprehension, with scores ranging from 0-40 (Ly et al., 2020; Rothman, 2005). During the assessment, the interviewer verbally presents a word, and the child selects the image that best represents its meaning or indicates “don't know”(Dunn & Dunn, 1997). The task starts with two training items with four distinct images displayed, respectively. If both are answered correctly, the test of a Core Set of 20 items proceeds. Depending on the number of errors in the Core Set, children may receive an additional Basal Set if many errors occur or a Ceiling Set if few errors occur (Ly et al., 2020). The test concludes if the child chooses to stop or when their performance indicates no further items are needed. The scoring is based on the total number of correct responses on vocabulary, such as wrapping, exercising and horrified.

The current research used the PPVT-III to assess children's receptive language skills in English, consistent with the prior GUiNZs research (Galvin et al., 2020; Neumann et al., 2019). Data for the PPVT-III were available as a single variable derived from the 3-parameter IRT model analysis conducted by the GUiNZs (Ly et al., 2020). Item-level

data were not available to secondary data users; therefore, the current research was unable to test its factor structure and assess MI across the child's gender and ethnic groups. A detailed discussion of the scoring methods, validity as a measure of receptive vocabulary in young children and cultural biases of the original PPVT-III, the administration and scoring of the adapted PPVT-III in the GUiNZs and its use in the current research is provided in 8.6 Appendix 6.

4.6.2.5 The Luria Hand Clap Task.

The Luria Hand Clap task, adapted from the Luria-Nebraska Neuropsychological battery (Bialystok et al., 2010; Golden et al., 1980; Plaisted et al., 1983), aims to test children's executive function skills (Buckley et al., 2019; Diamond & Taylor, 1996). The task measures young children's inhibitory control and other executive function skills (i.e., working memory, cognitive flexibility) by testing the child's ability to focus on the number of claps that the interviewer clapped, and the child responded by doing the opposite number of claps, indicating the executive component. The number of correct responses across sixteen test tasks provides a score ranging from 0 to 16. In GUiNZs, Galvin et al. (2020) examined 54-month-old children's executive function using a hand-clap task. Due to the nature of its discrete, skewed distributional properties and non-negative values, scores are treated as a count distribution and represented as medians in the current research. Additionally, the current research could not confirm its factor structure and test MI across the child's gender and ethnic groups. A detailed description of the distributions of the Hand Clap Task scores, both in the overall sample

and by gender and ethnic groups, along with the use of the scores, is provided in 8.7 Appendix 7.

4.7 Procedure

The following section outlines details for conducting the present study, which represents a secondary analysis of data obtained from child observations and mother interviews collected at the GUiNZs 54-month data collection wave. This procedure includes information about data access, project consultation, an overview of the analytic process, each phase described in detail, and the treatment of missing data.

4.7.1 Data Access

As the nature of the secondary data analyses in the current research, access to these existing datasets in the GUiNZs was obtained (reference number: 22DA001477) via the established process known as the GUiNZs dataset internal access application. This process involved the preparation of an initial research proposal (including the justification for the research, research design, variable selection, analysis plan, cultural consultation, and ethical considerations) for peer review/consultation by the Centre for Longitudinal Research (C4LR). Informed by this process, a final proposal was developed, reviewed, and approved by the GUiNZs Data Access Committee (DAC), and evaluated at the Doctoral Confirmation event at Massey University. Throughout my research, I have engaged in data due diligence and appropriate partnership and consultation for the research design, analysis, and interpretation.

4.7.2 Project Consultation

The project consultation process has occurred in several phases. Initially, considerations and discussions around research design, data analysis and results interpretation were discussed with supervisors during fortnightly supervision meetings. Then, research design, data access application, and ethical considerations were discussed in consultation with Associate Professor Elizabeth Peterson and Associate Professor Annette Henderson from the University of Auckland. Ethical issues around ethnicity identification and the appropriate use of ethnicity variables were specifically discussed in a cultural consultation with Professor Te Kani Kingi and Dr Pania Te Maro. Thereafter, questions around the conceptual framework and the appropriate use of ethnicity variables were discussed with researchers from the Centre for Longitudinal Research, University of Auckland. Suggestions were also given at the doctoral confirmation event at Massey University. Ongoing fortnightly supervision meetings have also been used to review and discuss the results and findings. Finally, findings and thesis chapters were submitted to the GUINZs for review and feedback by the research team and Data Access Committee, consistent with GUINZs data use procedures, before the submission of the thesis.

4.7.3 Analytical Process

As noted above, the research was conducted in two phases: an initial phase assessing construct validity and an exploration of profiles of capabilities that may be evident in the overall sample of children at the 54-month assessment, and a second

phase in which the generalisability and validity of this model across child gender and ethnic groups were assessed. Figure 4-1 outlines these phases in more detail. The following section describes the methodological procedures used to address the research questions examined in each phase. Table 4-3 provides a more detailed overview of the analytical process used. Table 4-4 outlines the measures used and scores applied in each step by measures. Notably, CFA and MG-CFA were conducted only when constructs were indicated by indicators (prosocial behaviours, emotional symptoms, emotional knowledge, and expressive language), and item response data of these measures were available for secondary analyses in the current research. Together, both tables provide contexts that guide the methodological process in both phases of the research, key decisions on the analytical techniques, data sources and scores used in the analyses, and an overview of measures and scores by measures that aligned with the two phases of the studies in the thesis. The data analyses in both phases were carried out using SPSS Version 28.0.1.1 for data file preparation and Mplus version 8.10 for the subsequent analyses. All procedures and analyses were conducted under the GUINZs data use agreement and consulted with the project advisors and research team from GUINZs.

Figure 4-1

Methodological Process

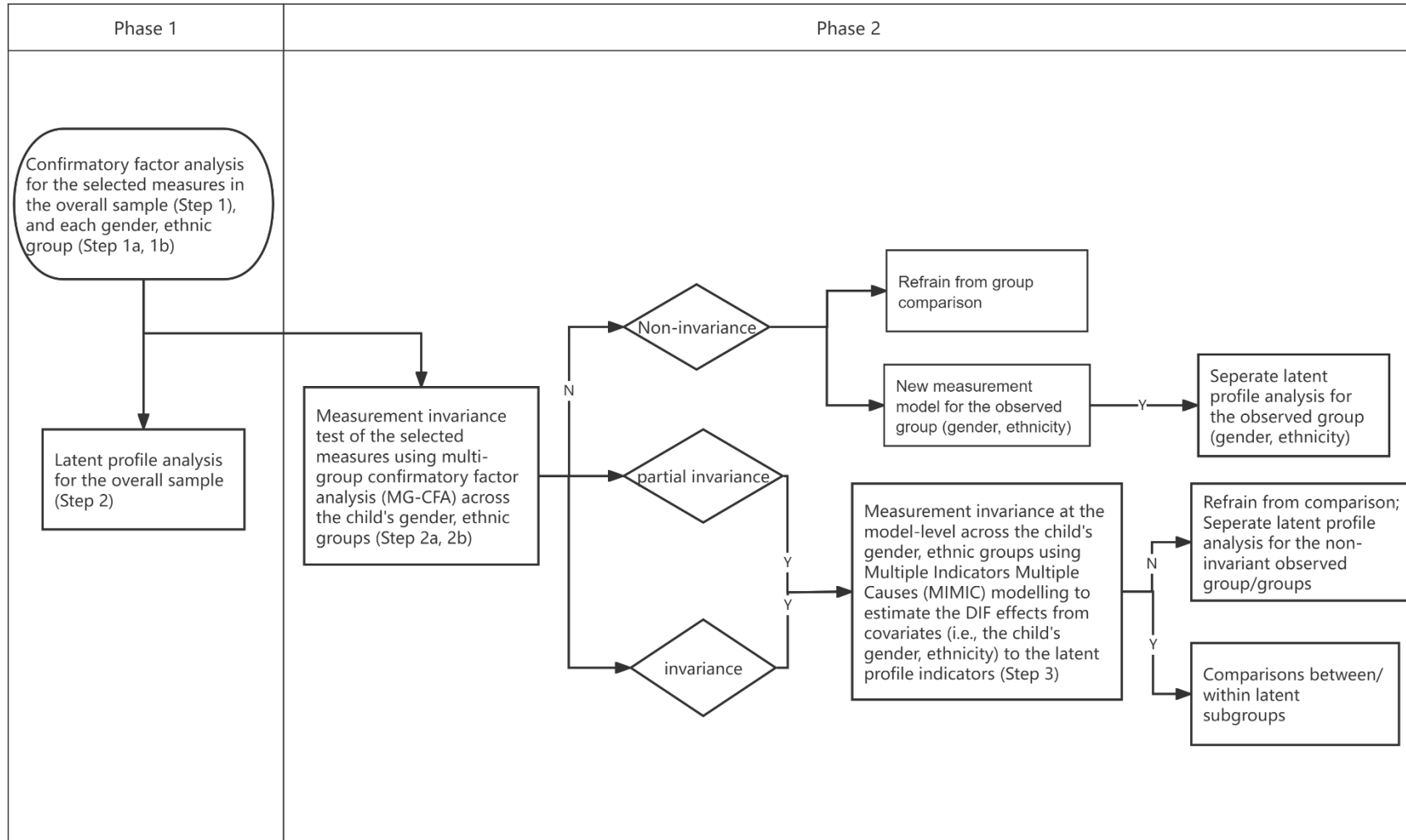


Table 4-3

Overview of Analytical Process

Phases	Phase 1		Phase 2				
	Step 1	Step 2	Step 1a	Step 2a	Step 1b	Step 2b	Step 3
Analysis techniques	CFA	Latent Profile Analysis (unconditional LPA models)	Separate CFA for each gender group	MI of the selected measures across gender groups using MG-CFA	Separate CFA for each ethnic group	MI of the selected measures across ethnic groups using MG-CFA	MI of latent profile factors was tested using the DIF analysis through Multiple Indicators Multiple Causes (MIMIC) modelling: adding gender and ethnic groups as covariates of the latent profile factors
Data source	All data of the overall sample	Reduced sample	Data for each gender group	All data of the overall sample	Data for each ethnic group	All data of the overall sample	Reduced sample
Scores for analyses	Item scores of the available measures	Factor scores generated in step 1, and factor scores and count scores generated by the GUiNZs team, where item scores were not available	Item scores of the available measures by each gender group	Factor scores generated in step 1 and factor and count scores generated by the GUiNZs team, MG-CFA by gender	Item scores of the available measures, CFA by Ethnicity	Factor scores generated in step 1 and factor and count scores generated by the GUiNZs team, MG-CFA by ethnicity	Factor scores generated in step 1, and dummy-coded scores of the child's gender and ethnicity

Table 4-4

Overview of Measures and Scores Used by Measures in Each Step

Phases		Phase 1		Phase 2					
Steps		Step 1	Step 2	Step 1a	Step 2a	Step 1b	Step 2b	Step 3a	Step 3b
Model indicators	Selected measures	Factor of the indicators in CFA	Score entered in LCA	Factor of the indicators in CFA	MG-CFA Results	Factor of the indicators in CFA	MG-CFA Results	Scores in step 3a	Covariates and factors in MIMIC modelling
Prosocial behaviour	SDQ prosocial (5 items)	1 factor	Factor scores	1 factor	MI	1 factor	Partial MI	MG-CFA across gender threshold invariance test factor scores	Using the observed grouping variables (gender, ethnicity) as covariates of the optimal latent group solution and its factors generated in Phase 1, Step 2
Emotional symptoms	SDQ emotional symptom (5 items)	1 factor	Factor scores	1 factor	MI	1 factor	Partial MI	MG-CFA across gender threshold invariance test factor scores	
Emotional knowledge	AKT (6 items)	1 factor	Factor scores	1 factor	MI	1 factor	Partial MI	MG-CFA across gender threshold invariance test factor scores	
Expressive language	PROL (5 items)	1 factor	Factor scores	1 factor	MI	1 factor	Partial MI	MG-CFA across gender threshold invariance test factor scores	
Receptive language	PPVT- III	n/a	GUINZs calculated score	n/a	n/a	n/a	n/a	GUINZs calculated score	
Executive function	Hand Clap Task	n/a	Count Score	n/a	n/a	n/a	n/a	Count Score	

4.7.3.1 Phase 1: Confirmatory Factor Analysis.

As noted in Phase 1 Figure 4-1, the research initially focused on validating the selected indicators of children's capabilities through Confirmatory Factor Analysis (CFA: identified as step 1 in Table 4-3 and Table 4-4). Phase 1, step 1, involved using item scores of each of the available measures to conduct separate confirmatory factor analyses for the overall sample. This process helped confirm the factor structure of the selected measures (5-item prosocial behaviour and emotional symptoms subscales in the SDQ, the 6-item AKT, and the 5-item PROL). This step also helped ensure indicators that reflect young children's social, emotional, and language capabilities were theoretically sound and appropriate for the sample, prior to examining the LPA of the overall sample.

Phase 1, step 1 confirmed one-factor structure and generated factor scores in the CFA for each of the available measures, which were further used in the subsequent LPA. The advantages of using factor scores are to provide a more robust alternative to scale scores by giving more weight to items with lower measurement errors, thus offering partial implicit control for measurement errors (Morin et al., 2016; Morin & Marsh, 2015). Detailed reports for the CFA results for the available measures are described in **Appendices 3-5**. Notably, the present research was unable to conduct CFA for the PPVT-III due to data unavailability, nor for the Luria Hand Clap tasks, as it was represented by a single count indicator in the GUiNZs 54-month dataset.

4.7.3.2 Phase 1: Latent Profile Analysis.

As discussed in Chapter 3 and noted in Phase 1 Figure 4-1, following the examination of the construct validity of the available measures, a holistic and person-centred analysis approach was adopted to explore the potential for subgroups of young children to display distinct shared patterns of social, emotional, language, and executive function capabilities. In Latent Profile Analysis (LPA), individuals are classified into high probabilistic classes, with low probabilities of being assigned to other latent subgroups (Harring & Hodis, 2016; Jago et al., 2018). This enables substantive latent subgroups of young children's social, emotional, language and executive function capabilities to be explored and described.

In Phase 1, step 2 (Table 4-3 and Table 4-4 by measures), indicator variables in the LPA were the factor scores generated in CFA for each of the available measures in the current research (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language). In addition to these factor scores, the present research used factor scores of the PPVT-III and count scores in the Luria Hand-Clap task generated by the GUiNZs team to represent young children's receptive language and executive function, respectively.

For the present research, factor scores representing prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language can be interpreted as standardised mean scores. This means that indicators of profiles in latent group solutions are broadly described in relation to the standard

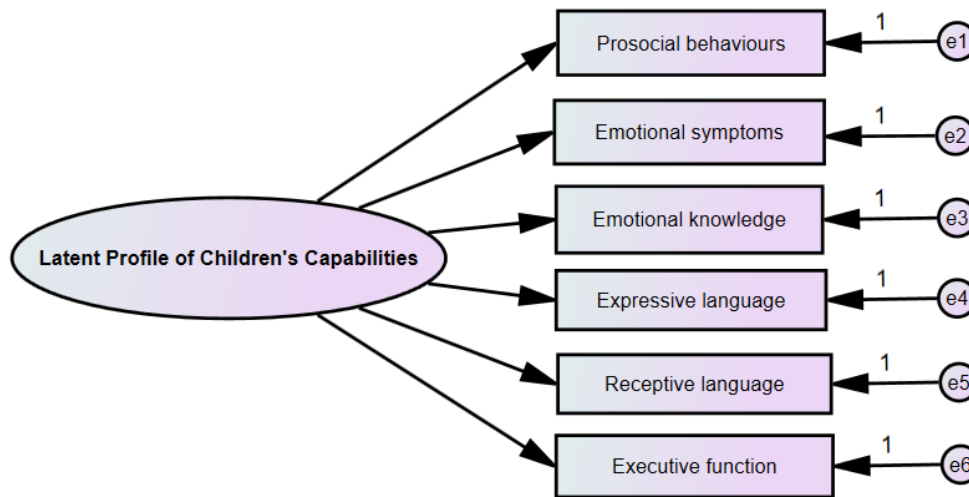
deviation below or above the overall sample mean. Median scores of the Luria Hand Clap task are reported with the grand median score of 13 used as the reference on the X-axis in the illustrations of latent group models in the Results section of Chapter 5. The use of grand mean/median scores was employed as a reference to reflect young children's overall social, emotional, language and executive function capabilities within the sample, rather than indicating a pre-determined or expected level of performance on each measure. Scores below or above the mean should not be interpreted as having clinical implications.

The conceptual model of the LPA, illustrated in Figure 4-2 and presented again in Chapter 5, is designed to explore profiles of young children's capabilities. Using predefined criteria, these profiles are indicated through associations between children's capabilities. Composite scores (i.e., factor scores) were used to indicate prosocial behaviours, emotional symptoms, emotional knowledge, expressive language and receptive language, and count scores were used to assess executive function. The latent profile refers to the unobserved latent variables, which are represented as estimated group solutions in Chapter 5, the 'Results' section within the current research.

Unconditional LPA models were estimated for group solutions ranging from one to six groups, with variances fully constrained and unconstrained for examining model stability. In variances fully constrained models, variances were fixed to be equal across group solutions in a model, whereas in variance-unconstrained models, variances were freely estimated across group solutions.

Figure 4-2

Visualisation of the LPA Conceptual Model in Phase 1



Note. The latent profile of children's capabilities represents the unobserved latent variable in the ellipse. The square boxes represent indicators or measured variables that include prosocial behaviour, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function. e1 to e6 represent the error variances in the model.

The optimal group enumeration process was carried out alongside the estimation of successive LPA models, with both model fit indices and theoretical considerations informing the optimal solution. Specifically, model fit indices used to determine the number of profiles that fit the data were information criteria, including Sample-size-adjusted Bayesian Information Criterion (SABIC), Lo-Mendell Rubin Likelihood Ratio Test (LMR), Bootstrap Likelihood Ratio Test (BLRT), entropy, as well as the group sizes and interpretability of the emergent subgroups (Muthén, 2004; Nylund et al., 2007). A detailed LPA description of the LPA methods and findings is provided in Chapter 5.

4.7.3.3 Phase 2: Multiple-group Confirmatory Factor Analysis.

In Phase 2, Figure 4-1, MI across the observed gender and ethnic groups were evaluated for prosocial behaviours, emotional symptoms, emotional knowledge, and expressive language. The analyses evaluate the construct validity of these measures across observed grouping variables (i.e., children's gender and ethnic groups), commonly used as socio-demographic variables in GUiNZs (Meissel et al., 2019; Morton et al., 2020). When the parameters of a measurement model are statistically equivalent across the child's gender and ethnic groups, MI is established, meaning the construct of the selected measures is understood in the same ways across the child's gender and ethnic groups. This was explored by testing the hypothesised CFA model separately in each gender and ethnic group using CFA, followed by further analyses of testing a multiple-group model using MG-CFA to see whether MI across child gender and ethnic groups were established.

As recommended by Dimitrov (2010) and Bowen and Masa (2015), CFA was conducted separately to assess the best factor model for each group before the invariance tests prior to the MG-CFA. This process is shown as Phase 2 in Table 4-3 and Table 4-4 for assessing CFA models by the child's gender (step 1a) and ethnicity (step 1b). Multi-group analyses of MI for each of the CFA models of the selected measures were conducted using ordered categorical variable methodology and weighted least squares means-variance (WLSMV) estimation and the Theta parameterisation as implemented in Mplus v8.10. In a CFA model, each item is assumed to indicate the

same generic true score, but with a unique contribution to the composite score and an item-specific error variance (Dragovic, 2004).

MG-CFA was conducted to test the MI of the selected measures across child gender and ethnic groups, as seen in Steps 2a and 2b in Table 4-3 and Table 4-4. Due to the limited data access, the available selected measures included in the MG-CFA are the prosocial behaviour and emotional symptoms subscale in the SDQ, the AKT, and the PROL. Specifically, the group invariance of these one-factor models of the selected measures across the child's gender and ethnic groups was tested by assessing the relative fit of nested models with increasingly restrictive constraints on the factor structure, factor loadings and item thresholds. Decrements in model fit were assessed with reference to changes (Δ) in fit indices with increasingly restrictive models, with critical values of Δ CFI [0.01], SRMR ($\Delta < 0.02$ for configural vs. metric; $\Delta < 0.01$ metric vs. threshold), and RMSEA ($\Delta < 0.015$) used to evaluate whether a null hypothesis of group invariance should be rejected (Chen, 2007). Due to its sensitivity to sample size, changes in chi-square for increasingly restricted models assessed using the DIFFTEST function were reported but not considered in the current evaluation of model fit. Detailed reports for MI tests were described in Appendices 3-5.

Testing MI of the selected measures in the current research included configural, metric and threshold invariance. Configural invariance was conducted by specifying the factor structure to be consistent across groups, while the factor loadings and thresholds were freely estimated across groups. The metric invariance was assessed by

constraining all factor loadings to be equal across groups, fixing residual variances at one in the reference group while freeing them in the other groups, and setting factor means at zero in the reference group while estimating them freely in the remaining groups. The threshold invariance was conducted by constraining factor loadings and thresholds to be equal across groups.

Where model evaluation indicates a lack of metric or threshold invariance, partial invariance analysis was performed. A backward estimation approach freed parameters displaying the greatest model miss-fit across groups. Partial invariance was identified if acceptable levels of decline in model fit could be achieved, where more than half of the model indicators were invariant across groups.

Results of the MI tests across the child gender and ethnic groups are shown in steps 2a and 2b in Table 4-4. Full MI was found for the selected measures across the child gender groups, while partial invariance was found for the selected measures across the child ethnic groups. These results are used to guide the following MIMIC modelling across the child gender and ethnic groups.

4.7.3.4 Phase 2: DIF Analyses of the Optimal Latent Group Solution Factors.

A notable aspect of the current research was the exploration of MI within latent profile solutions in person-centred approaches. The consistency of profiles derived from the analysis was assessed across the child's gender and ethnic groups to ensure appropriate interpretation of findings. Initially, this was assessed by testing the hypothesised LPA models separately in each gender group using LPA in Step 3a and

Step 3b in Table 4-3, followed by further analyses of estimating potential biases from the covariates (i.e., the child gender, ethnicity) to the latent profile factors using MIMIC modelling in Table 4-3 to see whether MI of the latent profile factors at the model level would be established.

Once MI tests of the available measures were completed, the current research proceeded with data analysis through Step 3 DIF approach using MIMIC modelling as it enabled to examine MI at the construct level of the two selected indicators (i.e., receptive language measured by the PPVT-III, executive function measured by the Hand Clap task) across the child's gender, and five single-ethnicity groups. Specifically, as shown in Table 4-3 Step 3, the DIF approach was used to assess whether the optimal model construct was equivalent across the child's gender and ethnic groups. The DIF approach through MIMIC modelling uses child's gender, and five single-ethnicity groups as covariates of the optimal latent group solution and its factors to extend the examination of MI on the profiles and assesses the direct effect on each of the factors (i.e., prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function) in the optimal latent group solutions. A significant direct effect of the covariates on a factor suggests that the factor mean differs across the levels of the covariate. Where direct effects are not significant, it can be inferred that the model is equally valid levels of the covariate and that group differences can be meaningfully interpreted (Harring & Hodis, 2016).

The results from MI at the model level determine whether comparisons of the mean and probabilities of the optimal latent group solutions across the child gender and ethnic groups are proceeded or refrained, as noted in Figure 4-1. This exploration is significant as it helps mitigate biased estimations in LPA, where the measurement of non-invariance of latent profiles exists (Cole, 2017). Further, assessing the MI of the latent profile solution itself across critical demographic groups allows for the evaluation of whether these groups shared similar theoretical concepts or typologies (Bettencourt et al., 2022; Lukac et al., 2019).

4.7.3.5 Missing data.

Of the 6,156 participants who had data on the demographic variables, some had incomplete data on the selected measures (indicators) of young children's social, emotional, language and executive function capabilities. As multiple conceptually and statistically distinct stages were conducted in the current research, entries for participants were considered according to analysis purposes in each stage. In CFA in Phase 1 and MG-CFA in Phase 2, the numbers of participants were included to confirm the constructs of selected measures where there was data in each measure. In the LPA in Phase 1 and MIMIC modelling in Phase 2 to explore shared patterns of young children's social, emotional, language and executive function capabilities, cases of missing data on all indicators and cases were excluded (7.3%) and treated as missing completely at random where there was a missing observation for at least one of the

following tasks: Luria (executive function), PPVT-III (receptive language) or AKT (emotional knowledge).

4.8 Summary

Secondary analyses of the Growing Up in New Zealand study (GUiNZs) 54-month dataset were conducted in the current research. The primary focus of the current research was to explore the shared patterns of young children's social, emotional, language, and executive function capabilities. In addition, MI tests of the available selected measures were tested to ensure the constructs of these measures were understood equivalently across the child's gender and ethnic groups, which lays the foundations for the further examination of the MI on the profile model level. Research questions and hypotheses guided the specific analyses. In the current research, statistical techniques include descriptive analysis of demographic information and each measure, CFA, LPA, MG-CFA and the DIF approach using MIMIC modelling.

Each technique in the analytical procedures was optimised to align with the research aims and questions. Phase 1, Step 1, item scores from each of the available measures were used to conduct separate CFA for the overall sample, confirming the factor structure of the selected measures: the 5-item prosocial behaviour and emotional symptoms subscales in the SDQ, the 6-item AKT, and the 5-item PROL. In Step 2, a person-centred approach using the LPA was adopted to identify shared patterns of young children's social, emotional, behavioural, and executive function. Phase 2 evaluated whether the MI of the optimal latent group solution generated in

Phase 1 Step 2 can be established across the child's gender and ethnic groups to facilitate generalisation and comparisons of findings. In Phase 2, Step 3, the DIF approach using MIMIC modelling assessed whether the optimal latent group solution was equivalent across the levels of covariates (i.e., the child's gender and five single-ethnicity groups). The child's gender and five single-ethnicity groups were used as covariates of the optimal latent group solution and its factors to estimate the DIF effects from the covariates to the factors. This extended the examination of MI in the profiles and assessment of the DIF effects on each indicator to ensure fairness in the interpretation of the results.

Chapter 5 Phase 1: Methods and Findings

The purpose of the present chapter is to address the first phase of my research. This includes a more detailed description of the methods, describing the findings, and briefly discussing the Phase 1 implications and limitations. Phase 1 was to explore whether distinct latent subgroups with shared patterns of children's social, emotional, language and executive function capabilities could be identified in the 54-month dataset in GUiNZs. These capabilities were specifically measured by prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function. Based on multiple statistical criteria and theoretical justifications, a latent profile solution was identified, and the dominant patterns of children's social, emotional, language and executive function capabilities were interpreted in each latent subgroup.

5.1 Introduction

Young children's social, emotional, language, and executive function are often not directly observable and are conceptualised as latent variables, which cannot be directly observed but are inferred through mathematical models based on related observable variables (Bollen, 2002). Methods incorporating categorical latent variables are commonly known as finite mixture modelling (Harring & Hodis, 2016). Finite mixture modelling aims to identify subgroups of cases, also referred to as latent classes or profiles, based on distinct statistical characteristics, such as means, variances and covariances among variables of interest (Harring & Hodis, 2016). Mixture models assess

and classify cases in a probabilistic way. This means that the probability of an individual belonging to each latent class in a given model is computed, and then the individual is assigned to that class for which they have the highest probability of membership.

Latent class analysis (LCA) and latent profile analysis (LPA) fall under finite mixture modelling (FMM) techniques and are used as a person-centred approach to latent variable analysis. LCA refers to the probabilistic modelling technique to identify subgroups within the population using multivariate categorical data (Collins & Lanza, 2010). In contrast, LPA assesses how well underlying latent profiles explain the variance among a set of observed continuous variables (Bergman & Magnusson, 1997; Collins & Lanza, 2010; Masyn, 2013).

LPA is a rigorous, statistical method that enables the investigation of qualitative differences—reflected in the patterns of associations across key factors—alongside quantitative differences, such as the numbers and proportions of subgroups within the population. This makes LPA a powerful tool for generating a nuanced understanding of the diverse characteristics of the populations. LPA identifies individual differences within an effect or phenomenon through patterns and posits that a small number of unique but meaningful patterns in associations between variables occur within a given sample (Bergman & Magnusson, 1997; Bergman et al., 2003; Sterba, 2013). Patterns that emerge within and across groups might be described as homogenous or heterogeneous. When the pattern is homogenous, it refers to an overall trend in how individual factors relate to each other across groups. When the pattern is

heterogeneous, it refers to unique patterns of how individual factors relate to each other across groups, such that the patterns of relationship across factors may be distinct for each group.

As noted in the person-centred analysis methods section in Chapter 3, LPA has been extensively applied to describe patterns in the development of young children's social, emotional, language, and executive function characteristics (Collie et al., 2019; Russell et al., 2022; Sandilos et al., 2019; Thomson et al., 2017; Thomson et al., 2019). LPA enables researchers to identify groups of individuals with shared characteristics, describe them based on their indicators, and explore factors associated with group membership. It is also undoubtedly valuable for a holistic understanding of young children's capabilities and common patterns in strengths and weaknesses across diverse populations, datasets, and indicators used in the research.

A latent profile analysis involves model assumptions, specification, estimation, evaluation, selection, and interpretation. Steps for estimating, evaluating, and selecting an LPA involve fitting and comparing models specifying varying numbers of groups to evaluate their relative fit to the data and substantive interpretation of solutions to identify an optimal model (Masyn, 2013). The current research does not discuss the mathematical foundations of the LPA, which can be found in the literature (Masyn, 2013; McCutcheon, 1987; McCutcheon, 2002; McLachlan & Peel, 2000).

5.2 Analysis

Latent profile analyses (LPA) were conducted using Mplus 8.10 to identify patterns in young children's social, emotional, language, and executive function capabilities in the GUiNZs 54-month dataset. As noted in Figure 5-1, indicators of children's profiles are composite scores (i.e., factor scores) representing the child's prosocial behaviours, emotional symptoms, emotional knowledge, expressive and receptive language ability, and count scores for executive function. After establishing the construct validity of the available measures, composite variables, such as factor scores, are used as indicators to characterise profiles of children's capabilities. The robust maximum likelihood (MLR) estimator was used to enhance the accuracy and reduce bias in parameter estimates (Muthén & Muthén, 1998-2017). Sets of random starting values were specified between 100 and 500 to locate the highest log-likelihood values as opposed to a local solution and minimise statistical inaccuracy or biasing parameter estimates (Jung & Wickrama, 2008; Wickrama et al., 2016). The initial identification of an optimal model is performed using unconditional latent profile analysis (i.e., without covariates included). Specifically, the analyses start with assessing variance-constrained unconditional LPA models in which indicator variances are constrained across groups, followed by assessing variance-unstrained unconditional LPA models that release parameter constraints in variances.

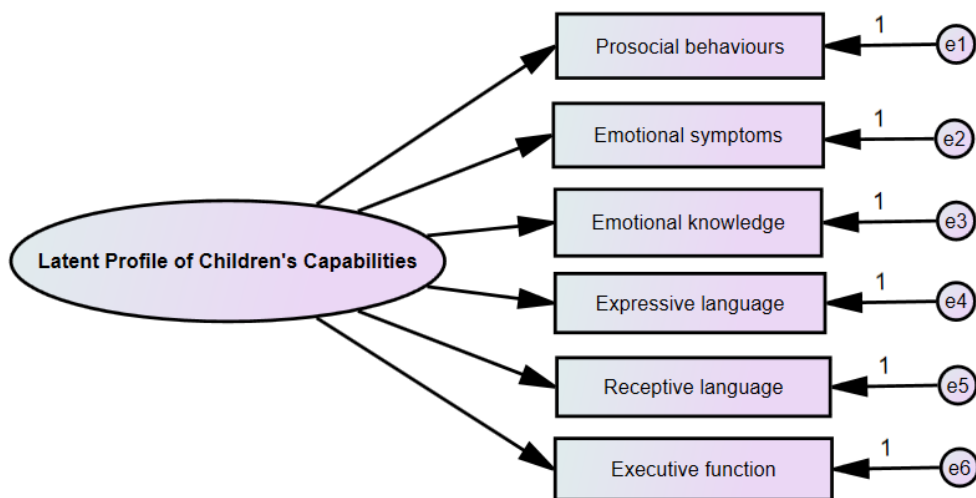
LPA models with increasing numbers of solutions were fitted to the data to identify the optimal number of latent profiles that best represent the shared patterns of

children's capabilities in the sample. A combination of data-driven and child developmental theory-based approaches was used to select the optimal LPA solutions. Specifically, the optimal number of groups was determined with reference to the following criteria: group size, group interpretability, entropy and average posterior probabilities of group separation, the Sample-size-adjusted Bayesian Information Criterion (SABIC), and significant Lo-Mendell Rubin Likelihood Ratio Test (LMR) and Bootstrap Likelihood Ratio Test (BLRT) statistics. A minimum group size, representing at least 50 cases or 5% of the sample, was required alongside the interpretability of the groups. Higher entropy values ($>.80$) were necessary to indicate greater class separation (and greater heterogeneity in the sample) and precision/certainty in group assignment (Muthén, 2004), with a value close to 1.0 signifying higher classification precision (Jung & Wickrama, 2008; Peugh & Fan, 2013). Model fit was evaluated by changes and plateaus in SABIC values with an increasing number of groups. SABIC values represent a statistical evaluation of the fit of a model while accounting for the number of parameters and the sample size, with lower values indicating a greater fit to the data (Henson et al., 2007). Improvements in model fit were compared between k and $k-1$ models using the SABIC values, and LMRT and BLRT statistics (Berlin et al., 2013; Lo et al., 2001; Muthén, 2004; Nylund et al., 2007). Once the models were identified, posterior probabilities were then used for tracking movement between the k vs. $k-1$ model, and a 0.7 was used as a minimum rule-of-thumb (Daniel, 2005). Bar charts based on estimated profile membership were presented to understand the

patterns of each model. Cross-tabulations of profile membership transitions between models based on the most likely profile membership were presented to interpret the changes in probabilities between the k vs. $k-1$ model.

Figure 5-1

Visualisation of the LPA Conceptual Model in Phase 1



Note. The latent profile of children's capabilities represents the unobserved latent variable in the ellipse. The square boxes represent indicators or measured variables that include prosocial behaviour, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function. e1 to e6 represent the error variances in the model. Figure 5-1 is the same as Figure 4-2.

5.3 Results

Results for this phase of research are presented in three parts. First descriptive statistics for the data are presented and described below. This is followed by a section to examine group enumeration for variance-constrained unconditional LPA models, followed by variance-unconstrained LPA models. Results are then presented for the

process to determine and interpret the optional solution identified for the present research.

A total of 5707 participants with data on either observation task (the Luria Hand Clap Task, the Peabody Picture Vocabulary Test- III, or the Affect Knowledge Task) were included in the current latent profile analysis. Descriptive statistics included means, variances, median, skewness and kurtosis, minimum and maximum values of the selected indicators in Table 5-1 and correlations between the selected indicators in Table 5-2.

Table 5-1 showed approximately symmetric distributions of prosocial behaviours, expressive language, and receptive language. In contrast, continuous variables fit within generally accepted values for approximate normal distributions regarding skewness and kurtosis. This further showed that moderately skewed distributions were observed for emotional symptoms and emotional knowledge represented by continuous variables and executive function represented by count variables. It is not surprising that these indicators were generally normally distributed across populations due to the factor scores generated as Z scores. As observed in Table 5-2, correlation coefficients between scores of the selected indicators showed that emotional symptoms were negatively correlated with other characteristics (prosocial behaviours, emotional knowledge, expressive language, receptive language, and executive function). Positive correlations were observed among these indicators except for emotional symptoms. These correlation coefficients, ranging from -0.21 to 0.44, showed zero to a moderate

negative/positive correlation (Akoglu, 2018; Dancey, 2007). These descriptive analyses provided a solid foundation for person-centred analysis.

Table 5-1

Descriptive Statistics for Scores on the Indicators

	n	Mean/variances	Median	Skewness/Kurtosis	Minimum/maximum
Prosocial behaviours	5705	-0.013(0.280)	0.054	-0.384/-0.512	-2.103/0.689
Emotional symptoms	5705	0.029(0.181)	-0.054	0.681/-0.183	-0.480/1.645
Emotional knowledge	5605	-0.072(0.424)	0.010	-0.707/0.859	-2.154/1.225
Expressive language	5702	-0.011(0.235)	0.033	-0.358/-0.395	-2.328/0.677
Receptive language	5606	-0.001(0.812)	0.038	-0.128/-0.209	-3.587/2.548
Executive function	5447	11.136(25.293)	13	-0.966/-0.342	0/16

Note. Overall observations of cases N=5707. The data included cases where there is no missing observation for the following tasks: Luria (executive function), PPVT-III (receptive language) or AKT (emotional knowledge). Factor scores of prosocial behaviours, emotional symptoms, emotional knowledge, expressive language and receptive language, and count scores of the executive function were used throughout the research. Original count scores of the executive function were used in the analysis in Table 5-1.

Table 5-2

Correlations Between Scores of the Selected Indicators

	Prosocial behaviours	Emotional symptoms	Emotional knowledge	Expressive language	Receptive language	Executive function
Prosocial behaviours	1.000	-	-	-	-	-
Emotional symptoms	-0.069	1.000	-	-	-	-
Emotional knowledge	0.063	-0.145	1.000	-	-	-
Expressive language	0.269	-0.182	0.226	1.000	-	-
Receptive language	0.051	-0.205	0.441	0.333	1.000	-
Executive function	0.085	-0.114	0.273	0.217	0.365	1.000

Note. Overall observations of cases N=5707. The data included cases where there was no missing observation for the following tasks: Luria (executive function), PPVT-III (receptive language) or AKT (emotional knowledge). Factor scores of prosocial behaviours, emotional symptoms, emotional knowledge, expressive language and receptive language, and count scores of the executive function were used throughout the research. Original count scores of the executive function were used in the correlation coefficient analysis in Table 5-2.

5.3.1 Group Enumeration

5.3.1.1 Variance-constrained unconditional LPA model.

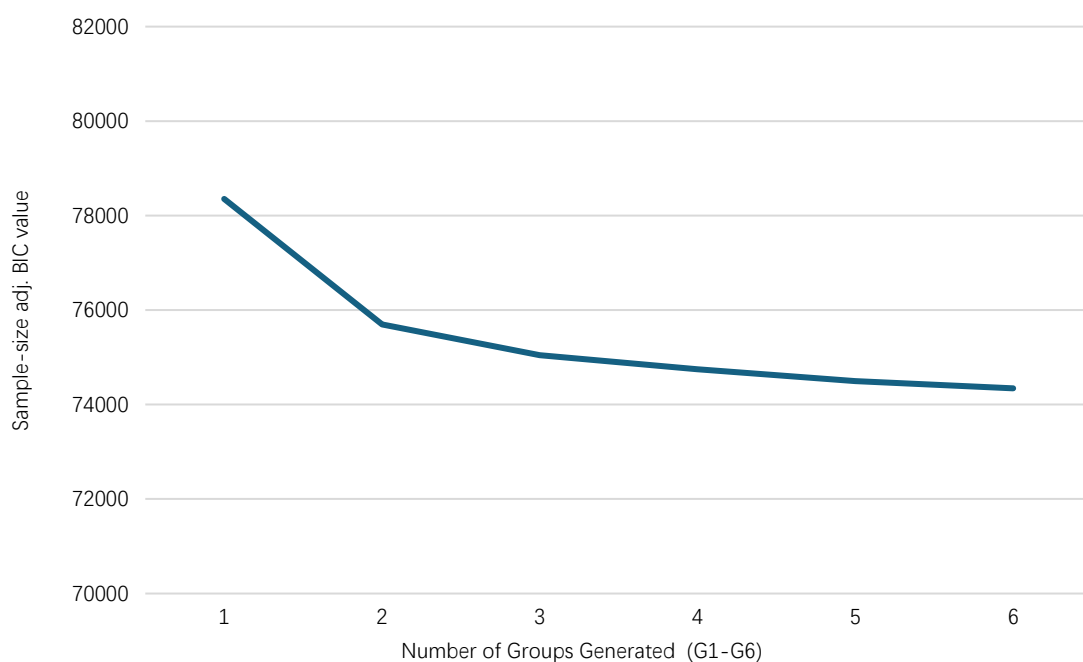
Group enumeration was conducted to identify the optimal number of group solutions represented in the sample. LPAs for 1- through 6-group models were estimated using unconditional models with variances fully constrained. Model fit statistics for each model were presented in Table 5-3. The entropy values for all models ranged from 0.64 to 1, indicating medium to high levels of group separation for these models. The posterior probabilities with a 0.7 minimum were achieved for 1-group to 3-group models and the 5-group model. The p-values in LMRT were all significant at the $\alpha < .01$ level, indicating that solutions with a greater number of profiles provided a better fit than the k-1 profile solutions. Separate descriptions and interpretations for 2- to 6-group models were given in the following sections.

Model fit improved with the number of profiles estimated based on changes in SABIC in Figure 5-2. Changes in SABIC in Figure 5-2 continued to decrease as additional profiles were added after extracting the 3-group model. The rate of change in SABIC values slowed and was not statistically significant after the 3-group model, as the slope of the line from the 3-to 4-group model appeared visually identical to that from the 4- to 5-group model. Also, the 'elbow' in the plot indicated that further increases in model complexity (i.e., more group solutions) did not yield the same decrease in SABIC. Although SABIC values continued to decrease, the 3- and 4-group models were identified as the most appropriate solutions, providing the best fit of the

data with entropy values indicating adequate group separation and relatively high accuracy of class assignment based on the six capabilities assessed (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language and executive function) scores.

Figure 5-2

The Elbow Plot of Sample-Size Adjusted Bayesian Information Criterion (SABIC) From Unconditional Latent Profile Models With Variances Fully Constrained of Children's Capabilities in the GUINZs 54-Month Dataset



Note. For the LPA analyses, count scores of the executive function in the LPA above were reversed, i.e., the number of correct responses vs. the number of errors were reversed, using a simple linear transformation (1-16 to 16-1) to address skewness in the Hand Clap task presented in 8.7 Appendix 7. Scores were then converted back to their original direction to ensure consistency and ease of interpretation across sample and grouping variables (i.e., gender and ethnic groups).

Table 5-3

Fit Statistics and Group Proportions for 1- to 6-Group Models With Variances Fully-Constrained of Children's Capabilities at 54-Month Assessment (n =5707)

Fit statistics	G1	G2	G3	G4	G5	G6
Log-Likelihood value	-39143.550	-37795.160	-37450.983	-37282.974	-37138.156	-37043.311
No. estimated parameters	12	19	26	33	40	47
<i>Information criteria</i>						
SABIC	78352.761	75694.283	75044.232	74746.515	74495.182	74343.794
Entropy	1.00	.707	.644	.708	.693	.671
Av. Posterior Probabilities range	1.00	0.86-0.93	0.81-0.84	0.64-0.90	0.73-0.84	0.68-0.81
<i>Relative model fit (k – 1)</i>						
LMR	-	2652.96, p < .001	1081.058, p < .001	330.560, p < .001	284.929, p < .001	186.609, p < .001
BLRT	-	p < .001	p < .001	p < .001	p < .001	p < .001
<i>Profile membership (most likely)</i>						
G ₁	5707 (100.0%)	1466 (25.7%)	2613 (45.8%)	611 (10.7%)	713 (12.5%)	493 (8.6%)
G ₂	-	4241 (74.3%)	518 (9.1%)	721 (12.6%)	2176 (38.1%)	718 (12.6%)
G ₃	-	-	2576 (45.1%)	1804 (31.6%)	477 (8.4%)	1195 (20.9%)
G ₄	-	-	-	2571 (45.1%)	1165 (20.4%)	2118 (37.1%)
G ₅	-	-	-	-	1176 (20.6%)	782 (13.7%)
G ₆	-	-	-	-	-	401 (7.0%)

Note. Groups represented profiles among children in the Growing Up in New Zealand Study from Latent Profile Analyses (fully constrained models across groups) for 1-6 group solutions. LPA = latent profile analysis; SABIC = Sample-size-adjusted BIC; the LMR test and the BLRT compare the current model to a model with k – 1 profiles. Av. posterior probabilities = average posterior probabilities. LMR = Lo-Mendell Rubin; BLRT = bootstrap likelihood ratio test; the LPA model constrained variances to be equal between groups.

Below were descriptions of the profile means in each profile with graphs (Figure 5-3 to Figure 5-7) and narrative descriptions of these means in 2- to 6-group models.

Profile Means refer to standardised mean scores that result in all profiles described concerning standard deviation below or above the sample mean.

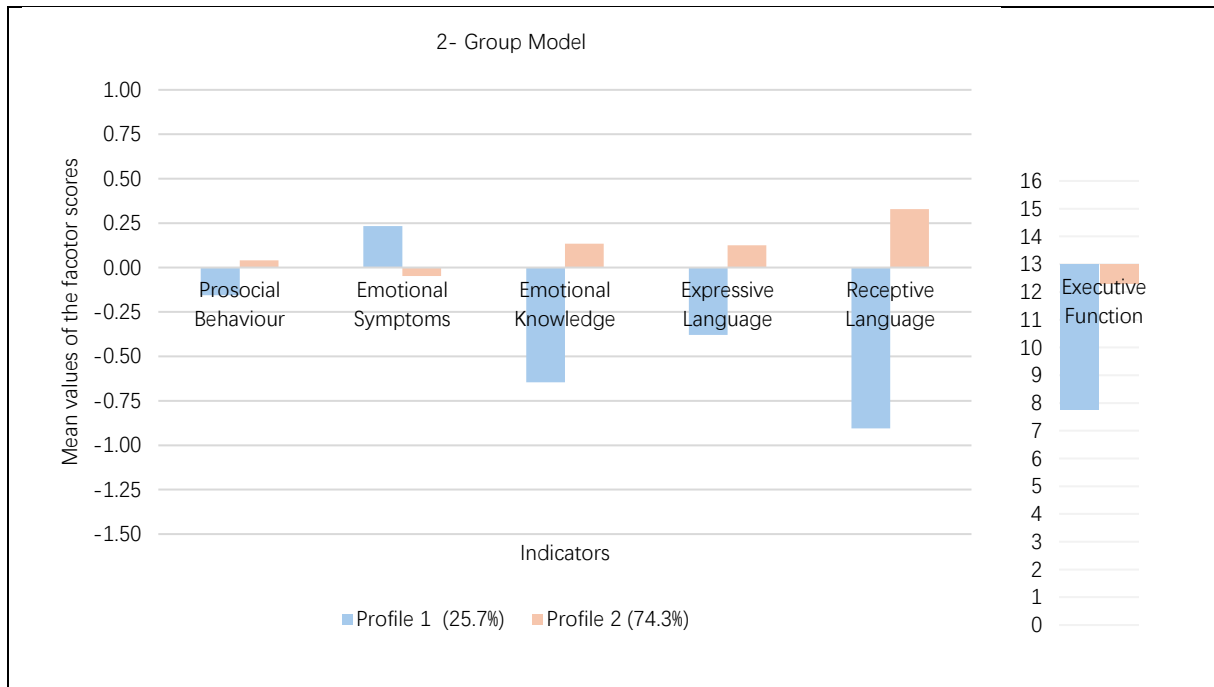
5.3.1.1.1 2-Group Model.

Fit statistics for the 2-group model in Table 5-3 showed the largest rate of change in $SABIC_{(2\text{-group vs. } 1\text{-group})} = 2658.48$, entropy = .71, with average posterior probabilities ranging from 0.86 to 0.93, and LMR and BLRT with significant p -values. Based on the rule-of-thumb of the criteria mentioned above, fit statistics for the 2-group model showed a good fit of the data.

As observed in Figure 5-3, the 2-group model indicated 25.3% of young children were best characterised by Profile 1 and 74.3% by Profile 2. Children in Profile 1 scored below the sample mean on prosocial behaviours, emotional knowledge, expressive language, receptive language, and executive function, and above the mean in emotional symptoms. Notably, scores for emotional knowledge and receptive language were more than half a standard deviation below the sample mean. In contrast, children in Profile 2 were characterised by scores that were close to or above the mean/median on prosocial behaviours, emotional knowledge, expressive language, receptive language, and executive function, and below the mean in emotional symptoms. Overall, these two profiles represented low and medium-high achievement in these capabilities.

Figure 5-3

Clustered Column With Standardised Scores of the 2-Group Model of Children's Capabilities Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Four of the five indicators represented by z-scores, The x-axis was set at zero such that scores above zero approximate group means that are above the sample average, and below zero represented scores below the sample average; For executive function (count range 0-16), bars represented group median scores charted relative to the grand median (sample median = 13).

5.3.1.1.2 3-Group Model.

Fit statistics for the 3-group model in Table 5-3 showed a smaller rate of change in $SABIC_{(3\text{-group vs. } 2\text{-group})} = 650.05$, entropy = .64, with average posterior probabilities ranging from 0.81 to 0.84, and LMR and BLRT with significant p -values. Based on the rule-of-thumb of the criteria mentioned above, fit statistics for the 3-group model showed a good fit of the data.

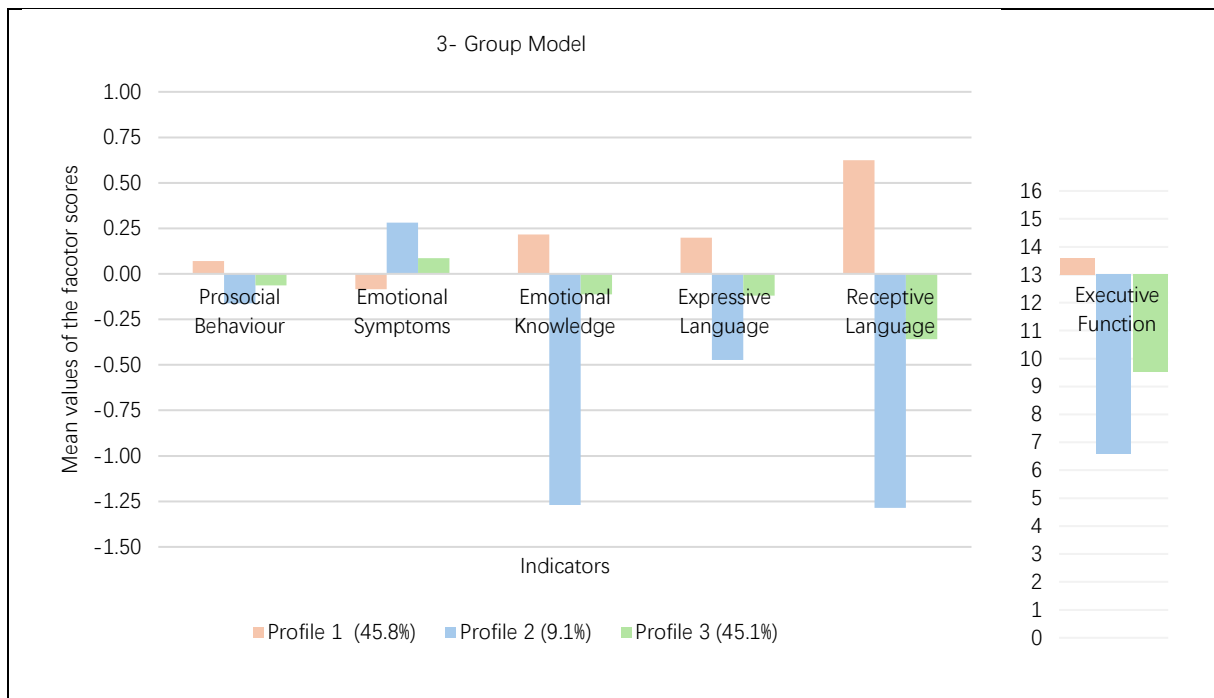
Overall, 45.8% of young children were identified in Profile 1 in Figure 5-4.

Compared to the other two profiles, young children's characteristics in Profile 1 scored highest in indicators of capabilities (above the means): factor scores' means of prosocial

behaviours, emotional knowledge, expressive language, receptive language, above the median score of executive function, and lowest in emotional symptoms. Whereas 9.1% of young children in Profile 2 showed the lowest scores in indicators of capabilities (below the means): prosocial behaviours, emotional knowledge, expressive language, receptive language, below the median of executive function, and highest in emotional symptoms. Profile 3 identified 45.1% of young children. Compared to Profile 1 and Profile 2, young children's capabilities in Profile 3 scores were close to the means/median, indicating means scores in prosocial behaviours, emotional knowledge, expressive language, receptive language, and median score of executive function were lower than Profile 1 but higher than that of Profile 2, and children in this profile score higher in emotional symptoms than Profile 1 but lower than that of Profile 2. Overall, these profiles represented groups of children who were below, at and above the mean on measured capabilities relative to their peers. Extraction of the smaller group refined the group characterised by below-average achievement in the 2-group model.

Figure 5-4

Clustered Column With Standardised Scores of the 3-Group Model Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as median count scores relative to grand median count scores (13).

5.3.1.1.3 4-Group Model.

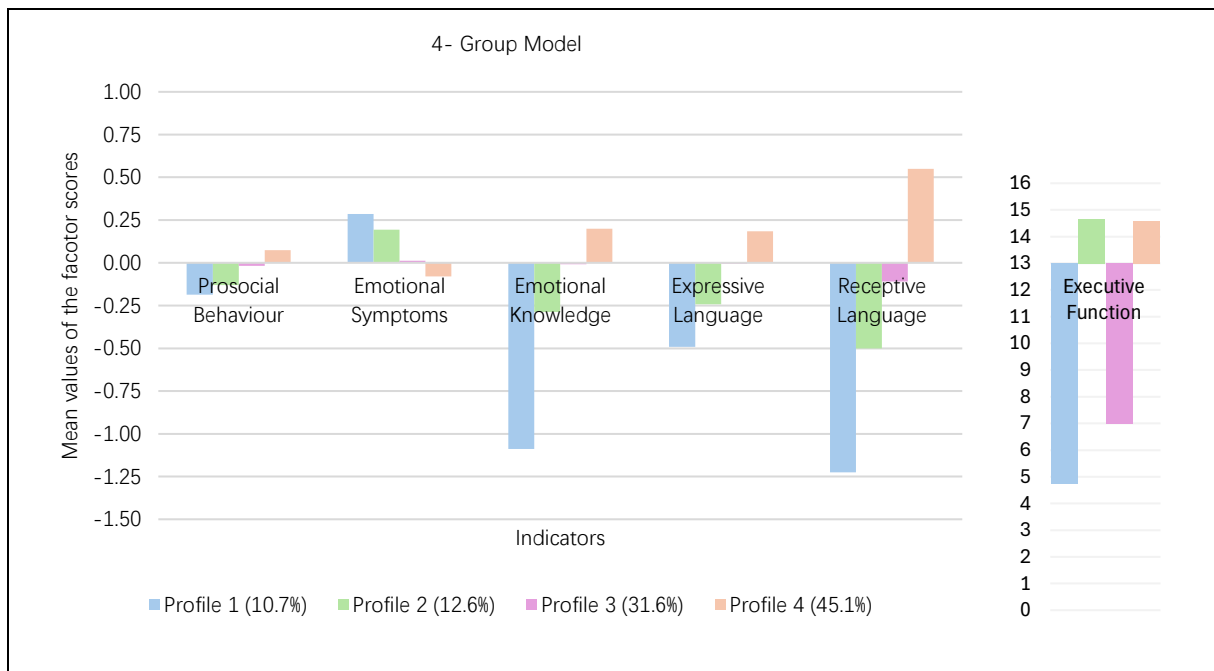
Fit statistics for the 4-group model in Table 5-3 showed the largest rate of change in $SABIC_{(4\text{-group vs. } 3\text{-group})} = 297.72$, entropy = .71, with average posterior probabilities ranging from 0.64 to 0.90, and LMR and BLRT with significant p -values. Based on the rule-of-thumb of the criteria mentioned above, fit statistics for the 4-group model showed a good fit of the data. As noted in 8.8 Appendix 8, profile membership transitions from 3-group and 4-group models, 45.8% of the total samples (5707) in Profile 1 (the largest group in the 3-group model) were split into Profile 4 and Profile 3 in the 4-group model. 7.7% of the total samples in Profile 2 of the 3-group model remained in Profile 1 of the 4-group model. Profile 3 of the 3-group model (the second largest group) was

split into Profile 3, Profile 2, and Profile 4 (from largest to smallest percentage) in the 4-group model.

As Figure 5-5 4-group model noted, 10.7% of children were identified in Profile 1. Compared to children in other profiles, children in this profile were characterised by the lowest scores in prosocial behaviours, emotional knowledge, expressive language, receptive language, and executive function. Moreover, they had the highest scores in emotional symptoms. Profile 2 identified 12.6% of children as being characterised by the second-lowest scores in prosocial behaviours, emotional knowledge, expressive language, and receptive language, but with the highest executive function and second-highest scores in emotional symptoms. Strengths within domains included executive function, with emotional symptoms and receptive language as relative areas of need. In Profile 3, scores of 31.6% of children's characteristics, including prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language, were characterised as close to the mean but with the second lowest executive function. Children in Profile 3 had relative strength in receptive language and relative needs in prosocial behaviours, emotional knowledge, and expressive language. Profile 4, the largest group with 45.1% identified in this group, showed that children had the highest prosocial behaviours, emotional knowledge, expressive language, and receptive language, the lowest emotional symptoms, and the second highest executive function.

Figure 5-5

Clustered Column With Standardised Scores of the 4-Group Model Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as median count scores relative to grand median count scores (13).

5.3.1.1.4 5-Group Model.

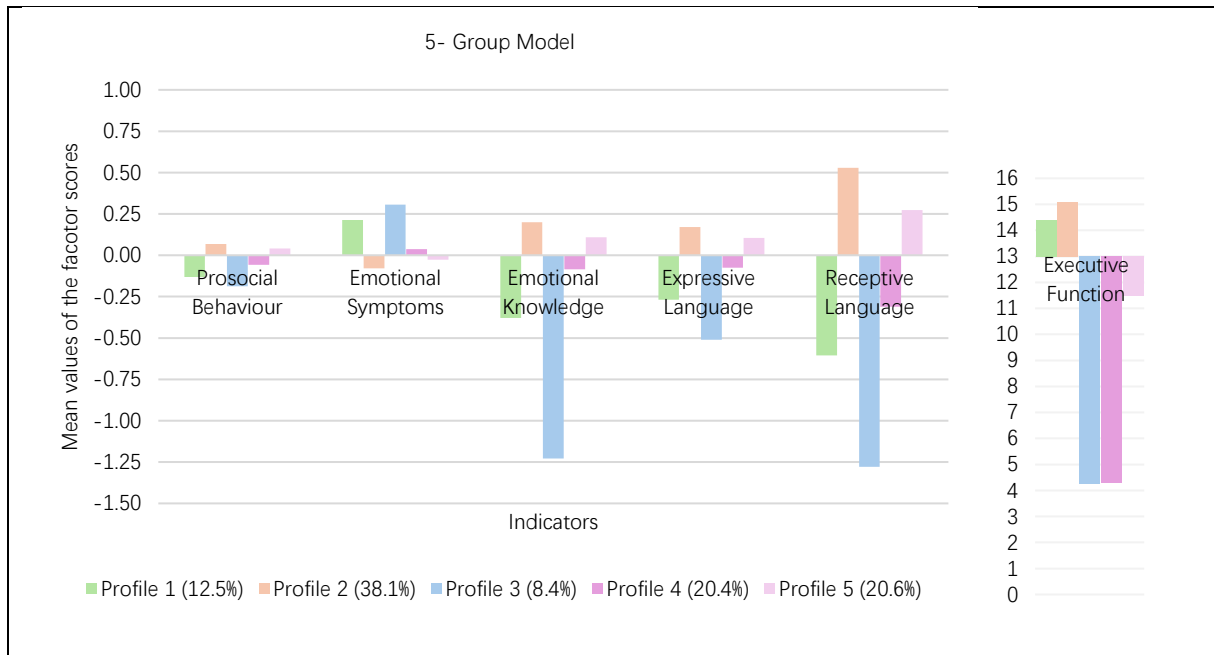
Fit statistics for the 5-group model in Table 5-3 showed a small rate of change in $SABIC_{(5\text{-group vs. } 4\text{-group})} = 251.33$, entropy = .69, with average posterior probabilities ranging from 0.73 to 0.84, and LMR and BLRT with significant p -values. Based on the rule-of-thumb of the criteria mentioned above, fit statistics for the 5-group model showed a good fit of the data. 8.8 Appendix 8 presents profile membership transitions from the 4-group and 5-group models. Profile 1 and Profile 2 in the 4-group model were retained in the 5-group model. The emerging group - Profile 5 was split from Profiles 3 and 4 in the 4-group model. Whereas the distinctions in the 5-group model were

mainly driven by executive function and emotional symptoms, scores of these capabilities do not differ significantly.

Under the 5-group model in Figure 5-6, Profile 1 identified 12.5% of children with the second lowest prosocial behaviours, emotional knowledge, expressive language, receptive knowledge, second highest executive function, and second highest emotional symptoms. In Profile 2, 38.1% of children were identified as having the highest prosocial behaviours, emotional knowledge, expressive language, receptive language, and executive function, but the lowest emotional symptoms. 8.4% of children were recognised in Profile 3 as having the lowest prosocial behaviours, emotional knowledge, expressive language, receptive knowledge, executive function, and the most significant emotional symptoms. The sizes of Profile 4 (20.4%) and Profile 5 (20.6%) were close but had different patterns. Differences in the patterns of Profile 4 and Profile 5 scores of children's capabilities were that children in Profile 4 had scores below the means of prosocial behaviours, emotional knowledge, expressive language, and receptive language, whereas scores of these capabilities in Profile 5 were above the means. Additionally, children in Profile 4 presented emotional symptoms above the mean and lower executive function than children in Profile 5.

Figure 5-6

Clustered Column With Standardised Scores of the 5-Group Model Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as median count scores relative to grand median count scores (13).

5.3.1.1.5 6-Group Model.

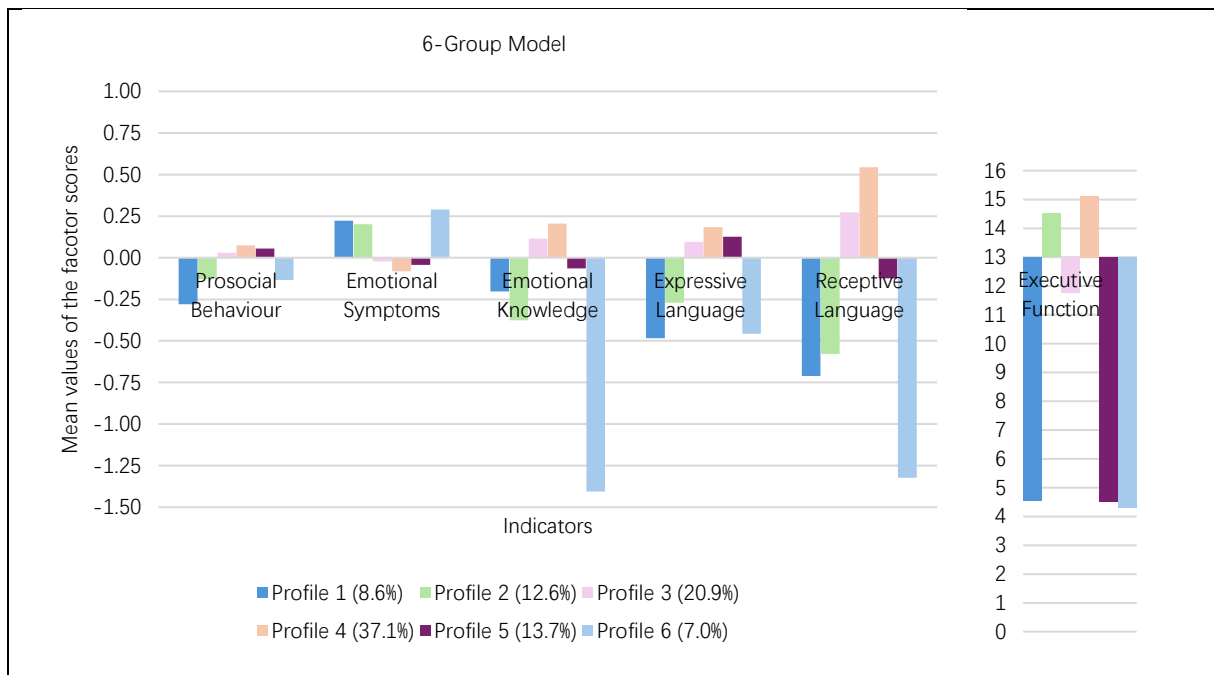
Fit statistics for the 6-group model in Table 5-3 showed a small rate of change in $SABIC_{(6\text{-group vs. } 5\text{-group})} = 151.39$, entropy = .71, with average posterior probabilities ranging from 0.68 to 0.81, and LMR and BLRT with significant p -values. Based on the rule-of-thumb of the criteria mentioned above, fit statistics for the 6-group model showed a good fit of the data. The 6-model presented distinct patterns of children's capabilities, but scores of these capabilities did not differ significantly, meaning that the distinction may not provide additional meaningful information.

Under the 6-group model in Figure 5-7, Profile 6 identified 7.0% of children with the lowest emotional knowledge, receptive language, and executive function as having

the highest emotional symptoms, close to the mean prosocial behaviours and the second lowest expressive language. Profile 1 identified 8.6% of children with the lowest prosocial behaviours, expressive and receptive language, second lowest executive function, second highest emotional symptoms, and close to the mean emotional knowledge. Noticeably, Profiles 1 and 2 shared similar patterns but did not display significant distinctions, with Profile 1 displaying lower executive function. 20.9% of children were recognised in Profile 3 as having prosocial behaviours, emotional knowledge, expressive language, receptive knowledge, executive function, and emotional symptoms close to the means. The similarities between Profile 4 (37.1%) and Profile 5 (13.7%) were that prosocial behaviours and expressive language were above the means, and emotional symptoms were below the means. Children in Profile 4 scored highest in emotional knowledge, receptive language, and executive function, while these characteristics in Profile 5 were below the means.

Figure 5-7

Clustered Column With Standardised Scores of the 6-Group Model Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as median count scores relative to grand median count scores (13).

5.3.1.2 Variances-unconstrained LPA model.

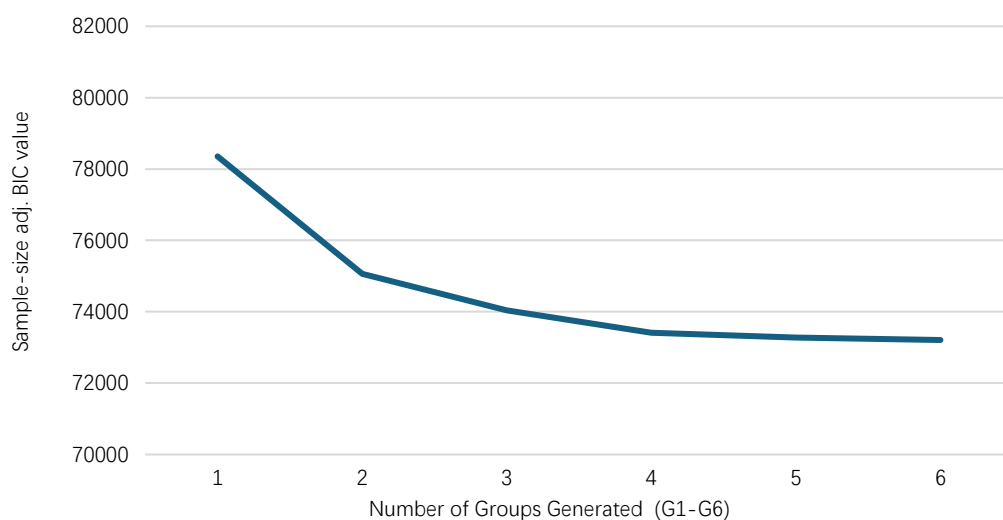
Table 5-4 presented fit statistics and proportions for model selection with variances freed from 1- to 6- Models. As noted in Table 5-4, starting from the 4-group model, model estimation outputs showed error messages for fixing the dispersion parameters for Luria Hand Clap task scores in Profile 1 and Profile 4. This indicated that the 4-group model with variances freed may not converge well with the data from multiple measures. The estimation of the 5-group and 6-group variance-constrained models required constraints on estimating the dispersion parameter associated with the Luria Hand Clap Task that were not consistent with observed parameters from prior models. Thus, these models were not considered further. Here, due to the error messages for

fixing parameters occurring from 4- to 6-group models, it is considered that the 3-group model was highly defensible. Compared to unconditional LPA models, models with variance freed were not considered a better fit for the data.

Model fit for variance-freed unconditional LPA models with the number of profiles estimated based on changes in SABIC was presented in Figure 5-8. Similar to the variances-constrained unconditional LPA models, changes in SABIC in variance-freed unconditional LPA models continued to decrease as additional profiles were added after extracting the 3-group model. The rate of change in SABIC values was also slowed and not significant after the 3-group model. Additionally, the 'elbow' in the plot indicated that further increases in model complexity (i.e., more group solutions) did not yield the same decrease in SABIC.

Figure 5-8

The Plot of Sample-Size Adjusted Bayesian Information Criterion (SABIC) From Unconditional Latent Profile Models With Variances Freed of Children's Capabilities in the GUINZs 54-Month Dataset



Note. count scores of the executive function in the LPA above were reversed, i.e., the number of correct responses vs. the number of errors were reversed.

Table 5-4

Fit Statistics and Group Proportions for 1- to 6-Group Models With Variances Freed of Children's Capabilities at 54-Month Assessment (n =5707)

Fit statistics	G1	G2	G3	G4	G5	G6
Log-Likelihood value	-39143.550	-37462.947	-36917.611	-36567.660	-36463.063	-36393.357
No. estimated parameters	12	25	38	51 [†]	64 ^{††}	77 ^{†††}
<i>Information criteria</i>						
SABIC	78352.761	75062.689	74043.149	73414.378	73276.319	73208.039
Entropy	1.00	.702	.689	.710	.653	.668
Av. Posterior Probabilities range	1.00	0.87-0.93	0.79-0.90	0.75-0.89	0.72-0.81	0.62-0.81
<i>Relative model fit (k – 1)</i>						
LMR	-	3331.58, p < .001	1081.058, p < .001	694.843, p = .004	207.594, p = .025	138.347, p = .188
BLRT	-	p < .001	p < .001	p < .001	p < .001	p < .001
<i>Profile membership (most likely)</i>						
G ₁	5707 (100.0%)	1585 (27.8%)	1017 (17.8%)	914 (16.0%)	1035 (18.1%)	468 (8.2%)
G ₂	-	4122 (72.2%)	1987 (34.8%)	2367 (41.5%)	532 (9.3%)	516 (9.0%)
G ₃	-	-	2703 (47.4%)	1857 (32.5%)	824 (14.4%)	451 (7.9%)
G ₄	-	-	-	569 (10.0%)	1854 (32.5%)	1762 (30.9%)
G ₅	-	-	-	-	1462 (25.6%)	1018 (17.8%)
G ₆	-	-	-	-	-	1492 (26.1%)

Note. Groups represented profiles among children in the Growing Up in New Zealand Study from Latent Profile Analyses (variances freely estimated between groups) for 1-6 group solutions. LPA = latent profile analysis; SABIC = Sample-size Adjusted BIC; LMR = Lo-Mendell Rubin; BLRT = bootstrap likelihood ratio test; the LMR test and the BLRT compared the current model to a model with k – 1 profiles; the LPA model allowed unequal indicator variances across groups; † dispersion parameter for Luria clap task estimate was constrained to zero in model G4 profile 1 (group with the highest indicator mean = 2.490, p < .001, with dispersion fixed at 0.000) and profile 4 (group with the Luria indicator mean = 0.291, p = .001, with dispersion fixed at 0.000) to support estimation – G4 profile 1 parameter did not significantly vary from zero in the proceeding G3 profile-varying variance structure model for group G3 profile 1, from which the majority of G4 profile 1 group members were derived i.e., G4 profile 1 indicator mean = 2.490, p < .001 and dispersion = 0.000, p < .001; †† dispersion parameter for Luria hand clap task estimate was constrained to zero in group G5 profile 2 (group with the highest indicator mean = 2.504, p < .001, with dispersion fixed at 0.000) and profile 3 (group with the 2nd highest indicator mean = 2.492, p < .001, with dispersion fixed at 0.000) to support estimation, with these groups largely representing a split of individuals from the G4 profile 2 and profile 3, in which this parameter significantly differed from zero i.e., G4 profile 2 Luria task indicator mean = 1.360, p < .001 and dispersion = 1.024, p < .001 and G4 profile 3 indicator mean = 1.367, p < .001 and dispersion = 1.128, p < .001; ††† dispersion parameters for Luria hand clap task scores were constrained to zero in groups G6 profile 1, G6 profile 2 and G6 profile 3.

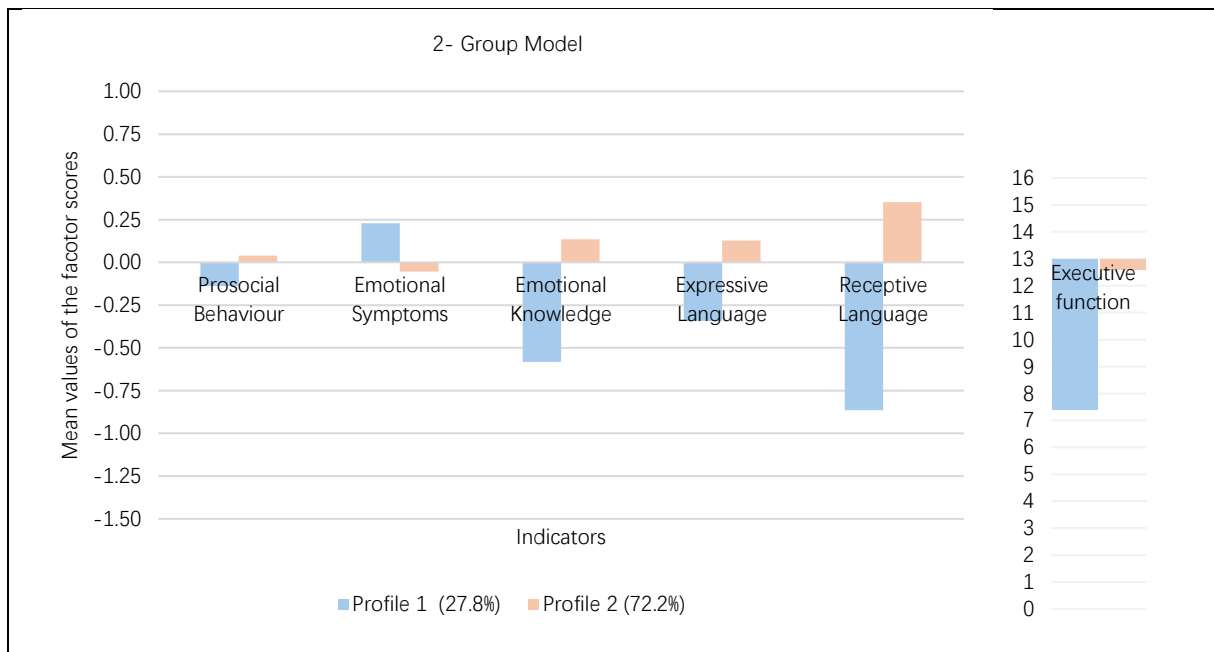
5.3.1.2.1 2-Group Model.

Fit statistics for the 2-group model in Table 5-4 showed the largest rate of change in $SABIC_{(2\text{-group vs. } 1\text{-group})}=3290.07$, entropy = .70, with average posterior probabilities ranging from 0.87 to 0.93, and LMR and BLRT with significant p -values. Based on the rule-of-thumb of the criteria mentioned above, fit statistics for the 2-group model showed a good fit of the data.

As observed in Figure 5-9, 7.8% of young children in Profile 1 and 72.2% in Profile 2 were identified in the 2-group model. Children in Profile 1 scored lowest in indicators of strengths-based capabilities (below the mean): mean scores of prosocial behaviours, emotional knowledge, expressive language, receptive language, executive function, and highest in emotional symptoms. Whereas 72.2% of young children in Profile 2 show the highest scores in indicators of strengths-based capabilities (above the mean): mean scores of prosocial behaviours, emotional knowledge, expressive language, receptive language, executive function, and lowest in emotional symptoms.

Figure 5-9

Clustered Column With Standardised Scores of the 2-Group Model With Variances Freed Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as median count scores relative to grand median count scores (13).

5.3.1.2.2 3-Group Model.

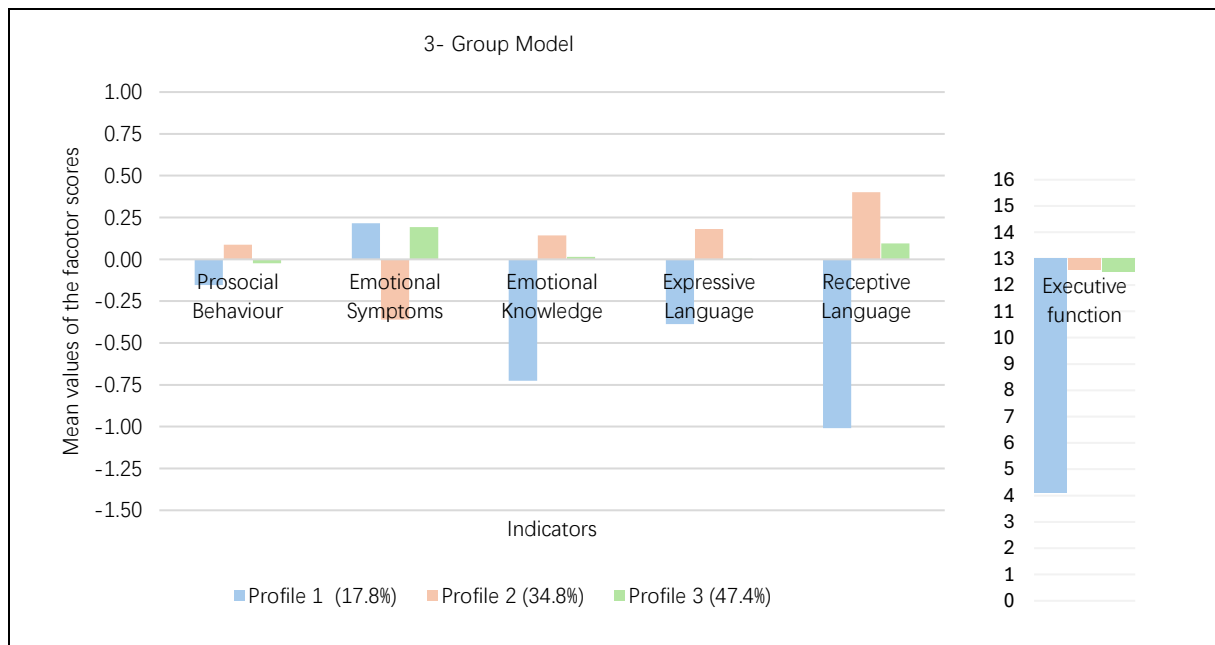
Fit statistics for the 3-group model in Table 5-4 showed a smaller rate of change in $SABIC_{(3\text{-group vs. } 2\text{-group})} = 1019.54$, entropy = .69, with average posterior probabilities ranging from 0.79 to 0.90, and LMR and BLRT with significant p -values. Based on the rule-of-thumb of the criteria mentioned above, fit statistics for the 2-group model showed a good fit of the data.

In the 3-group model with variances freed, 17.8% of young children were identified in Profile 1 in Figure 5-10. Compared to the other two profiles, young children's capabilities in Profile 1 scored the lowest (below the means) in indicators of prosocial behaviours, emotional knowledge, expressive language, and receptive language, below the median scores of executive function, and highest emotional symptoms. 34.8% of

young children in Profile 2 showed the highest scores (above the means) in indicators of prosocial behaviours, emotional knowledge, expressive language, and receptive language, close to median scores of executive function, and lowest in emotional symptoms. Profile 3 identified 47.4% of young children. Compared to Profile 1 and Profile 2, young children's capabilities in Profile 3 scores were close to the means/median, indicating means scores of prosocial behaviours, emotional knowledge, expressive language, and receptive language were lower than Profile 2 but higher than those of Profile 1. Median scores of executive function in Profiles 2 and 3 were similar. Children in Profile 3 scored higher in emotional symptoms than Profile 2 but lower than those in Profile 1.

Figure 5-10

Clustered Column With Standardised Scores of the 3-Group Model With Variances Freed Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as median count scores relative to grand median count scores (13).

5.3.1.2.3 4-Group Model.

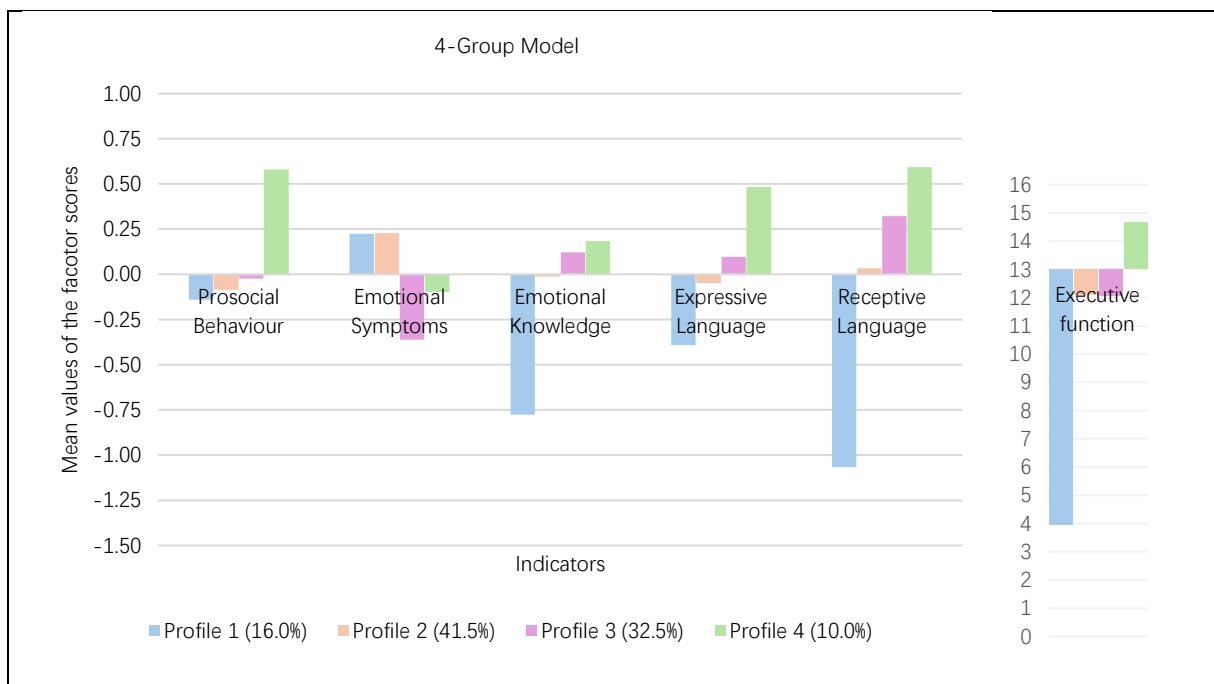
Fit statistics for the 4-group model in Table 5-4 showed the largest rate of change in $SABIC_{(4\text{-group vs. } 3\text{-group})} = 628.778$, entropy = .71, with average posterior probabilities ranging from 0.75 to 0.89, and LMR and BLRT with significant p -values. Based on the rule-of-thumb of the criteria mentioned above, the fit statistics indicated that the 2-group model provided a good fit of the data. However, the 4-group model with variances freed may not converge well with the data from multiple measures associated with fixing error parameters presented in Table 5-4. Error parameters were also observed in estimating the 5-group and 6-group variances-constrained models, which required constraints on estimating the dispersion parameter associated with the Luria Hand Clap Task. Thus, these models were not considered further.

As Figure 5-11 noted, 16.0% of children were identified in Profile 1. Compared to children in other profiles, children in this profile were characterised by the lowest scores in prosocial behaviours, emotional knowledge, expressive language, receptive language, and executive function and the highest scores in emotional symptoms. All the above capabilities were the key focus of this profile, and children in this group may need support across these areas of development. Profile 2 identified 41.5% of children as being characterised by the second-lowest scores in prosocial behaviours. Still, children's emotional knowledge, expressive language, and receptive language scores were close to the mean. Children in Profile 2 had the second-highest scores in executive function and emotional symptoms, identified as relative areas of need. In

Profile 3, 32.5% of children’s capabilities were characterised as emotional knowledge, expressive language, and receptive language above the mean, with prosocial behaviours below the mean and executive function below the median and the lowest in emotional symptoms. Children in Profile 3 may have relative needs to support their emotional symptoms. Profile 4, the smallest group with 10% identified in this group, showed that children had the highest scores in prosocial behaviours, emotional knowledge, expressive language, receptive language, and executive function, with the lowest emotional symptoms.

Figure 5-11

Clustered Column with Standardised Scores of the 4-Group Model With Variances Freed Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as median count scores relative to grand median count scores (13).

5.3.2 Optimal Solution

Figures 5-3 to 5-7 presented emerging profiles in the 2- to 6-group model using variances-constrained unconditional LPA. Figures 5-9 to 5-11 presented emerging profiles in the 2- to 4-group model using variance-freed unconditional LPA. Consistencies were observed in the models generated in both ways. In references to Tables 5-3 and 5-4, model fit indices using information criteria and group sizes, and theoretical justifications were assessed for the optimal solution. Groups generated in the unconditional LPA models with variances unconstrained were similar to the unconditional LPA models with variances fully constrained. Due to the model converging issues in the parameters of executive function, variances-unconstrained models were not considered in the optimal solution. Overall, in the variances-constrained unconditional LPA, the rate of changes in SABIC slowed after the 3-group model, entropy values for all models were medium to high, and LMR and BLRT showed significant p -values. Separate evaluations for each model were conducted with the conceptual considerations.

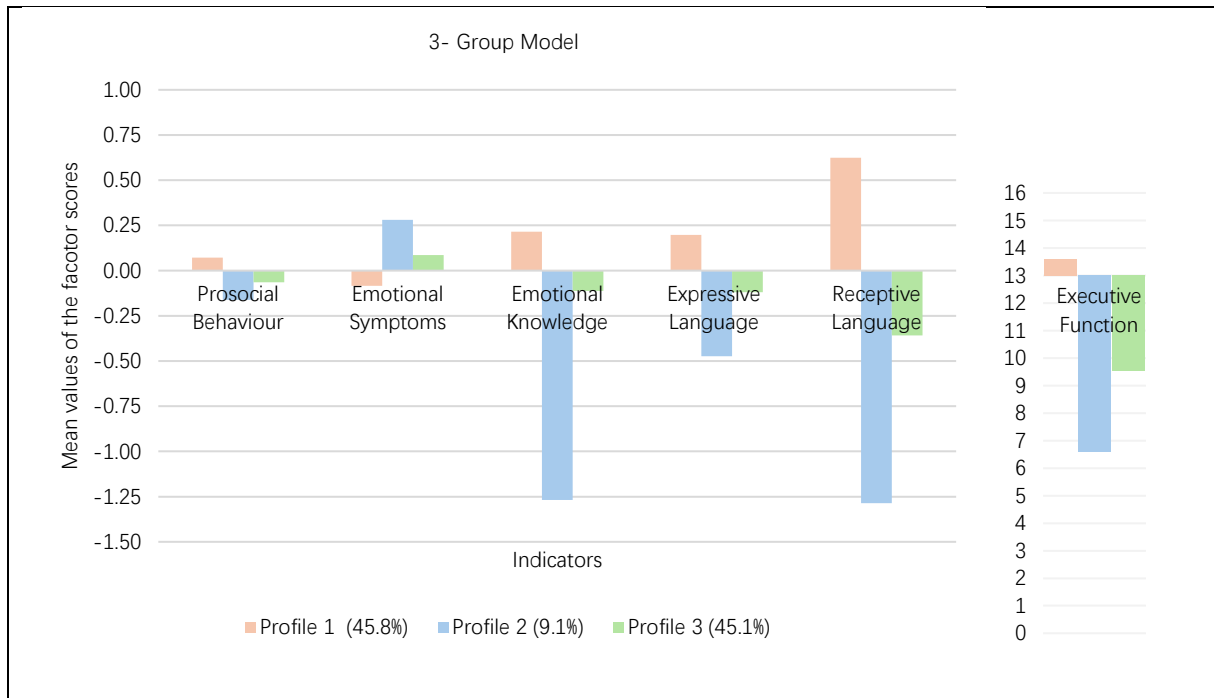
Differences in the unconditional LPA models with variances fully constrained were observed. The 2-group model children emerged with above and below the means of indicators, except for high or low median executive function scores. However, this model may not adequately represent the groups, as it may not capture the heterogeneity of children's capabilities. In contrast, the 3-group model grouped children into three profiles reflecting overall differences in children's capabilities (i.e.,

above, below, and close to the mean), which is broadly consistent with a previous study by Denham et al. (2012). The 4-group model indicated a significant difference between the 3-group model and the 4-group model in executive function. There was evidence for a sub-population in the variance-constrained 4-group model, with relative strength in executive function among a group (12%), who were otherwise performing more poorly relative to peers on other indices. However, this model was not supported in the variance-unconstrained model. In the 5-group model, the emerging group reflected a split of Profile 3 and Profile 4 from the 4-group models—both characterised by strengths across all areas of development—into a single large group representing 20.6% of the sample. This group did not distinguish significantly, with Profile 2 representing children displaying higher scores on executive function. The 6-group model presented distinct patterns of children's capabilities, but the distinction may not provide additional meaningful information to inform decision-making. The group sizes of the smallest groups comprising 7.0%- 10.7 % of the sample in 3- through 6-group models did not differ significantly. Thus, based on the above considerations that include fit indices and conceptualisation of children's capabilities, particularly the 2-, 4- through 6-group models were seen to add little substantive information to the models, the 3-group model was chosen as the more parsimonious representation of the data that represented genuine heterogeneity within the population, as illustrated in Figure 5-4 and shown again below as Figure 5-12. In addition to being statistically and conceptually defensible, the 3-group model was consistent with the hypothesis for the

current work based on prior literature (Collie et al., 2019; McLaughlin et al., 2024; Russell et al., 2022).

Figure 5-12

Clustered Column With Standardised Scores of the 3-Group Model Across Six Indicators



Note. The figure showed the mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language) and the executive function count scores. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as median count scores relative to grand median count scores (13).

The description of the 3-group model was presented by sub-group number or profile number. Each sub-group represented a distinct latent profile of children's development capabilities associated with a percentage of the sample population. It is important to note that the extent to which each profile was a good fit for any given pattern of capabilities can be examined through the posterior probabilities generated for each case. Average posterior probability scores ranged from 0.81 to 0.84, which suggested that assigned group membership matched the measured capabilities, while remembering that each child still had their unique profile of scores.

Labelling latent subgroups has been commonly used to characterise the most salient features of the profile efficiently. Here, I labelled the three profiles in the 3-group model as *Emerging* (profile 2), *Progressing* (profile 3), and *Competent* (profile 1) capabilities. These labels have been chosen because they correspond with the descriptions and expectations associated with children's capabilities outlined in *Te Whāriki* (Ministry of Education, 2017). In the *Emerging* profile, children showed the lowest prosocial behaviours, emotional knowledge, expressive language, receptive language, and executive function, and the highest emotional symptoms. In the *Progressing* profile, children showed close to average prosocial behaviours, emotional knowledge, expressive language, and receptive language, with executive function higher than children in the *Emerging* profile but lower than that of the *Competent* profile. Children in this profile scored higher in emotional symptoms than children in the *Competent* profile but lower than those in the *Emerging* profile. In the *Competent* profile, children showed the highest prosocial behaviours, emotional knowledge, expressive language, receptive language and executive function, and the lowest emotional symptoms.

5.4 Discussion

The first phase of the research aimed to understand patterns of young children's social, emotional, language and executive function capabilities in the New Zealand context using 54-month datasets from the GUiNZs. A person-centred approach through latent profile analysis supported the holistic view of children's capabilities.

Based on examining the optimal solution, children in the *Competent* profile showed significant strengths from same-aged peers in several domains assessed. While the group of children in the *Progressing* profile was close to the average, it is important to note that some children within this group still showed lower levels in specific domains. Other profile memberships suggested some children may require additional support in specific domains, such as receptive language and executive function capabilities, based on their unique profiles of capabilities. The *Emerging* profile characterised 9.1% of the 5707 young children in the sample by displaying prosocial behaviours, emotional knowledge, and expressive and receptive language below the sample mean, with emotional symptoms higher than the sample mean and the lowest executive function. This suggests a small proportion of children may benefit from additional support across areas of development to enhance their emerging capabilities.

Given that most previous studies predominantly adopted a variable-centred approach to assessing young children's characteristics in New Zealand contexts, the current study considered the characteristics from a person-centred approach and identified overall differences across the six capabilities of preschool-aged children. Specifically, all capabilities showed an overall trend in patterns of association across the profiles. The current research suggests that the nature of these capabilities may represent general capabilities as a whole. The results were similar to but not precisely the same as the 3-profile model of young children's social-emotional learning in the previous research (Denham et al., 2012).

Importantly, the current research focuses on a comprehensive set of young children's strengths-based capabilities (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function), which can be used for the extraction of empirically reliable profiles of young children's capabilities. Consistent with the study by McLaughlin et al. (2024), the research supports a positive view of children's development while considering multiple capabilities through a holistic lens. This also aligned with the strengths-based and holistic perspectives in *Te Whāriki*, which underscores the importance of delineating strengths-based capabilities at the preschool stage.

5.5 Implications

A key strength of characterising young children's capabilities is to guide early childhood practices and intervention programmes. Children in the *Competent* profile showed positive development across the six capabilities. Interventions and programmes for children in the *Progressing* profile may benefit from boosting these strengths-based capabilities and additional strategies to minimise the impact of emotional symptoms in their development. In contrast, children in the *Emerging* profile would likely benefit from practices and intervention programmes that focus on all capabilities included in the study (prosocial behaviours, emotional symptoms, emotional knowledge, expressive and receptive language, and executive function). Strategies need to focus on building strengths-based capabilities and reducing emotional symptoms. Each profile has been described in scores across and within profiles to understand each profile's comparative

and relative strengths and needs. It is also possible, and potentially helpful, to consider how these profiles provide implications for support in early childhood practices.

The current research reflects the strengths-based, holistic view of development in *Te Whāriki*, where knowledge associated with young children's social, emotional, language and executive functions is woven through all strands of *Te Whāriki*. The interconnectedness of these domains from a holistic lens helps develop an understanding of each child. Rather than focusing on problems or deficits in single domains of development, this research highlights children's capabilities, clustered together in the current research. Results have indicated that children display an overall trend in capabilities, meaning that the child may be more or less capable across areas of development (prosocial behaviour, emotional knowledge, expressive and receptive language, and executive function). This finding is consistent with the view that variation in children's capabilities is expected and that each child is on their own learning trajectory. Instead of focusing on a single domain of child development, this research examined six domains of capabilities holistically. This may also guide the teaching practice of a comprehensive early childhood curriculum and help shape a rich learning environment that nurtures children's capabilities across all domains (McLachlan, 2018; Ministry of Education, 2017).

Implications for education services based on the 3-group model can be considered within the evidence-informed, multi-tiered approach outlined in *Te Tūāpapa o He Pīkorua* (Ministry of Education, 2020). The information about children's strengths,

needs, and progress supports collaborative decision-making within the multi-tiered approach. The 3-tiered support framework increases in intensity to support inclusive learning opportunities for all children, depending on the needs and the context. The model is dynamic and integrated from the universal (Tier 1), targeted (Tier 2) and individualised (Tier 3). Tier 1 involves universal interventions and supports, with the notion of the needs of all children. Tier 2 represents targeted support for those who may need extra support (a child who needs support in receptive language and executive function). Tier 3 provides additional support for children with unique needs. Overall, the approach sees practices and systems as a continuum based on assessing the needs of children and identifying systemic patterns.

Current tailored intervention programmes provided by the Ministry of Education focus on single targeted domains. For example, the Incredible Years programme as part of Positive Behaviour for Learning (PB4L), is designed to reduce challenging behaviours and increase the social and behavioural skills of children aged 3-8. The ENGAGE programme focuses on self-regulation/self-control skills for children aged 3 to 5. The OLLi programme mainly focuses on 3- and 4-year-old children's expressive language and literacy. These programmes tend to be offered to all children, with the OLLi offering tiered services to children with tailored language support. Given the current findings, these selective programmes may be most appropriate for children in the *Progressing* and *Emerging* profiles. Moreover, resources and supports that enable educators to recognise and make connections across domains of learning and

development may strengthen their ability to support children's interconnected and holistic learning and development.

5.6 Limitations and Future Directions

Findings are not without limitations. First, although this current work has centred on identifying young children's strengths-based characteristics (prosocial behaviour, emotional knowledge, expressive and receptive language, executive function) with minimal emphasis on their difficulties, it is important to acknowledge that behavioural difficulties (conduct problems, peer problems, etc.) do have an impact on children's engagement and learning in early childhood education and school. The investigation of these areas is an important part of supporting children's learning now and in the future, but was outside the scope of the present research. In addition, the research only looked at 54-month-old children's characteristics. The research did not examine whether and how patterns of young children's characteristics change as they enter early childhood settings and progress through school. Moreover, the implications for future academic success are not known, nor should they be inferred from profiles presented in the current research.

Second, issues are drawn from measures. First, the measures used in the current research were from parent-reported and direct observations, while there were no teacher-reported measures within the early childhood settings. It is essential to explore how patterns of young children's characteristics identified in the research compare with those derived from other informants' reports, such as teachers. Limitations of parent-

reported measures may involve biased views of child development and parents' limited knowledge of the characteristics measured. Limitations of child observation tasks, including the Affect Knowledge Test (AKT) and Hand Clap Task, may include limited relevance of tasks and activities that can impact a child's performance. Second, the current research could not validate the raw scores of PPVT-III due to limited data access. The PPVT-III has been validated only for receptive vocabulary in English; it does not assess children's bilingual or multilingual abilities. This suggested that children who are English as second-language learners, may be impacted on their performance on this English vocabulary assessment. Additionally, the current research could not validate the Hand Clap Task observation because its scoring relies on count variables.

Third, there are limitations of sample size and the ability to detect low-frequency profiles. As noted in past studies (Collie et al., 2019; McLaughlin et al., 2024; Russell et al., 2022), large proportions of samples may represent homogeneous populations, with relatively small proportions displaying distinctive patterns of association on indicators of development. For instance, the 4-group model for young children's social and emotional behaviours observed in the larger Australian study by Collie et al. (2019) displayed population homogeneity, but with small proportions of the population displaying distinct profiles. In the current research, the three groups are not truly distinct in terms of patterns within and across the profiles. Only 3 profiles were extracted in the current research—*Emerging*, *Progressing* and *Competent* capabilities profiles within a relatively homogeneous population. Moving forward, it would be

beneficial to expand the sample sizes, as larger samples may enable the detection and extraction of lower frequency profiles with a more distinct pattern of development.

Fourth, the replicability and interpretability of the findings are limited in terms of sample, time, place, and analytical method. This study only looked at children aged 54 months in the GUiNZs dataset, and the sample represents children born in the Auckland, Manukau, and Waikato District Health Boards in 2009-2010. Children today may display different patterns of characteristics in other samples, and children who participated in the GUiNZs may display different patterns of capabilities at different ages. Additionally, critiques have been raised concerning the person-centred analysis technique's replicability and interpretability due to its exploratory, data-driven nature. Specifically, the results generated in the current research may not be replicable in other samples, as the study is based on a cohort broadly representative of New Zealand children born between 2009 and 2010. Researchers determine the class enumeration and optimal solution based on fit indices and theoretical considerations. However, it is possible that different researchers, or those from other disciplines, might choose and report on models with 4- to 6-group or interpret the patterns of young children's characteristics differently. Hence, caution is needed when generalising the results beyond the population examined in the research and this specific research purpose. Although this model has strengths in the examination of the invariance of individual indicators of development across children's gender and ethnicity, a notable limitation

relates to the invariance of the model as a whole across these observed groups. This issue is explored in greater depth in later chapters.

5.7 Conclusion

Based on a combination of statistical and theoretical considerations, the first phase of the thesis has chosen a 3-group model of young children's capabilities (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function) in the GUiNZs 54-month dataset. These findings are informative for early childhood education practice and policy in the New Zealand context. Noticeably, caution is needed when generalising beyond the population examined in the study regarding the replicability of the person-centred analysis techniques to other samples or the similarities and differences between diverse groups in the cohort. Thus, the second phase of the thesis was to check the measurement invariance (MI) at the person-centred model level across the child's gender and ethnic groups. Prior to that, the MI of the selected measures was assessed across the child's gender and ethnicity. This was to ensure the examination of comparability on the person-centred model across the diverse groups (i.e., gender and ethnicity). The methods and results for the second phase are presented in Chapter 6.

Chapter 6 Phase 2: Methods and Findings

The purpose of the present chapter is to address the second phase of my research, which focused on exploring the Differential Item Functioning (DIF) effects of the six profile factors across the child's gender (i.e., boys and girls) and ethnicity (i.e., European, Māori, Pacific Peoples, Asian, Other) in the person-centred model. This is important to better understand the nature of the 3-group model generated in Phase 1 and factors that may compromise its validity/generalisability, such as child gender and ethnicity. Recognising and addressing potential biases introduced by these factors is essential to ensure measurement quality and fairness in interpreting the results.

To prepare for Phase 2, I first explored the potential biases by assessing the MI of each measure used as an indicator of young children's social, emotional, language and executive function in the current research by the child's gender and ethnicity group. As preparatory background work for the Phase 2 analysis, these analyses are presented in **Appendices 3** through **7**. For the primary analyses of Phase 2, the child's gender and ethnicity were included in this chapter as covariates of the 3-group model and its six profile factors, to assess and model the DIF effects on latent profile factors through Multiple Indicators and Multiple Causes (MIMIC) modelling.

DIF, a form of measurement non-invariance (MNI), addressed with an item response theory (IRT) framework, is a statistical method that detects item bias, compares respondent groups at the item level and identifies non-invariant items while accounting for differences in the latent construct of interest (Dorans & Holland, 1993;

Wallin et al., 2024). DIF arises when an item measures more than one latent construct across different observed groups. For example, an item demonstrates DIF if it functions differently for participants in two or more observed subgroups. This introduces construct-irrelevant variance into item scores, which is attributable to group membership rather than the intended construct (Nugent, 2017). There are several approaches to evaluate covariates as potential sources of DIF within the latent variable modelling. Among these, MIMIC modelling has been the most commonly employed approach, as it offers greater flexibility in accommodating various types of variables as potential sources of DIF than other methods (Bettencourt et al., 2022). Therefore, the current thesis chose the MIMIC modelling technique to analyse the DIF effects from covariates to the latent profile factors.

The MIMIC modelling represents a novel stepwise approach for the DIF effects from covariates on both a latent variable and its indicators (Finch, 2005). Models tested in the MIMIC modelling framework include a measurement model, which assesses the relationship between a latent variable and its indicators, and a structural model, which examines the direct effect of a covariate (e.g., gender) that compromises the group membership on factor means and indicators (Tsaousis et al., 2020). These effects can be understood through two distinct types of DIFs: uniform and nonuniform DIFs. Uniform DIF occurs when the probability of responding correctly to an item is consistently higher for the reference group across all levels of ability, meaning the item is related to the group membership at every level of the latent trait (Finch, 2005; Tsaousis et al.,

2020). In contrast, nonuniform DIF occurs when items discriminate differently across groups. It is important to detect DIF to ensure the assessments are fair and unbiased.

This chapter includes a more detailed description of the methods, describing the findings, and brief discussions of Phase 2 implications and limitations. The chapter starts with a summary of MI in available measures at the construct level across the child's gender and ethnicity, before moving into a detailed assessment of the DIF effects in the latent profile indicators within the LPA model. In addition to assessing the MI of indicators in the context of the latent model of child development, the assessment of DIF effects in the latent profile indicators allows for the examination of MI of the indicators that could not be evaluated at the construct level, specifically, receptive language measured by PPVT-III and executive function assessed by the Luria Hand Clap task in the current research.

6.1 Summary of MI in Available Measures Across the Child's Gender and Ethnic Groups

As discussed in Chapter 4, MI at the construct level was tested for three of the available measures used as indicators of young children's social, emotional and language capabilities in the current research. **Appendices 3-5** provide detailed results of MI tests using MG-CFA, where applicable (i.e., prosocial behaviours and emotional symptoms in the SDQ, AKT, and PROL), across the child's gender and ethnicity.

Appendices 6 and 7 presented reports for PPVT-III and Luria Hand Clap tasks, for which MI could not be assessed at the construct level. Each appendix included introductions of the measurement, the purposes of the assessment of each measure in

current research, descriptions of the methods, and detailed results and conclusions.

Results of the MI tests of the available measures have supported the evaluation of MI in the person-centred model by validating the constructs of each measure and providing an initial assessment of potential sources of MI in the mixture model.

Results showed that across the child's gender groups, partial MI of the prosocial behaviours and emotional symptoms factors of the SDQ (Appendix 3), full invariance of a one-factor model of the AKT (specifying a single factor model with an error covariance between children's responses to the 'AKT7' yucky/disgusted face and 'AKT6' scared face, and between the 'AKT6' scared face and 'AKT3' surprised face: Appendix 4), and full invariance of the PROL (Appendix 5) were established. Across ethnicity, partial MI was established for the prosocial behaviours and emotional symptoms factors of the SDQ, the one-factor model of the AKT, and the PROL. Partial or full MI indicated that these measures displayed reasonable consistency in measuring constructs across the child's gender and ethnic groups. However, the restrictive nature of MG-CFA may be overly sensitive to trivial non-invariance, potentially identifying insignificant differences as meaningful (Khademi et al., 2023). This may unnecessarily complicate the interpretation of results. Additionally, due to the restricted availability of the PPVT-III and the count data based nature of the Luria Hand Clap, it was not feasible to employ MG-CFA to assess MI for these two indicators. To supplement these findings, MI of the broader 3-group model was appropriately assessed across the child's gender and ethnic groups. This remained viable even with partial invariance identified, as further

elaborated in the subsequent sections.

6.2 Testing MI of the Person-Centred Model Across the Child's Gender and Ethnic

Groups

Testing MI of the latent group solutions across observed groups is required when the research aims to determine whether these observed groups exhibit the same set of latent group solutions (Collins & Lanza, 2010; Finch, 2015). Once MI at the model level was established, the generated latent group solutions were understood equivalently across the observed groups. As such, researchers can be justified in making generalisations and comparisons across the groups in terms of the patterns and interpretation of the latent group solutions (Olivera-Aguilar & Rikoon, 2018).

Compared to assessing MI for the measures, evaluating it in the person-centred model offers distinct advantages. It allows the identification of potential impacts of MI on the overall conclusions drawn from the model. Additionally, it enables the assessment of MI in cases where individual indicators could not be examined for the measures within the current research (i.e., PPVT-III, Luria Hand Clap task). This approach provides a broader, more flexible framework for ensuring valid and reliable comparisons across observed groups.

Aligning with the literature on the comparability of the person-centred model in Chapter 3, ways of examining MI in the person-centred model across observed subgroups vary. Similarly to the approach to assess MI at the construct level, researchers may adopt a multi-group approach to the evaluation of MI in latent profile

models, assessing whether performance on the indicators (e.g. means of the latent profiles) is equivalent across groups (Collins & Lanza, 2010; Morin et al., 2016; Olivera-Aguilar & Rikoon, 2018). This approach is particularly applicable where the source of invariance may be considered categorical. Furthermore, while MG-CFA models are constrained to the use of observed categorical sources of MI and require a sufficient sample size in each subgroup, the DIF approach through MIMIC modelling has frequently been used as an alternative method for testing MI, relaxing such limitations. The application of the DIF tests, such as the use of MIMIC modelling, includes covariates as predictors of the latent profile factor, providing insights into potential sources of MI (Masyn, 2017). Each of these approaches is described in more detail.

One of the approaches to examining MI in the person-centred model is the multiple-group approach (Collins & Lanza, 2010; Morin et al., 2016; Olivera-Aguilar & Rikoon, 2018). Collins and Lanza (2010) recommended estimating a separate model for each group as the simplest way to conduct a multiple-group analysis. Morin et al. (2016) proposed a three-step process for comparing group similarities and differences in latent profile analysis (LPA). This involves first ensuring the identical number of latent profiles in each group, then testing the means and variances of continuous profile indicators within a specific profile, and finally confirming the consistency of latent profiles across groups. However, most studies using the multiple-group approach focus on adult participants in U.S. contexts and deficit-based characteristics (Brooks et al., 2017; Chung & Anthony, 2013; Min et al., 2023). Also, limitations of the multiple-group

approach include issues common in MG-CFA. These include MI of the indicators that cannot be tested at the construct level, which may impact interpreting MI when full MI is not achieved, as well as model estimation. To address these challenges, the MIMIC approach has been proposed as an alternative methodological solution.

Testing covariates as potential sources of DIF through MIMIC modelling can help validate the construct underlying the extracted latent group solutions and provide deeper insights into their characteristics. Masyn (2017) suggested using the latent class regression model, where the direct effects from latent class variable predictors to latent class indicators are considered in mixture models. As a part of the latent class model, these effects were examined using the MIMIC modelling technique, recommended as the 7-step procedure extending the 3-step approach from Masyn (2017). The 3-step approach involves identifying latent classes in an unconditional model using observed latent class indicators without incorporating covariates, estimating classification errors, and using latent group membership as an outcome variable to be predicted by covariates. The presence of the association represents that the covariates significantly predict the likelihood of individuals being assigned to a specific latent profile, indicating the differences in the profile compared to the reference group (Lanza et al., 2013). Conversely, the absence of association implies that the covariates do not significantly impact latent group membership, with no meaningful differences between the profile and the referral group (Collins & Lanza, 2010).

The 3-step approach has been critiqued for its limitations in ensuring robustness,

particularly in mitigating misspecifications during the latent class enumeration process and producing accurate estimates of the covariates' effects on latent class membership (Kim et al., 2016; Masyn, 2017). Specifically, the approach assumes that the direct effects representing the DIF effects from covariates to latent profile factors are null. This may result in a poorly fitting latent class regression model, thereby raising concerns about its adequacy in capturing complex associations within the data (Nylund-Gibson & Masyn, 2016). To support the detection of direct effects of covariates on latent profile factors in mixture models, Masyn (2017) developed the MIMIC approach to detecting and accounting for DIF effects through a 7-step procedure. These tests would help examine the covariates as potential sources of DIF and account for any significant direct effects of those covariates on the indicators within the latent group solutions (Masyn, 2017). Therefore, the current study employed the MIMIC modelling technique to assess MI within the person-centred framework to enhance the robustness and validity of the analysis.

6.3 Analysis

Covariate analyses of the optimal 3-group model as described by Masyn (2017), were first evaluated using Mplus 8.10 to indicate proportional differences in covariates (i.e., the child's gender and ethnic groups) between the profiles, followed by the MIMIC approach to test the direct effects of covariates on profile indicators. The child's gender and ethnicity were used as covariates of the 3-group model. A significant direct effect of the child's gender or ethnicity on profile membership suggests notable differences in

the likelihood of different groups being associated with different profiles. The magnitude of these differences can be assessed using the Odds Ratio (OR) in the covariate analyses, which quantifies the association between predictor and outcome variables and compares the relative odds of group membership to the largest group (Moser, 2010; Rudas, 1998). An OR of 1.0 suggests no effect, values > 1.0 indicate increased odds of the effects, and values < 1.0 represent decreased odds. However, the observed difference may be attributed to DIF, necessitating further investigations of DIF effects through MIMIC modelling. In the current work, the DIF effects were assessed using standardised coefficients for continuous factor scores, and unstandardised coefficients for count-based scores. According to (Cohen, 1988; Fey et al., 2023), effect sizes are: $|\beta| < 0.20$ (small effect), $0.2 \leq |\beta| < 0.50$ (medium effect), $|\beta| \geq 0.50$ (large effect). A significant direct effect of either the child's gender or ethnicity on the factor suggests a degree of the DIF effects associated with that factor, indicating the implications should be explored within the context of the broader LPA model. Non-significant direct effects may indicate MI, suggesting the model is equally valid across the sample and allowing meaningful interpretation of group differences (Harring & Hodis, 2016; Masyn, 2017). The use of the two key analytical techniques in the thesis is detailed below.

The covariate analyses tested the direct effects from the covariates to the 3-group model to assess proportional differences between the child's gender and ethnic groups across the profiles through the 3-step approach. The unconditional 3-group model was initially estimated based on the preceding enumeration process. Individuals were then

classified into the most likely class membership and the classification error rates for the model class assignments were computed. Lastly, the child's gender and ethnicity were used as covariates to examine the direct effects on the optional 3-group model. Here, variables representing the child's gender and ethnicity were dummy-coded before assessing the association of gender and ethnicity with each of the 3 latent profiles using multinomial regression. The 3-group model was specified as a nominal variable with 3 categories, class-specific measurement parameters were fixed for the nominal variable indicator to reflect the classification error rates, and regressions of the 3 categories on covariates (i.e., the child's gender, ethnicity) were estimated.

The second part was included to evaluate the direct covariate effects on the latent profile factors (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, receptive language and executive function). This part addresses the limitations in the previous covariate analyses, which did not account for the DIF effects across profiles and ignored the direct covariate effects on the latent profile factors. Specifically, the child's gender and ethnicity were analysed separately as potential sources of DIF to latent profile factors of the 3-group model. A detailed description of assessing MI and DIF effects can be found in the supplemental material of the Masyn (2017) article. For each covariate, seven steps in the thesis were tested as outlined below.

Step 1 started with an omnibus test comparing a null model (no-DIF model, M1.0) with no association of the latent profile factors with the covariates (the child's gender or

ethnicity) and an alternative model with class-specific direct effects on the covariates (nonuniform DIF, M1.1). Based on the 3-group unconditional variances-constrained model (the optimal group solution in Phase 1), the null model M1.0 estimated a complete invariance model, indicating MI across the levels of the covariates. M1.0 included regression of the 3-group model on the covariates (i.e., gender or ethnicity), assuming these covariates impact the latent profiles but have no direct effects on any latent profile factors. The alternative model was then tested, representing the nonuniform DIF model (M1.1). M1.1 added class-specific direct effects from the covariates to every latent profile factor, indicating that DIF effects differ for one or more profiles. The likelihood ratio of the omnibus test (M1.0 vs. M1.1) was subsequently compared. When the omnibus test was rejected, indicating there was evidence of DIF for at least one factor, and the covariates are sources of the DIF effect for at least one factor in at least one of the profiles. Then, a factor-by-factor sequential test proceeded to identify which of the factors functioned differently until a model with DIF that was not statistically significantly worse fitting than the full DIF model.

Step 2 tested each factor for nonuniform DIF (i.e., DIF effects are different for each latent profile), assessing class-specific direct effects from the covariates to each model factor. This step tested each factor for nonuniform DIF with a stepwise procedure.

Based on the 3-group unconditional variance-constrained model and the estimated average classification errors for the class assignment. Twelve models were tested and compared separately for gender and ethnicity.

Step 3 estimated all statistically significant nonuniform DIF effects found in Step 2. Step 4 tested nonuniform DIF versus uniform DIF for each factor identified in Step 2. Step 5 constrained all the nonuniform direct effects from Step 3 and not significantly nonuniform in Step 4 to uniform direct effects. Step 6 involved the evaluation of the practical and substantive impact of the DIF specified in the optimal model in Step 5. Step 7 involved an omnibus test of the association between the covariates and the latent profile. It compared the optimal model from Step 5 and the MIMIC model in Step 7. This step displayed the results of the omnibus test of the association between the covariates and the group membership. If the MIMIC model was significantly worse than the optimal model from Step 5, it indicated that the covariates were related to the group membership.

6.4 Results

6.4.1 Multivariate Analysis of the Optimal 3-Group Model, When Not Accounting for the DIF Effects in Model Indicators

This section describes the results of assessing associations of gender and ethnicity with group membership using multinomial regression without accounting for the DIF effects in factors of latent group membership. Full results are presented in Table 6-1. The *Competent* Profile (Profile 1, above the means), boys and European children were used as the reference groups for gender and ethnicity, respectively. Results show significant direct effects of the child's gender on the *Emerging* profile (Profile 2, below the means) and the *Progressing* profile (Profile 3, close to the means). The four

ethnicities (Māori, Pacific, Asian, and Other ethnicities) had significant direct effects on the *Emerging* and *Progressing* profiles. These findings suggested that the child's gender and ethnicity were important factors in determining the likelihood of being assigned to the *Emerging* and *Progressing* profiles. For example, gender significantly impacted the profile membership, indicating that girls were 0.56 times as likely as boys to be assigned to the *Emerging* profile and 0.77 times as likely as boys to be assigned to the *Progressing* profile.

Table 6-1

Models Assessing Associations of Gender and Five Ethnic Groups With Group Membership From Multinomial Regression

Covariate	Observed grouping variables	Latent group solutions		
		Competent Profile (Profile 1, 45.8%) REF	Emerging Profile (Profile 2, 9.1%) OR (95% CI)	Progressing Profile (Profile 3, 45.1%) OR (95% CI)
Gender	Boy (REF)	-	-	-
	Girl	-	0.56 (0.45, 0.71)	0.77 (0.65, 0.90)
Ethnicity	European (REF)	-	-	-
	Māori	-	3.39 (2.34, 4.91)	2.43 (1.72, 3.44)
	Pacific	-	86.53 (26.81, 279.32)	24.02 (7.00, 82.48)
	Asian	-	7.72 (5.03, 11.85)	5.72 (3.70, 8.84)
	Other ethnicities	-	0.39 (0.24, 0.64)	0.65 (0.50, 0.86)

Note. OR values > 1.0 indicate increased odds, and values < 1.0 represent decreased odds. Ethnicity was prioritised from single/combo prioritisation, with the dual ethnicities group and single-ethnicity group prioritised as Māori > Pacific Peoples > Asian > Other ethnic groups except for European > European, detailed report in **Appendix 2**. Covariate analyses for the child's gender and ethnicity were performed using Mplus v10.0 AUXILLARY multinomial logistic regression procedure in which auxiliary variables do not influence group formation during LPA analyses. Bolded estimates represented a significant direct effect of covariates on the latent group membership. REF represented that the *Competent* profile, boys, and European children were used as the reference groups in the multinomial regression. Gender variables were dummy-coded (boys = 0 and girls = 1). Ethnicity variables were dummy-coded and saved as a new variable, for instance, Māori = 1 and non-Māori = 0; this recode also applies to the European, Pacific, Asian, and Other ethnicities groups.

However, these effects did not account for the potential sources of the DIF effects

on the latent profile factors, which were further assessed using the 7-step approach with the results of the direct covariate effects on each of the latent profile factors. These additional tests were completed by testing the child's gender and ethnicity as separate sources of MNI.

6.4.2 Testing Gender as a Source of the DIF Effects in the Latent Profile Model

Table 6-2 presents model comparisons of DIF tests accounting for the child's gender as a source of latent profile factors representing young children's social, emotional, language, and executive function capabilities. The analyses in Table 6-2 were conducted following the 7-step approach by Masyn (2017) and Bettencourt et al. (2022). The results were interpreted as follows:

Step 1 compared the null, no-DIF model (M1.0), which included the regression of the 3-group model on gender ($LL_0 = -37411.81$, $npar_0 = 28$), to the alternative model, nonuniform DIF model (M1.1) that added class-specific direct effects of gender on each latent profile factor ($LL_1 = -37304.33$, $npar_1 = 46$). The likelihood ratio test of the omnibus test (M1.0 vs. M1.1) showed a rejection of M1.0 ($LRTS = 214.95$, $df = 18$, $p < .001$). This suggested that gender was a source of the DIF effects for at least one of the six latent profile factors in at least one of the three latent profiles. Therefore, Step 2 was proceeded.

Step 2 identified which of the six factors function differently (nonuniform DIF effects) for boys and girls. This step compared no-DIF models to nonuniform DIF models for each of the six factors, and results showed that no-DIF models for executive

function (M2.0.1, LRTS=7.74, $df = 3$, $p > .05$), emotional knowledge (M2.0.4, LRTS=3.83, $df = 3$, $p > .05$), expressive language (M2.0.5, LRTS=6.13, $df = 3$, $p > .05$), and receptive language (M2.0.6, LRTS=5.73, $df = 3$, $p > .05$) were not significantly worse than the models allowing for nonuniform DIF. No-DIF models were significantly worse than nonuniform DIF models for prosocial behaviours (LRTS=135.1, $df = 3$, $p = .00$) and emotional symptoms (LRTS=34.52, $df = 3$, $p = .00$). Hence, models without DIF were rejected for prosocial behaviours and emotional symptoms, which suggested these models may function differently for each latent profile with respect to gender.

Step 3 tested prosocial behaviours and emotional symptoms with the nonuniform DIF effects found in Step 2. M3.0, including the two nonuniform DIF effects from gender to prosocial behaviours and emotional symptoms, was compared to M1.0 and M1.1, respectively. Results showed that M3.0 is significantly better than M1.0 (LRTS=173.92, $df = 6$, $p < .01$) but significantly worse than M1.1 from Step 1 (LRTS=41.03, $df = 12$, $p < .01$). There was a significant improvement in model fit from model M1.0 to M3.0, but M1.1 was significantly better than M3.0. M3.0.1 to M3.0.4, then compared models with nonuniform DIF to models with and without uniform DIF. In addition to nonuniform DIF for prosocial behaviours and emotional symptoms, M3.0.1 to M3.0.4 sequentially added the nonuniform indicator executive function to M3.0.1, emotional knowledge to M3.0.2, expressive language to M3.0.3 and receptive language to M3.0.4. M3.0.1 to M3.0.4 were compared to M3.0 and M1.1. Results showed that M3.0 was significantly worse than M3.0.1, M3.0.3 and M3.0.4, but not

significantly worse than M3.0.2. M1.1 was significantly better than these four models. Therefore, M1.1 was retained as the final model to proceed to Step 4.

Step 4 tested models with and without uniform DIF. Each of the six factors was estimated in six additional models. In M4.1.1, prosocial behaviours were specified as class-invariant, while the remaining five indicators were allowed to be class-varying. The same specification was applied sequentially in M4.1.2 to M4.1.6, such that each indicator—emotional symptoms, emotional knowledge, expressive language, receptive language, and executive function—was specified to be class-invariant in turn, with the other five indicators specified as class-varying. These models were then compared to M1.1, which added class-varying direct effects from gender to each of the six factors. The results indicated that M4.1.1 to M4.1.4 did not fit significantly worse than M1.1, whereas M4.1.5 and M4.1.6 showed significantly poorer fit. The DIF effects of gender on receptive language and executive function are nonuniform DIFs; prosocial behaviours, emotional symptoms, emotional knowledge, and expressive language are uniform DIFs.

Based on results from previous steps, Step 5 specified the DIF effects of gender with nonuniform DIFs for receptive language and executive function, uniform DIFs for prosocial behaviours, emotional symptoms, emotional knowledge, and expressive language. M5.1 was compared with M1.1 in Step 5. Results found that M5.1, with the uniform and nonuniform DIFs constraints imposed, was not statistically worse fitting than M1.1. Hence, M5.1 was then used for Step 6 evaluation.

Figure 6-1 presented model-estimated latent profile indicator distribution by gender in Step 6, which assessed the practical and substantive impact of the DIF identified in M5.1. This step ultimately determined whether sufficient invariance existed across gender groups, allowing a valid comparison of the distribution of the latent group membership among boys and girls. The estimated standardised coefficients of the uniform DIF effects of gender on prosocial behaviours ($\beta= 0.17$), emotional symptoms ($\beta= 0.05$), and expressive language ($\beta= 0.07$) were significant ($p < .05$) across the three latent profiles. This means gender has a small effect on prosocial behaviours and considerably negligible effects on emotional symptoms and expressive language, with girls having higher scores on these three factors compared to boys. Uniform DIF effects from gender on emotional knowledge were non-significant. The nonuniform DIF effect of gender on receptive language was not significant ($p > .05$) across the three latent profiles, meaning gender is not significantly related to receptive language in the model. The unstandardised coefficients were statistically significant ($p < .05$) for both the *Competent* profile (unstandardised estimate = -0.33, $SE=0.15$) and the *Progressing* profile (unstandardised estimate = -0.10, $SE=0.04$). These findings may suggest a medium effect of gender on estimating the count-based scores for executive function in the *Competent* profile and a small effect in the *Progressing* profile. Visual inspection of the 3-group model across gender in Figure 6-1 suggested that the 3-group model was substantively comparable for boys and girls. The nonuniform DIF effects did not shift the separation or interpretation of the latent group solutions for boys and girls.

Based on the estimated DIF effects and the evaluation of the gender-specific distribution in Figure 6-1, I concluded that the 3-group model was comparable for boys and girls. Hence, Step 7 compared the distribution of the 3-group model across the child's gender groups.

Step 7 involved an omnibus test of the association between gender and latent group membership. This was achieved by comparing the final model from Step 5, relabelled as M7.1, where the multinomial logistic slope coefficients for the regression of the latent group membership on gender were freely estimated. This model was compared to M7.0 with multinomial logistic slope coefficients for the regression of the latent group membership solution on gender fixed at 0. M7.1 had a significantly better fit than M7.0, indicating gender was associated with latent group membership.

Table 6-2*Model Comparison for Stepwise DIF Tests for the Child's Gender With the 3-Group Latent Profile Model of Young Children's Capabilities*

Steps	Models	Model description	LL	#free parameters	Model comparison	LRTS	df _{LRTS}	p	
1	M1.0	No DIF	-37411.806	28					
	M1.1	All DIF	-37304.331	46	M1.0 vs M1.1	214.95	18	0.000	
2	M2.0.1	EF: No DIF	-19024.686	8					
	M2.1.1	EF: Nonuniform DIF	-19020.814	11	M2.0.1 vs M2.1.1	7.744	3	0.052	
	M2.0.2	Pro: No DIF	-9688.119	8					
	M2.1.2	Pro: Nonuniform DIF	-9620.569	11	M2.0.2 vs M2.1.2	135.1	3	0.000	
	M2.0.3	EM: No DIF	-8268.668	8					
	M2.1.3	EM: Nonuniform DIF	-8251.408	11	M2.0.3 vs M2.1.3	34.52	3	0.000	
	M2.0.4	EK: No DIF	-9270.881	8					
	M2.1.4	EK: Nonuniform DIF	-9268.964	11	M2.0.4 vs M2.1.4	3.834	3	0.280	
	M2.0.5	EL: No DIF	-8514.682	8					
	M2.1.5	EL: Nonuniform DIF	-8511.615	11	M2.0.5 vs M2.1.5	6.134	3	0.105	
	M2.0.6	RL: No DIF	-10570.508	8					
	M2.1.6	RL: Nonuniform DIF	-10567.644	11	M2.0.6 vs M2.1.6	5.728	3	0.126	
	3	M3.0	Pro and EM with nonuniform DIF	-37324.844	34	M1.0 vs M3.0	173.924	6	0.000
						M3.0 vs M1.1	41.026	12	0.000
M3.0.1		nonuniform DIF: EF, Pro, EM symptoms	-37320.367	37	M3.0 vs M3.0.1	8.954	3	0.030	
					M3.0.1 vs M1.1	32.072	9	0.000	
M3.0.2		nonuniform DIF: EK, Pro, EM symptoms	-37323.633	37	M3.0 vs M3.0.2	2.422	3	0.490	

Steps	Models	Model description	LL	#free parameters	Model comparison	LRTS	df _{LRTS}	p
					M3.0.2 vs M1.1	38.604	9	0.000
	M3.0.3	nonuniform DIF: EL, Pro, EM symptoms	-37317.128	37	M3.0 vs M3.0.3	15.432	3	0.001
					M3.0.3 vs M1.1	25.594	9	0.002
	M3.0.4	nonuniform DIF: RL, Pro, EM symptoms	-37318.619	37	M3.0 vs M3.0.4	12.45	3	0.006
					M3.0.4 vs M1.1	28.576	9	0.001
4	M4.1.1	all items nonuniform except Pro	-37304.415	44	M4.1.1 vs M1.1	0.168	2	0.919
	M4.1.2	all items nonuniform except EM	-37306.008	44	M4.1.2 vs M1.1	3.354	2	0.187
	M4.1.3	all items nonuniform except EK	-37305.797	44	M4.1.3 vs M1.1	2.932	2	0.231
	M4.1.4	all items nonuniform except EL	-37306.577	44	M4.1.4 vs M1.1	4.492	2	0.106
	M4.1.5	all items nonuniform except RL	-37310.191	44	M4.1.5 vs M1.1	11.72	2	0.003
	M4.1.6	all items nonuniform except EF	-37308.477	44	M4.1.6 vs M1.1	8.292	2	0.016
5	M5.1	nonuniform DIF for RL and EF, and uniform DIF for Pro, EM, EK, EL	-37310.199	38	M5.1 vs M1.1	11.736	8	0.163
6		Evaluation of the practical and substantive impact of DIF in M5.1 (Figure 6-1)						
7	M7.0	C on gender @0	-37314.411	36				
	M7.1	C on gender free, same as M5.1	-37310.199	38	M7.0 vs M7.1	8.424	2	0.015

Note. Pro = prosocial behaviours, EM= emotional symptoms, EK= emotional knowledge, EL= expressive language, RL= receptive language, EF= Executive function. LL= Log Likelihood. npar= Number of free parameters estimated in the model. LRTS = likelihood ratio test statistic. df= degrees of freedom. Nonuniform DIF indicates that DIF effects are different for one or more latent profiles. Uniform DIF indicates DIF effects are the same across the latent profiles. No DIF indicates there is no DIF effect.

Figure 6-1

Model-Estimated Latent Profile Factor Distribution by Gender for Factors With DIF Effects



Note. N = 5704, 3 cases missing on gender were not estimated. The figure showed the model-estimated mean of the five-factor scores (prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, and receptive language). Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. The figure also showed the median count scores of the executive function relative to its grand median count scores (13) in Chapter 5.

Figure 6-1 showed the model-estimated latent profile factor means and median distribution by gender. Results suggested that the child's gender was related to latent group membership. Gender was a potential source of the DIF effects for at least one of the latent profile factors. Specifically, uniform DIFs were identified for prosocial behaviours, emotional symptoms, expressive language, and emotional knowledge, with nonuniform DIFs identified for receptive language and executive function. As noted in Table 6-2, significant uniform DIF effects were observed for prosocial behaviours, emotional symptoms, and expressive language across the three latent profiles, with gender having a small effect on prosocial behaviours and negligible effects on emotional symptoms and expressive language. Uniform DIF effects from gender on emotional knowledge were not significant. Nonuniform DIFs were identified for receptive language and executive function, suggesting gender was a source of nonuniform DIF effects for these two factors. The nonuniform DIF effects for receptive language were not significant. They were only significant for executive function in both the *Competent* and *Progressing* profiles, with a medium effect in the *Competent* profile and a small effect in the *Progressing* profile identified. In sum, despite the presence of both uniform and nonuniform DIF effects, the magnitude of these effects was small or negligible, providing sufficient invariance to support the comparability of the 3-group model for the boys and girls.

6.4.3 Testing Ethnicity as a Source of the DIF Effects in the Latent Profile Model

Table 6-3 presented model comparisons for stepwise DIF testing for ethnicity as a source of latent profile factors representing young children's social, emotional, language and executive function capabilities. The analyses in Table 6-3 were conducted following the 7-step approach by Masyn (2017) and Bettencourt et al. (2022). The results were interpreted as follows:

Step 1 compared the null model (M1.0), which included regression of the 3-group model on ethnicity ($LL_0 = -36882.86$, $npar_0 = 34$), to the alternative nonuniform DIF model (M1.1), which added class-specific direct effects on each latent factor regarding ethnicity ($LL_1 = -36525.34$, $npar_1 = 106$). The likelihood ratio test of the omnibus test (M1.0 vs. M1.1) showed a rejection of M1.0 (LRTS = 715.04, $df = 72$, $p < .001$), suggesting that ethnicity is a source of the DIF effects for at least one of the six latent profile indicators in at least one of the three latent profiles.

Step 2 assessed how these six factors function differently across different ethnic groups. This step compared no-DIF models to nonuniform DIF models for each factor. Results showed that no-DIF models for executive function (LRTS = 194.36, $df = 12$, $p < .01$), prosocial behaviours (LRTS = 72.89, $df = 12$, $p < .01$), emotional symptoms (LRTS = 104.22, $df = 12$, $p < .01$), emotional knowledge (LRTS = 29.28, $df = 12$, $p < .01$), expressive language (LRTS = 85.48, $df = 12$, $p < .01$), and receptive language (LRTS = 164.00, $df = 12$, $p < .01$) were significantly worse than the models allowing for nonuniform DIF. Hence, no-DIF models were rejected for all six factors, which might

function differently with respect to ethnicity.

Based on the results from Step 2, Step 3 tested a model that included all six factors with nonuniform DIF effects (M3.0), and six models sequentially assessed one of the six factors with no-DIF effects while specifying the remaining five factors with nonuniform DIF effects (M3.0.1 to M3.0.6). M3.0, including the nonuniform DIF effects from ethnicity to the six factors, was the same as M1.1, with all DIF indicating class-varying direct effects from ethnicity to each of the six factors. M3.0 was then compared to M1.0 (repeating the omnibus test of model invariance, M1.0 vs M1.1). The results indicated that M3.0 provided a significantly better fit to the data than M1.0 (LRTS=715.04, df =72, $p < .01$). M3.0.1 to M3.0.6 were then tested and compared to M3.0. Results indicated that M3.0.1 (model with all nonuniform DIF factors, except for executive function with no-DIF effects; LRTS=20.17, df =12, $p = .06$) and M3.0.2 (model with all nonuniform DIF factors, except for prosocial behaviours with no-DIF effects; LRTS=18.34, df =12, $p = .11$) displayed an improved model fit than M3.0. M3.0.3 through M3.0.6 had significantly worse fitting than M3.0. Based on the results from these steps, an additional model (M3.0.7) was tested. This model specified executive function and prosocial behaviours with no-DIF effects, while the remaining four factors were modelled with nonuniform effects. M3.0.7 showed improvements in model fit compared to M3.0.1 (LRTS=19.28, df =12, $p = .08$), but was significantly worse than M3.0 (LRTS=39.44, df =24, $p < .05$) and M3.0.2 (LRTS=21.10, df =12, $p < .05$). Therefore, M3.0.2 was retained as the final MIMIC model to proceed to Step 4.

Step 4 tested models for uniform DIF. Following M3.0.2, each of the factors (executive function, emotional symptoms, emotional knowledge, expressive language, and receptive language) identified in Steps 2 and 3 was estimated in five additional models: M4.1.1 and M4.3.1 to M4.6.1. These models sequentially specified all factors with nonuniform DIF, except uniform DIF for executive function, emotional symptoms, emotional knowledge, expressive language, and receptive language and all with no-DIF prosocial behaviours. Each of the models (M4.1.1, M4.3.1 through M4.6.1) was then compared to M3.0.2. The results showed that M4.3.1 provided an improved model fit compared to M3.0.2 (LRTS=14.31, df =8, p= .07), the other four models had significantly worse fitting than M3.0.2. Hence, M4.3.1 was retained as the model for Step 5.

Step 5 specified a model M5.0.1, the same as M4.3.1 with no-DIF prosocial behaviours, uniform emotional symptoms, and nonuniform DIFs for executive function, emotional knowledge, expressive language, and receptive language. This model was retained as the model for Step 6, evaluating the practical and substantive impact of DIF. Figure 6-2 presents the model-estimated distribution of latent profile factors by ethnicity in Step 6. This step ultimately evaluated whether sufficient invariance existed across ethnicity, allowing for valid comparisons of their latent group membership. Step 7, which involved the comparisons of the distribution of the 3-group model for ethnicity, did not proceed due to insufficient invariance for factors.

Table 6-3*Model Comparison for Stepwise DIF Testing for Ethnicity With the 3-Group Latent Profile Model of Young Children's Capabilities*

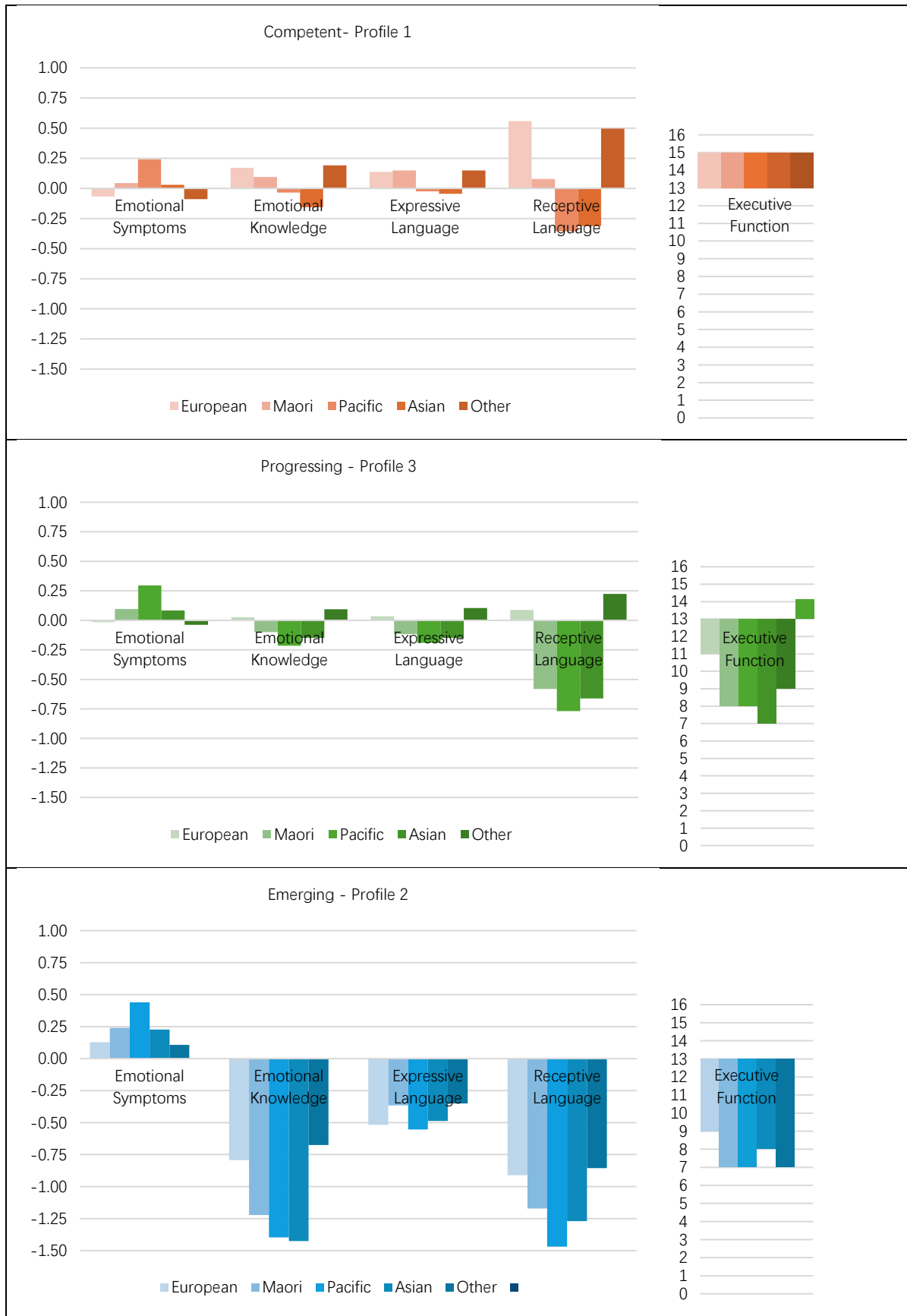
steps	Models	Model description	LL	#free parameters	Model comparison	LRTS	df _{LRTS}	p
1	M1.0	No DIF	-36882.855	34				
	M1.1	All DIF	-36525.336	106	M1.0 vs M1.1	715.038	72	.000
	M2.0.1	EF: No DIF	-18731.954	14				
	M2.1.1	EF: Nonuniform DIF	-18634.773	26	M2.0.1 vs M2.1.1	194.362	12	.000
	M2.0.2	Pro: No DIF	-9282.461	14				
	M2.1.2	Pro: Nonuniform DIF	-9246.017	26	M2.0.2 vs M2.1.2	72.888	12	.000
	M2.0.3	EM: No DIF	-7786.158	14				
2	M2.1.3	EM: Nonuniform DIF	-7734.049	26	M2.0.3 vs M2.1.3	104.218	12	.000
	M2.0.4	EK: No DIF	-8858.487	14				
	M2.1.4	EK: Nonuniform DIF	-8843.849	26	M2.0.4 vs M2.1.4	29.276	12	.004
	M2.0.5	EL: No DIF	-8130.366	14				
	M2.1.5	EL: Nonuniform DIF	-8087.626	26	M2.0.5 vs M2.1.5	85.480	12	.000
	M2.0.6	RL: No DIF	-9995.703	14				
	M2.1.6	RL: Nonuniform DIF	-9913.703	26	M2.0.6 vs M2.1.6	164.000	12	.000
	M3.0	All nonuniform DIF (same as M1.1)	-36525.336	106	M1.0 vs M3.0	715.038	72	.000
	M3.0.1	All nonuniform except no-DIF EF	-36535.419	94	M3.0.1 vs M3.0	20.166	12	.064
3	M3.0.2	All nonuniform except no-DIF Pro	-36534.508	94	M3.0.2 vs M3.0	18.344	12	.106
	M3.0.3	All nonuniform except no-DIF EM	-36674.166	94	M3.0.3 vs M3.0	297.660	12	.000
	M3.0.4	All nonuniform except no-DIF EK	-36578.995	94	M3.0.4 vs M3.0	107.318	12	.000

steps	Models	Model description	LL	#free parameters	Model comparison	LRTS	df _{LRTS}	p
	M3.0.5	All nonuniform except no-DIF EL	-36584.288	94	M3.0.5 vs M3.0	117.904	12	.000
	M3.0.6	All nonuniform except no-DIF RL	-36680.028	94	M3.0.6 vs M3.0	309.384	12	.000
	M3.0.7	All nonuniform except no-DIF for both EF and Pro	-36545.058	82	M3.0.7 vs M3.0	39.444	24	.025
					M3.0.7 vs M3.0.1	19.278	12	.082
					M3.0.7 vs M3.0.2	21.1	12	.049
	M4.1.1	All items nonuniform except EF, and no-DIF Pro	-36542.695	86	M4.1.1 vs M3.0.2	16.374	8	.037
	M4.3.1	All items nonuniform except EM, and no-DIF Pro	-36541.661	86	M4.3.1 vs M3.0.2	14.306	8	.074
4	M4.4.1	All items nonuniform except EK, and no-DIF Pro	-36552.515	86	M4.4.1 vs M3.0.2	36.014	8	.000
	M4.5.1	All items nonuniform except EL, and no-DIF Pro	-36546.240	86	M4.5.1 vs M3.0.2	23.464	8	.003
	M4.6.1	All items nonuniform except RL, and no-DIF Pro	-36550.577	86	M4.6.1 vs M3.0.2	32.138	8	.000
5	M5.0.1	No-DIF Pro, uniform EM, nonuniform DIF for EF, EK, EL and RL (same as M4.3.1)	-36541.661	86	M5.0.1 vs M3.0.2	14.306	8	.074
6		Evaluation of the practical and substantive impact of DIF in M5.0.1 (See Figure 6-2 and Table 6-4)						
7	-	-	-	-	-	-	-	-

Note. Ethnicity was prioritised from single/combination prioritisation, with the dual ethnicities group and single-ethnicity group prioritised as Māori > Pacific Peoples > Asian > Other ethnic groups except for European > European, detailed report in **Appendix 2**. Pro = prosocial behaviours, EM= emotional symptoms, EK= emotional knowledge, EL= expressive language, RL= receptive language, EF= Executive function. LL= Log Likelihood. npar= Number of free parameters estimated in the model. LRTS = likelihood ratio test statistic. df= degrees of freedom. Nonuniform DIF indicates that the DIF effects are different for one or more latent profiles. Uniform DIF indicates that the DIF effects are the same across latent profiles. No DIF indicates there is no MNI effect. Significant effects between models indicate that the baseline model is significantly better than the compared model. Non-significant effects between models indicate the new model is not worse than the baseline model, and the new model should be retained.

Figure 6-2

Model-Estimated Latent Profile Factor Distribution by Ethnicity for Factors With DIF Effects



Note. N = 5704, missing 3 cases on gender variables. No-DIF prosocial behaviours were not graphed in the figure above due to limitations with graphing means of no-DIF factor within mixture models in Mplus output. The figure showed the model-based mean of the four-factor scores (emotional symptoms, emotional knowledge, expressive language, and receptive language) and the median count scores of the executive function. Scores above zero represented scores higher than the average, and scores below zero represented scores lower than the sample average. Executive function was charted as absolute median count scores relative to grand median count scores (13).

The visualisations in Figure 6-2 presented model-estimated means or medians of the latent profile factor distribution by ethnicity. DIF effects were characterised as:

Prosocial behaviours were identified with no-DIF effect, suggesting invariance across ethnicity. Significant differences in means were observed in the *Competent* and *Emerging* profiles. However, the means of prosocial behaviours in each ethnic group across the three profiles were not presented in Figure 6-2.

Significant uniform DIF effects were found for emotional symptoms among Māori, Pacific, and Asian children across the three profiles, with small effect sizes observed for these ethnic groups on emotional symptoms ($|\beta| < 0.20$). Non-significant DIF effects were identified for Children from Other ethnicities. These findings suggested ethnicity had no effect or only small effects on emotional symptoms. Specifically, Māori, Pacific and Asian children had slightly higher factor scores for emotional symptoms compared to the reference group, European children, across the three profiles.

Significant nonuniform DIF effects were identified for emotional knowledge among Pacific and Asian children across the three profiles, and Māori children in the *Emerging* profile. Small effect sizes ($|\beta| < 0.20$) were observed for Māori children in the *Emerging* profile and Pacific and Asian children in the *Competent* and *Progressing* profiles. Medium effect sizes ($0.2 \leq |\beta| < 0.50$) were found for Pacific and Asian children in the

Emerging profile. Non-significant nonuniform DIF effects were observed in emotional knowledge among children from Other ethnicities.

Significant nonuniform DIF effects were observed for expressive language. Pacific and Asian children showed small effect sizes in the *Competent* and *Progressing* profiles. Children from Other ethnicities exhibited a small effect size in the *Emerging* profile, while Māori children had a small effect size in the *Progressing* profile ($|\beta| < 0.20$).

Significant nonuniform DIF effects were identified for receptive language. Pacific children had medium effect sizes across the three profiles ($0.2 \leq |\beta| < 0.50$). Māori children exhibited small effect sizes across the three profiles. Asian children showed small effect sizes in the *Emerging* and *Progressing* profiles ($|\beta| < 0.20$) but a medium effect size in the *Competent* profile ($0.2 \leq |\beta| < 0.50$).

Unstandardised coefficients for executive function were only significant for Asian children in the *Progressing* profile (estimate=0.16, $p < .01$). These coefficients were non-significant in the *Competent* and *Emerging* profiles across the ethnic groups.

Since the DIF effects altered the separation and interpretation of the latent group solution, and there was insufficient full measurement invariance for factors like executive function, emotional knowledge, and expressive and receptive language, I concluded that the 3-group model was not comparable across ethnicity.

6.5 Discussion

The study aimed to explore whether the 3-group model was invariant across the child's binary gender (i.e., boys and girls) and ethnicity (i.e., European, Māori, Pacific

Peoples, Asian, Other) using a DIF approach through MIMIC modelling to invariance testing (Bettencourt et al., 2022; Masyn, 2017). This approach used in the current work accounted for DIF effects on the latent profile factors of young children's social, emotional, language and executive function capabilities, as ignoring these DIF effects may bias estimated parameters for the measure and structural model of the mixture models (Masyn, 2017). This inclusion of covariate effects on the latent profile factors could help identify whether the covariates (i.e., gender and ethnicity) were potential sources of DIF for the latent profile factors and mitigate the risks of introducing biases in the prediction model (Kim & Wang, 2019).

The 7-step MIMIC modelling procedure was used to test gender and ethnicity as separate sources of DIF effects in latent profile factors, with results in Tables 6-2 and 6-3. Table 6-4 summarised findings for uniform and nonuniform DIF effects associated with the six latent profile factors in each gender and ethnicity.

Results suggested that the child's gender was a source of the DIF effects for all six factors. Uniform DIF effects on prosocial behaviours, emotional symptoms, emotional knowledge, and expressive language were identified, indicating that the DIF effects are the same across the three latent profiles. Specifically, a small effect on prosocial behaviours and negligible effects on emotional symptoms and expressive language were found significant across profiles, with non-significant uniform DIF effects on emotional knowledge observed. In contrast, nonuniform DIF effects were observed for receptive language and executive function, indicating the DIF effects are different for at

least one or more of the latent profiles. Significant medium effects in the *Competent* profile and small effects in the *Progressing* profile for executive function were identified, with non-significant nonuniform DIF effects for receptive language observed. The effect sizes, together with visualisations of the model-estimated latent profile factor mean and median distribution by gender, indicated sufficient measurement invariance of the 3-group model across the child's gender groups. This finding is consistent with the hypothesis presented in Chapter 4, which hypothesised that a child's gender (boys and girls) would have no direct effects on the latent developmental profile indicators.

Ethnicity was a source of DIF effects for five latent profile factors in at least one of the three latent profiles, except for prosocial behaviours, with no-DIF effects observed. Specifically, uniform DIF effects with small effect sizes were found in emotional symptoms among Māori, Pacific, and Asian children across the three profiles.

Nonuniform DIF effects on executive function, emotional knowledge, and receptive and expressive language were identified, with non-significant or small to medium effect sizes varying in one or more profiles. For example, significant nonuniform DIF effects for emotional knowledge were identified among Pacific and Asian children across the three profiles and Māori children in the Emerging profile, with small to medium effects. These small to medium effect sizes and visualisations of the model-estimated mean and median distribution of latent profile factors by ethnicity indicated insufficient measurement invariances of the 3-group model across ethnicity. Figure 6-2 visualisations suggested that emotional symptoms may be comparable across ethnicity

across the latent profile. Emotional knowledge, and expressive and receptive language show varying degrees of comparability, with differences specific to individual profiles. Hence, the 3-group model was not comparable across ethnicity, even with the partial DIF effects. These findings are inconsistent with the hypothesis presented in Chapter 4, which posited that the mother-prioritised child's ethnicity (Māori, Pacific, Asian, European, and Other ethnicities) would have no direct effect on the latent developmental profile indicators.

Findings from the study underscore the importance of data due diligence. Ignoring or avoiding DIF effects in mixture models may jeopardise the results since it provides biased estimates for the classification process and the prediction model (Masyn, 2017). Applying the DIF approach was necessary for addressing unbiased estimates, as it identified DIF effects across the child's gender and ethnicity. However, testing the direct effects of covariates on latent profile indicators through MIMIC modelling requires careful consideration, as the approach remains relatively new within mixture modelling and has been supported by a limited body of empirical evidence. The use of the DIF approach in the current work also highlighted the importance of acknowledging the limitations of the data, cautioning against the assumption that all latent profile indicators were equally interpreted across different levels of the child's gender and ethnicity. By examining covariates as potential sources of DIF effects in the current research, this research established MI across gender groups in a 3-group model, allowing for equitable interpretation of indicators, with no comparability across

ethnicity concluded due to the lack of sufficient invariances. These findings reflected the importance of addressing DIF effects to ensure fairness in interpretation and robust results across diverse demographic groups.

Table 6-4

Summary of Findings of Uniform and Nonuniform DIF Effects Using the Child's Gender and Ethnicity as Covariates

Factors	Gender ¹	Profile 1	Profile 2	Profile 3	Ethnicity ²	Profile 1				Profile 2				Profile 3			
						Māori	Pacific	Asian	Other	Māori	Pacific	Asian	Other	Māori	Pacific	Asian	Other
Pro	uniform	0.168	0.166	0.168	no DIF	-	-	-	-	-	-	-	-	-	-	-	-
EM	uniform	0.045	0.045	0.045	uniform	0.068	0.204	0.075	-0.016	0.086	0.313	0.072	-0.018	0.078	0.246	0.065	-0.015
EK	uniform	0.015	0.014	0.015	nonuniform	-0.036	-0.105	-0.195	0.011	-0.231	-0.428	-0.331	0.064	-0.069	-0.151	-0.092	0.036
EL	uniform	0.074	0.073	0.074	nonuniform	0.007	-0.098	-0.130	0.008	0.111	-0.035	0.021	0.125	-0.096	-0.165	-0.112	0.045
RL	nonuniform	-0.016	-0.046	0.092	nonuniform	-0.155	-0.322	-0.359	-0.023	-0.116	-0.327	-0.155	0.025	-0.248	-0.364	-0.265	0.050
EF	nonuniform	-0.332	-0.018	-0.095	nonuniform	0.045	0.007	-0.110	0.091	-0.043	0.051	-0.141	-0.038	0.082	0.086	0.159	-0.038
Conclusion	The results indicated sufficient measurement invariance of the 3-group model across the child's gender groups				Conclusion	even with the partial DIF effects, the results suggested that the 3-group model may not be comparable across the child's ethnicity.											

Note. 1 represents that the group 'boys' was used as the reference group for gender; 2 represents that the group 'European' was used as the reference group for ethnic groups. Pro = prosocial behaviours, EM= emotional symptoms, EK= emotional knowledge, EL= expressive language, RL= receptive language, EF= Executive function. Standardised coefficients were generated for prosocial behaviours, emotional symptoms, emotional knowledge, expressive and receptive language; unstandardised coefficients were used for executive function. Nonuniform DIF indicates that the DIF effects are different in one or more profiles. Uniform DIF indicates that the DIF effects are the same across profiles. No DIF indicates there is no DIF effect. The abbreviation ns indicates not statistically significant.

6.6 Implications

The second phase of the thesis focused on how best to understand measures that address young children's social, emotional, language, and executive function capabilities. Results from the DIF approach using the MIMIC modelling to assess DIF effects could inform fairness in understanding measures and the interpretation of children's capabilities. Accounting for the DIF effects may mitigate the risks of biased estimates of the associations between covariates and latent group solutions. The results of the DIF effects on latent profile factors may inform revisions to the existing factors or the addition of new factors. Specifically, nonuniform DIF was observed for receptive language and executive function by gender, and executive function, emotional knowledge, expressive language, and receptive language by ethnicity. This suggests that factor structure may require adjustment to ensure equivalent measurement across groups.

By accounting for DIF effects on the latent group solution and factors associated with the child's gender and ethnicity, this research can appropriately assess more complex models to understand latent profiles and better inform research and policy. Previous research in Chapter 3 explored gender and ethnic differences in individual domains. However, a holistic approach is needed to deepen the understanding of the complexity and interconnectedness of children's capabilities. The current research integrated observed subgroup variation (i.e., gender and ethnicity) with the person-centred model to ensure a robust understanding of children's capabilities, highlighting

subgroup differences and emphasising the interconnectedness of these capabilities across different subgroup levels.

Acknowledging the complexity of the analytical approaches employed in the thesis is important. The thesis employed the DIF approach, recognising its advantages in assessing DIF effects on the latent profile factors. Specifically, this DIF approach assumed configural invariance across different levels of the covariates in the study, and that general class-specific patterns of item endorsement were consistent across covariate values (Bettencourt et al., 2022). This approach has been easily adapted for different metrics of covariates and latent group indicators (Masyn, 2017). This may necessitate empirical guidelines to judge the magnitude of DIF in the mixture models. Future research may require a thorough and systematic study design and population features that impact the DIF in mixture models with covariates. This may provide an opportunity to consider additional covariates for future studies.

6.7 Limitations and Future Directions

The ethnicity categorisation in the thesis may not provide the most representative depiction of the populations under study. The child's ethnicity was used as a grouping variable to examine DIF effects in the DIF approach and generate results to support the potential utility of education policies for certain ethnicities. Although the research followed the single/combination ethnicity categorisation, including single-ethnicity groups did provide sufficient sample size for latent profile analysis and group comparisons. However, the complexities of ethnicity categorisation posed challenges

for statistical analysis among participants identified with dual- or multiple-ethnicities by their mother. The strengths and limitations of each ethnicity classification method, such as administrative prioritisation, single/combination, and self-prioritisation, were discussed by the GUINZs researchers (Yao et al., 2021; Yao et al., 2022). Further, the current research did not include Dual ethnicities due to the complex nature of Dual ethnicities. Careful considerations for ethnicity categorisation are needed to avoid jeopardising research results.

The thesis did not focus on the interaction effects of the child's gender and ethnicity on the latent profile indicators, as the literature review in the current thesis did not provide sufficient coverage of the interaction effects on children's capabilities according to the research purposes. This limitation can be addressed in future studies by exploring the combined sources of the child's gender and ethnicity of the DIF effects on the latent profile indicators. Given the insufficient invariance across ethnic groups, it would be important for future studies to explore how ethnicity may impact the latent profiles differently across diverse ethnic and other demographic backgrounds. This could help uncover potential ethnic, cultural, or societal factors that contribute to variations in children's capabilities, ultimately improving the generalisability of the findings. Additionally, this research did not explore the impact of other covariates on MI and the DIF effects due to the focus of the research, which may limit the understanding of the factors influencing children's capabilities.

The limitation in reporting the effect sizes may arise from an inconsistency in using

standardised and unstandardised coefficients as metrics for evaluating effect sizes. As the Mplus software does not generate standardised coefficients for count-based data in the MIMIC model in the current research, the unstandardised coefficients were used to assess the nonuniform DIF effect of gender on executive function. This may impact the interpretability of the results, whereas it addresses the appropriate data modelling and the research question (Pek & Flora, 2018).

The DIF approach is a starting point for methodological research related to MI in mixture modelling (Masyn, 2017). Limitations in research design and the selection of covariates may provide a limited understanding of the distal outcomes influencing the mixture models. For instance, literature suggested changes in capabilities over time (Collins & Lanza, 2010), while the current study was a cross-sectional study that was unable to examine how DIF effects may impact the measures and latent group solutions across the child's gender and ethnic groups over time.

6.8 Conclusion

Findings from the DIF approach concluded that the 3-group model demonstrated comparability across the child's gender groups. However, insufficient invariance observed across ethnicity limited the comparability of the 3-group model. The findings offer a strong step forward and an example for future studies in understanding young children's social, emotional, language, and executive function capabilities with advanced statistical analysis techniques. By exploring MI across diverse subpopulations and DIF effects using MG-CFA and the MIMIC modelling techniques, the thesis

emphasised the data due diligence, fairness in interpreting the constructs and model, and using data with care. The current work is important for the methodological exploration of reducing biases by incorporating DIF effects into the mixture models of young children's social, emotional, language, and executive function capabilities.

Chapter 7 General Discussion

The current chapter aims to provide a thorough synthesis of key findings in the current thesis, with critical discussions on both the consistencies and discrepancies with existing literature. It begins by presenting an overview of significant findings on the shared patterns of social, emotional, language, and executive function capabilities in young children. It then summarises the results of measurement invariance (MI) tests conducted for the available selected measures across the child's gender and ethnicity. Furthermore, it includes a summary of MI tests at the person-centred model level and the comparability of the optimal group solution across the child's gender and ethnicity. Findings are then discussed in an overall discussion section with a focus on the consistencies and variations in person-centred studies using the GUiNZs 54-month dataset. Key discussion points centre on the importance of holistic, strengths-based, and culturally connected ways to understand young children's capabilities and fairness in assessment.

Following the general discussion, the broader implications of the present research will address methodological, practical, research, and policy-related considerations. This includes the advantages of person-centred research compared to variable-centred studies to better understand children's capabilities from a holistic lens. A closer look at fairness in assessment and the role of measurement invariance testing is highlighted as a way to test and mitigate potential bias. Findings may enhance the appropriate use of data in understanding children's capabilities in early childhood practice, research, and

policy contexts, which contributes to equitable access to high-quality services in early childhood education in New Zealand. Limitations of the study are acknowledged, and recommendations for future research are proposed. This chapter concludes with a concise summary of the current thesis, underscoring the original contribution to advance the application of big data analysis that is holistic, strengths-based and responsive to children's emerging identities.

7.1 Overview and Summary of Significant Findings

The overarching purpose of the thesis was to characterise shared patterns of social, emotional, language and executive function capabilities among New Zealand children in the year prior to school entry through a holistic and strengths-based approach in alignment with key principles of *Te Whāriki* (Ministry of Education, 2017). Acknowledging the importance of culture, language and identity, as highlighted in *Te Whāriki*, further underscores the need to recognise the potential for biases in measurement that may unequally—and thus unfairly—characterise performance and associated patterns for children from different genders and ethnic groups. A necessary step in this process was to estimate the potential bias from the child's gender and ethnicity as the covariates to the latent profile factors and examine the comparability of the optimal group solution across the child's gender and ethnic groups. Aligned with these purposes, the research was structured in two distinct phases, each addressing questions designed to advance knowledge of children's learning and development using the data from the GUiNZs 54-month dataset. As children in New Zealand

generally enter formal schooling at 60 months (5 years) of age, this dataset provides information about young children's capabilities, which include prosocial behaviours, emotional symptoms, emotional knowledge, expressive and receptive language, and executive function before school entry. The period of early childhood is a critical and formative phase of learning and development, and children's early experiences and capabilities in this stage have the potential to shape life trajectories (Phillips & Shonkoff, 2000).

The research was conducted through secondary analyses of this GUiNZs dataset using confirmatory factor analysis (CFA), latent profile analysis (LPA), multiple group confirmatory factor analysis (MG-CFA), and Multiple Indicators Multiple Causes (MIMIC) modelling, aligned with the two phases of the thesis. Phase 1 confirmed the factor structure of the available selected measures (5-item prosocial behaviours and emotional symptoms subscales in the SDQ, the 6-item AKT, and the 5-item PROL) through CFA and supported the theoretical relevance and appropriateness of these constructs for the sample. Latent profile analysis was then used to examine how the capabilities clustered with and across groups of children. A 3-group latent profile of social, emotional, language, and executive function capabilities was identified. The three profiles, representing the *Emerging*, *Progressing*, and *Competent* capabilities, reflected overall differences in young children's capabilities.

Phase 2 initially assessed the MI of the selected measures to ensure the consistency of interpreting the constructs across gender and ethnic groups. Due to the data

accessibility for PPVT and the nature of count data for the Luria Hand Clap Task, the MG-CFA tests were limited to the following measures: prosocial behaviour and emotional symptoms subscales in the SDQ, the AKT, and the PROL in the current sample. Results suggested partial MI was established for the prosocial behaviours and emotional symptoms factors of the SDQ (8.3 **Appendix 3**) across the child's gender groups, with full invariance observed for the one-factor model of the AKT (8.4 **Appendix 4**) and the PROL (8.5 **Appendix 5**). Similarly, partial MI was achieved across ethnicity for each of these measures.

Phase 2 then treated the child's gender and ethnicity as the covariates and estimated the DIF effects from the covariates to the latent profile factors using MIMIC modelling. These tests revealed potential non-invariance in specific measures and across profiles in the 3-group model, suggesting the need for caution in the comparability and interpretation of findings across diverse populations. Specifically, findings suggested sufficient invariance to support the comparability of the 3-group model across the child's gender, but insufficient invariance across ethnic groups, indicating that directly comparing the 3-group model across ethnic groups would be inappropriate. Key insights from the research will contribute to the knowledge of the shared patterns of young children's social, emotional, language, and executive function capabilities in strengths-based, holistic, and culturally connected ways. The sections below provide detailed summaries of the research findings.

7.1.1 Profiles Observed

Given the nature of the sample and range of capabilities assessed (i.e., the existing interrelationships among social, emotional, language and executive function capabilities in young children), the current research hypothesised that subgroups within the sample would primarily reflect overall differences in children's capabilities across the domains of development (i.e., a homogenous pattern of association among factors), rather than distinct or unique profiles characterised by specific capabilities (i.e., a heterogeneous pattern of association among factors). The hypothesis was then evaluated using the LPA to identify shared patterns of social, emotional, language, and executive function capabilities in young children in the sample overall. Due to the exploratory nature of the latent profile analyses, the optimal number of group solutions was not predetermined and required a rigorous evaluation of statistical criteria alongside theoretical justification.

A 3-group model was chosen as the optimal solution due to its ability to provide a parsimonious yet meaningful representation of the data while capturing genuine heterogeneity within the population. The three groups, labelled as *Emerging* (9.1% capabilities lower than peers), *Progressing* (45.1% capabilities close to average), and *Competent* (45.8% capabilities higher than peers), reflected overall differences in young children's capabilities using the GUiNZs 54-month dataset. Specifically, children in the *Emerging* profile scored lowest across the factors (e.g., prosocial behaviours, emotional knowledge, language, executive function) and the highest in emotional symptoms.

Children in the *Competent* profile scored the highest in factors except for the lowest emotional symptoms, while the *Progressing* profile scored close to the mean or median across these six profile factors.

Noticeably, the homogeneity in association among children's capabilities was characterised by overall differences across all domains. The observation of profile homogeneity in the current work may be attributed to the sample sizes required to detect distinct profiles (i.e., a small subgroup of children for whom individual or groups of capabilities develop at different rates/degrees), as observed in previous studies such as the large Australian study by Collie et al. (2019). Of 100,776 participants in their study, 70% identified a homogeneous Social-Emotional (SE) prosocial profile, while smaller groups exhibited distinct profiles: SE-anxious (12%), SE-aggressive (13%), and SE-vulnerable (5%). In the current work, such distinct profiles may not be detectable in the smaller sample size of 5,707 participants, which is insufficient to capture nuances observed in larger datasets. Notably, the smallest profile identified in the Australian study required an estimated sample size of 7,460 to reliably detect distinct profiles, which is larger than the overall sample available in the GUiNZs dataset. The smallest *Emerging* profile identified in the thesis primarily differed in emotional symptoms, while other capabilities remained consistent, reflecting homogeneity in the association between children's capabilities.

The focus on identifying and understanding the pattern of associations among characteristics in the current thesis aligns with developmental theories that underscore

the holistic nature of children's capabilities (Bronfenbrenner & Ceci, 1994; Fischer & Bidell, 2006; Piaget, 1948; Vygotsky, 1978). Fischer and Bidell (2006) conceptualised children's development as evolving patterns that exhibit stability at any given point. Changes in areas of development are dynamic and holistic, contributing to the overall trend in patterns of child development (Magnusson, 1995, 2003; Magnusson & Allen, 1983). The overall differences across domains observed from three profiles—*Emerging*, *Progressing* and *Competent* suggest general patterns in children's capabilities, while variations among the three profiles are observed. For instance, children in the *Emerging* profile may require support across all six developmental domains rather than targeting a single domain. The holistic view of children's capabilities compensates for the single-lens view arising from variable-centred approaches by exploring the ways capabilities cluster together, contributing to children's overall capabilities related to learning and development. The findings carry important implications for early childhood education in New Zealand, particularly the holistic nature of children's capabilities implications related to the types and intensities of supports and services for young children (i.e., a multi-tiered service approach as presented in *Te Tūāpapa o He Pikorua* and discussed in Chapter 5).

7.1.2 MG-CFA Approach

In the current thesis, MI tests for prosocial behaviour and emotional symptoms subscales in the SDQ, AKT, and PROL through the MG-CFA approach aimed to ensure fairness in interpreting the constructs of the measures across gender and ethnic

groups. The partial or full invariance indicated that these measures display a reasonable degree of consistency in their measurement of constructs across children's gender and ethnic groups. Children from different genders and ethnic groups respond to these measures in a similar and comparable fashion, confirming the robust use of the measure in diverse subpopulations.

The MG-CFA tests offer several strengths for establishing MI or partial MI, allowing for meaningful score comparisons of a measure between groups (Brown et al., 2017). MI results validated the constructs for each of the selected measures and provided an initial assessment of potential sources of DIF in the mixture model. This helps address group differences while guarding against the assumption of population homogeneity, where all individuals were drawn from a single population (Muthén, 1989). These results further support the evaluation of the comparability of the person-centred model across the observed subpopulation. Given the diversity of the early childhood populations, the tests also help minimise potential biases from item responses and in interpreting constructs (Vandenberg & Lance, 2000).

While addressing its strengths, it is worth acknowledging the limitations of MG-CFA in the current sample and latent profile factors. These include data sparsity, limitations of criteria for models using ordered categorical methods and WLSMV estimators, and the limitation of potential over-sensitivity in MG-CFA tests. Data sparsity can significantly affect parameter estimates, particularly when subgroup sizes constitute less than 5% of the overall sample (Collins & Lanza, 2010). This may arise from overall or

group sample size and low endorsement of rating scale options, which may reduce the accuracy and reliability of the estimates from measurement models (Howe et al., 2024). Another limitation in the current thesis is inconsistencies in model fit criteria for models using ordered categorical methods and the WLSMV estimator. These inconsistencies may complicate invariance assessment due to shifting standards and evaluators, the inadequacy of a single criterion for all scenarios and the reliance on varied criteria, such as the significance of changes in CFI for nested models (Rutkowski & Svetina, 2014). Additionally, the restrictive nature of MG-CFA may be overly sensitive to trivial non-invariance, potentially identifying insignificant differences as meaningful (Khademi et al., 2023). However, this over-sensitivity may have a limited impact on the person-centred model, which necessitates the assessment of DIF effects on the latent profile factors through MIMIC modelling in the current thesis.

7.1.3 DIF Approach to the LPA Model

The DIF approach through MIMIC modelling estimated the substantial bias in the associations between the covariates (i.e., the child's gender and ethnic groups) and the latent profile factors. This approach also examined whether the nature of the shared patterns of capabilities is comparable across the child's gender and ethnic groups. The examination also helps mitigate the risks of introducing biases in the prediction model.

Results suggested that the child's gender was a source of DIF effects for all six profile factors, while ethnicity was identified as a source of DIF effects for five factors, except for prosocial behaviours. For gender, significant uniform DIF effects were

observed for prosocial behaviours, emotional symptoms, and expressive language, while non-significant uniform DIF effects were observed for emotional knowledge across profiles. Non-significant nonuniform effects for receptive language and executive function were observed. Despite the presence of these effects, their magnitudes were negligible or small, suggesting sufficient invariance allowing for the comparison of the 3-group model across gender groups. For example, a significant uniform DIF effect with a small magnitude was observed in prosocial behaviours across gender groups. This suggested that the DIF effects on prosocial behaviours were consistent across three latent profiles, allowing for meaningful comparisons across gender across and within the profiles.

However, DIF effects varied within and across profiles for ethnicity. Except for no-DIF prosocial behaviours, significant uniform DIF effects were found for emotional symptoms among Māori, Pacific, and Asian children across the three profiles. Nonuniform DIF effects were observed for executive function, emotional knowledge, and expressive and receptive language, with the significance of these effects varying across ethnic groups and profiles. For example, small effect sizes ($|\beta| < 0.20$) were observed for Māori children (*Emerging* profile) and Pacific and Asian children (*Competent* and *Progressing* profiles). Medium effect sizes ($0.2 \leq |\beta| < 0.50$) were found for Pacific and Asian children (*Emerging* profile). Non-significant nonuniform DIF effects appeared in emotional knowledge for Other ethnicities. These results highlight insufficient invariance to support the comparability of the 3-group model across ethnic

groups. Although the three latent profiles—*Emerging*, *Progressing*, and *Competent*—were identified in the overall sample, along with the overall differences observed between factors, the shared patterns of capabilities do not remain consistent across ethnic groups. Hence, reporting the overall patterns of capabilities by ethnicity would be inappropriate.

The DIF approach was key to addressing unbiased estimates of children's capabilities across diverse populations. Research that ignores or avoids DIF effects in mixture models may introduce bias into model comparisons (Masyn, 2017). In the current work, assessing the 3-group model of children's capabilities across ethnicities without accounting for DIF effects on latent profile factors may lead to inaccurate conclusions on the model comparisons. Such an approach may produce biased estimates for how ethnicity predicts the latent profile, potentially misinterpreting the true differences in profile factors across ethnic groups. The current work addressed the importance of acknowledging the limitations of the data by incorporating the DIF approach to account for DIF effects. The current work also cautioned against the assumption that all latent profile factors were presumably equally interpreted across gender and ethnicity among young children, thereby ensuring more rigour and avoiding biased interpretations.

7.1.4 Overall Discussion

The current research is supported by and extends on previous research that has used a holistic approach to examine children's learning, development, and health using

the GUiNZs 54-month dataset (i.e., McLaughlin et al., 2024; Russell et al., 2022). Table 7-1 presents an overarching summary of the McLaughlin et al. (2024) and Russell et al. (2022) research alongside the current research. Each study used the 54-month data, with variations in sample sizes likely related to study inclusion criteria and how each study addressed missing data. For example, my research included participants with data from at least one observational task completed in the 54-month dataset.

The table presents aspects of the studies that were similar or different regarding the analysis methods used, domains of interest, and the GUiNZs measures that were selected. The table also shows the number of profiles identified in the optimal solution, any names or labels applied to each profile, along with the proportion of the sample and a description of key features.

Table 7-1

Three Person-Centred Studies Using the GUIiNZs 54-Month Dataset

Citation	Analysis methods	Domains of interest	Measures of profile indicators	Numbers and labels of profiles	Percent of Sample	Description
Russell et al. (2022) n=6109	latent profile analysis (LPA); multinomial regression	physical health, motor development, socioemotional and behavioural development, communication, and learning	(1) Strengths and Difficulties Questionnaire (SDQ)	(1) healthy	52.6%	better than average mean scores in six of the seven health and development variables
			(2) The Affect Knowledge Test	(2) early social skills flourishing	14.5%	lowest mean scores for behavioural difficulties and highest mean scores for emotional knowledge skills
			(3) Hand Clap Task	(3) early learning skills flourishing	4.0%	highest mean scores for executive function skills and early literacy skills
			(4) Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	(4) early learning difficulties;	19.5%	mean scores for executive function skills and early literacy skills were lower than other profiles
			(5) Gross and motor skills scale	(5) physical health difficulties	5.6%	mean scores for acute health problems and chronic health problems were higher than other profiles
			(6) Number of chronic medical conditions and acute health problems	(6) difficulties in development	3.7%	low scores across all health and development variables
Mclaughlin et al. (2024) n=5725	confirmatory factor analysis (CFA); latent profile analysis (LPA); multinomial regression; Multiple group confirmatory analysis (MG-CFA)	behavioural strengths and difficulties, oral language ability, executive function, pre-academic skills, gross motor functioning, receptive language ability	(1) Prosocial behaviours and emotional symptoms subscales in the Strengths and Difficulties Questionnaire (SDQ)	Profile 1	14.9%	strengths in emotional security and preacademic skills, with relative needs in prosocial behaviour
			(2) Parent Rating of Oral Language (PROL)	Profile 2	15.0%	similarities to peers across all areas with relative needs in prosocial behaviour
			(3) Hand Clap Task	Profile 3	37.2%	strengths in receptive language and preacademic skills, with relative needs in gross motor and prosocial behaviour
			(4) DIBELS Letter Naming Fluency task, Name and Numbers tasks	Profile 4	17.3%	below-average scores across all areas, with relative strengths in prosocial behaviour, emotional security, and gross motor functioning
			(5) Gross and motor skills scale	Profile 4	17.3%	strengths above the sample means in all areas, particularly in prosocial behaviour and receptive language
			(6) Peabody Picture Vocabulary Test 3 rd edition (PPVT-III)	Profile 5	15.6%	strengths above the sample means in all areas, particularly in prosocial behaviour and receptive language
current work n=5707	confirmatory factor analysis (CFA); latent profile analysis (LPA); multinomial regression; multiple group confirmatory analysis (MG-CFA); differential item functioning (DIF)	prosocial behaviours, emotional symptoms, emotional knowledge, expressive language, executive function, receptive language	(1) Prosocial behaviours and emotional symptoms subscales in the Strengths and Difficulties Questionnaire (SDQ)	(1) emerging	9.1%	capabilities that were lower than their peers in other profiles
			(2) The Affect Knowledge Test	(2) progressing	45.1%	close to average capabilities
			(3) Parent Rating of Oral Language (PROL)	(3) competent	45.8%	the highest capabilities relative to peers
			(4) Hand Clap Task			
			(5) Peabody Picture Vocabulary Test 3 rd edition (PPVT-III)			

Consistencies and differences between the current research and two other person-centred studies within the GUiNZs provide a unique opportunity to explore key issues related to children's learning and development in New Zealand. This includes the importance and insights gained from using a person-centred approach to child data, the different patterns of characteristics observed when different combinations of indicators are included, and differences in overall focus when research is presented from a deficit or strengths-based perspective. This also includes an opportunity to gain an overarching takeaway about children's learning and development in New Zealand by drawing across the 3 studies. Finally, differences in the treatment and analysis of gender and ethnicity across the studies highlight the unique and important contribution of the present study to advance the methods used with children's learning and development data to ensure the appropriate use of measures and fairness in assessment.

First, all three studies are person-centred, adopting a holistic approach to explore how capabilities cluster together within a New Zealand preschool-aged cohort, differentiating from variable-centred studies. Prior to these three studies, research conducted with the GUiNZs 54-month data focused on specific aspects of children's learning and development and key associations cross-sectionally or longitudinally (cf., Buckley et al., 2019; D'Souza et al., 2019; Jan et al., 2021; Neumann et al., 2019). The focus on single domains of importance often arises from variable-centred approaches to exploring child outcomes (Howard & Hoffman, 2018; Laursen & Hoff, 2006). A

variable-centred approach focuses on the associations between individual variables, whereas a person-centred approach integrates the interconnectedness of variables through a holistic lens (Howard & Hoffman, 2018). For example, the study McLaughlin et al. (2024) identified five profiles that represented a snapshot of holistic learning and development in young children. They argued for the need for a nuanced approach to measuring and assessing children's data, one that is responsive to the holistic patterns of children's learning and development. Similarly, Russell et al. (2022) generated six profile patterns of children's health and development outcomes and highlighted the need to cluster children's health and development outcomes on a continuum, and related that to socioeconomic and ethnic inequities in children's developmental health. Consistent with these two studies, the 3-group model in the current research viewed social, emotional, language and executive function capabilities in young children holistically, reflecting the integrated and multifaceted nature of child development. Taken together, these studies collectively addressed the importance of investigating key areas of development in a holistic lens within health and education services, rather than targeting single domains.

Second, while all three studies focused on different combinations of indicators, only the two prior studies resulted in unique patterns. The current research identified overall differences in the patterns of children's capabilities, rather than distinct patterns. Different combinations of selected indicators allow for specific distinctions, which can lead to variations across LPA models and impact the selection of optimal group

solutions in person-centred studies (Bauer, 2022). In Table 7-1, studies by McLaughlin et al. (2024) and Russell et al. (2022) show consistency in social, emotional, language, executive function and motor skills indicators, with differentiation primarily arising from their focus on health outcomes and pre-academic skills. The current research examined the capabilities common to the two studies, including social, emotional, language and executive function, but without the pre-academic, motor skills, and health indicators. Given the established interconnections among social, emotional, language and executive functions in the existing literature, it is not surprising that homogeneity was observed across the capabilities from the 3-group model in the current work, which exhibits overall differences in children's capabilities.

Third, across the three studies, the focus on a strength-based perspective, which is strongly advocated for in early education settings, differed. McLaughlin et al. (2024) and the current research specifically highlighted a strengths-based perspective, aligned with perspectives in *Te Whāriki* and early childhood research. For McLaughlin et al. (2024), this involved the decision of not labelling the profiles and describing them based on learning support needs. For the current study, the labelling of profiles worked from the position that all children have growing capabilities and are in different places along their unique journey. In contrast, the Russell et al. (2022) study occurred in the context of health research, which is often more focused on risk and prevention of negative health outcomes. Within early education, the type of labelling of profiles as having difficulties might be characterised by a deficit-focused perspective. Deficit-

focused narratives provide insights into social, emotional, and academic development; however, such perspectives may overlook children's strengths, growth, and developmental changes over time while disproportionately addressing atypical development (Albritton et al., 2023). This approach may also reinforce bias and create challenges for promoting inclusivity in education settings (Black-Hawkins, 2017). The strengths-based perspective and assessments, as reflected in *Te Whāriki*, support positive views of child development that understand the diverse capabilities and needs in children's learning and development, rather than focusing on deficits or risks (Galloway et al., 2016; Madsen Sjö et al., 2021). Aligned with the strengths-based perspective in *Te Whāriki*, the selection of the measures to portray the latent groups in the current research centred on children's strengths, such as prosocial behaviours and emotional knowledge. The inclusion of the SDQ and particularly the emotional symptoms subscale was an exception to this strengths-based focus and is further discussed in the limitations section of this chapter. Nonetheless, a strength-based approach to working with children's learning and development data emphasises children's strengths across domains of development.

Fourth, a promising finding across all three studies is that the majority of children were identified as progressing well or with strengths in key areas of learning and development. Across all three studies, less than 20% of children were identified as needing more support for their health, learning or development. The study by McLaughlin et al. (2024) identified 17.3% of the children as below the average across all

areas of development, while the study by Russell et al. (2022) found 3.7% of children identified as having difficulties across the areas of development. Consistent with these studies (McLaughlin et al., 2024; Russell et al., 2022), the current research also suggests that most children demonstrate adequate capabilities, with only a small proportion (9.1%) of children in the *Emerging profile* identified as having lower capabilities than their peers across all areas of development. This provides implications for education services designed to accommodate the capabilities across areas and prioritise inclusivity for all children. Targeted services should be tailored to address the specific needs of a small proportion of children with lower capabilities, without neglecting the needs of the wider early childhood population. Related to this, McLaughlin et al. highlighted the multi-tiered system of support framework referred to as *Te Tūāpapa o He Pikorua* (Ministry of Education, 2020), which was also described in Chapter 5 of the present research, to highlight that it is typical and expected that different proportions of children will need varying levels of support within the context of high-quality early learning services.

Fifth, and finally, all three studies used ethnicity and/or gender as key covariates but approached them differently through distinct analytical techniques. The study by Russell et al. (2022) explored the association between ethnicity and latent profiles to help explain the underlying factors contributing to the inequalities in children's development and health. They reported the associations between optimal group solutions and ethnicity through multinomial regression analyses and compared the

group membership across ethnicities. They reported that children of Māori and Pacific ethnicity had increased odds of being classified with a sub-optimal health profile. Notably, they did not conduct MI tests to confirm that the measures were equitably understood and interpreted across ethnic groups. Instead, they highlighted the importance of using appropriate measures of fairness and equity in their discussion, but these were not systematically tested.

McLaughlin et al. (2024) conceptualised gender and ethnicity as children's emerging identities and explored MI before conducting latent profile analysis of children's learning and development to address the diversity and inclusivity in early childhood settings, while highlighting the effectiveness of measurements. Their work also identified full and partial MI for some of the measures and included a range of measures in which MI could not be tested. Due to this, the authors acknowledged that comparative statements about the distribution of children's learning and development profiles by gender and ethnicity without further tests to determine the comparability of profiles were not advisable.

The present thesis advances these studies by employing a more comprehensive mixture model that accounts for the potential biases introduced by gender and ethnicity as covariates to the latent profile factors. Building on prior research testing MI in person-centred models (Bettencourt et al., 2022; Masyn, 2017; McLaughlin et al., 2024; Tsaousis et al., 2020), the current research estimated the potential biases in the LPA using the DIF approach via MIMIC modelling, suggesting that the 3-group model

is comparable across gender groups but not across ethnicity. This means that participants from diverse backgrounds, such as gender and ethnic groups, may understand the constructs of the measures differently, which may pose biases in the responses. Given the impact of diverse identities on development capabilities, the effectiveness of measurements and the fairness in understanding children's learning and development, it may be challenging to develop comparable measures that account for their diverse identities. These findings have significant implications for methodological, practice, research, and policy implications, as elaborated in the implications section.

7.2 Implications

7.2.1 Methodological and Practical Implications

The current study highlights the need to characterise and quantify young children's social, emotional, language and executive function capabilities in strength-based, holistic, and culturally responsive ways. The optimal 3-group model reflecting overall differences in young children's capabilities across the domains of development, may be useful for early childhood educational research and practices, service design, and policy.

7.2.1.1 Person-Centred vs Variable-Centred Approach.

The holistic, person-centred approach reflects the shifting tide to explore methodologies to better understand and account for children's capabilities in holistic and culturally responsive ways. A person-centred approach provides a more

comprehensive understanding of child development than a variable-centred approach (Bergman et al., 2015). The variable-centred approach assumes that participants' characteristics are homogeneous, implying that participants belong to the same population (Ferguson et al., 2020). This approach often reflects a single-lens perspective, which may only capture information about the domains under investigation in a way that may not reflect the finer details of associations between domains (Bergman & Trost, 2006). In contrast, the person-centred approach considers domains holistically, rather than focusing solely on a specific domain. The approach allows for a broader yet more nuanced exploration of the association between domains and individual differences in child development (Bergman & Trost, 2006). Such holistic views of children's capabilities recognise children as complete entities and underscore the dynamic interaction between various aspects of child development alongside influences that drive learning and development.

The current research offers additional methodological and practical implications on fairness and MI and relevance to early childhood education, which are discussed separately in the following sections.

7.2.1.2 Fairness and MI.

A second methodological implication of this research concerns the fairness of assessments as a quality of inclusivity in the education context. Educational assessment is critical to informing educators about teaching and learning. Unbiased assessment is at the heart of upholding inclusivity and fairness in the education context, and

assessments must be conducted impartially (Tierney, 2017). First, the selection of measures must align with a rigorous and meaningful understanding of the capabilities of the diverse child population. The inability of the measures to accurately reflect the true characteristics of certain groups may undermine the transparency and reliability of information regarding children's capabilities (Putnick & Bornstein, 2016). Specifically, measures that are broadly applicable to the overall child population may not capture the unique characteristics of specific subpopulations that are explicitly considered. When measures are not well-suited to certain observed groups, the resulting conclusions may potentially be biased, thereby compromising the validity of the findings. Second, the presence of measurement bias may result in inaccurate or misleading conclusions about children's capabilities (Wu et al., 2017). For example, researchers may interpret differences observed as true differences in children's capabilities. However, the differences observed may result from the measurement bias, potentially leading to the misinterpretation of the findings. Hence, it is important to actively attend to bias in assessment to help ensure fairness in using data and interpretation of the results.

In the current work, MI tests through MG-CFA and the DIF approach via the MIMIC modelling technique provide insights into examining the presence of measurement bias. MI tests through MG-CFA assessed whether the constructs of the measures were understood equivalently across gender and ethnic groups. However, MG-CFA could not be applied to the PPVT-III due to data unavailability, nor to the Luria Hand Clap

tasks due to the nature of the count variable of the Luria Hand Clap task in the GUiNZs 54-month dataset. The absence of invariance tests for these two measures may impact the invariance of the 3-group model. To address the limitation, the DIF analysis was employed as it compensates for the constraints of the MG-CFA by accounting for potential biases in measures at the model level, especially the two measures that could not be tested through MG-CFA. The DIF analysis also allows for estimating and comparing models that account for nonuniform and uniform DIF effects from the covariates to the latent profile factors (Bettencourt et al., 2022; Masyn, 2017; Tsaousis et al., 2020), which helps address issues in measurement, such as equivalency and biased responses across groups (Masyn, 2017). Therefore, the use of MG-CFA and the DIF approach in the current work serves as an example of how assessment can incorporate appropriate use of measures to support fair interpretation and comparisons in future research.

Ensuring fairness in interpreting results through MI tests reflects a nuanced understanding of the diverse identities that may influence child development. Children's diverse cultural backgrounds and characteristics, such as ethnicity, may impact the capabilities assessed (Ministry of Education, 2017). The thesis identified gender and ethnicity as key aspects of children's emerging diverse identities and evaluated their educational relevance in development through MI tests, which offer insights into children's diverse identities by treating these identities in a fair and equitable manner. By assessing the measurement invariance at the construct- and

model-levels across these subpopulations, the current work draws meaningful conclusions that support generalisations and comparisons of the findings associated with these identities. For instance, the comparison of the 3-group model across ethnicity did not proceed due to the lack of invariance in the factors of the 3-group model across ethnic groups, leading to the incomparability of the model observed in Chapter 6. This mitigates the risks of overlooking the heterogeneity in the subpopulations, which could potentially mislead conclusions. This approach highlights the importance of considering the indirect impact of children's diverse identities on early learning services, local curricula, and assessments, which must be tailored to reflect the distinct cultures and ethnicities that shape children's upbringing in New Zealand. Such insights may contribute to supporting equity and inclusive practice by acknowledging how children's emerging identities impact their learning and the fair and unbiased assessment of their learning within the New Zealand context.

[7.2.1.3 Relevance to Early Learning.](#)

Children's social, emotional, language, and executive function capabilities represent key areas of learning and development, necessitating a strengths-based, holistic view to effectively recognise the interconnectedness of these areas (Magnusson, 1995). Part of the intricacy and interconnectedness is reflected in the current early childhood curriculum—*Te Whāriki*, which views the child as a 'whole learner' (Ministry of Education, 2017). Building on these perspectives, *Te Whāriki* states that assessment makes valued learning visible. While the statement is intended for early childhood

teachers, it could equally be informative to researchers and policymakers. What we assess, how we assess, and how we analyse the assessment of children matter. This thesis aimed to be guided by values relevant to young children's learning as identified in *Te Whāriki*. These included taking a holistic, strengths-based, and culturally responsive view of children.

Early childhood teachers should develop a holistic understanding of children's learning and development through assessments and evaluations. As highlighted in the Early Learning Action Plan (Ministry of Education, 2019b), high-quality early learning requires teaching staff to possess expertise and culturally responsive knowledge to support positive learning and development trajectories, ensuring children's holistic development. The current thesis helps teachers gain insights into the unobserved latent subgroups, recognising that children's learning and development generally follow a predictable sequence. The 3-group model with overall differences observed in young children's capabilities and a small proportion of children having lower capabilities than their peers suggests that teachers must address these capabilities in an integrated manner. This means that understanding how children's capabilities cluster together to support their daily function while also acknowledging individual variations in child development that some children may experience slower progress in certain areas and may require further assessment, targeted intervention and support (Ministry of Education, 2017). Teachers should recognise general development patterns in children's developmental stages and provide children with developmentally appropriate services,

as these insights into the skills young children are likely to acquire at different stages of their early years. This holistic view of child development can optimise children's learning outcomes and assure them of high-quality early learning experiences.

The person-centred, holistic approach to understanding young children's capabilities may have important implications for early childhood services. In the current thesis, the 3-group model revealed overall differences in young children's capabilities, which did not result in distinct profiles. This aligns with the multi-tiered approach of support in *Te Tūāpapa o He Pīkorua* (Ministry of Education, 2020), which is an evidence-informed, multi-tiered approach to supporting inclusive learning. The tiered support is dynamic and integrated, from universal (Tier 1) to targeted (Tier 2) and individualised (Tier 3), based on children's strengths, needs, and progress. Tier 1 addresses the support for all children. Consistent with Tier 1, children whose individual profiles align with the *Competent* and *Progressing* profiles in the current work may require support to strengthen their strengths. Tier 2 provides targeted support to those who require additional adaptation. Children whose individual profiles align with the *Progressing* profile may need support for receptive language, executive function or other areas of development based on their unique capabilities at the local level. Tier 3 offers individualised support for children with more learning support needs. Children whose individual profiles align with the *Emerging* profile may require support across all capabilities. Therefore, it is useful to consider the 3-group model in relation to *Te Tūāpapa o He Pīkorua* in terms of the support children need for their capabilities and

development (McLaughlin et al., 2024).

7.2.2 Research Implications

The data processing and analytical procedures in the present study were designed to ensure the rigour of research findings, reflected by including factor scores as indicators of the LPA and the DIF approach to testing DIF effects in the current thesis. In the thesis, due to the unknown invariance of the PPVT and the Hand Clap Task scores, the use of factor scores representing latent profile factors may improve the convergence and reduce the impacts of using non-invariant indicators in the mixture models (Nylund-Gibson & Masyn, 2016; Wang et al., 2022). By implicitly controlling for measurement errors, the use of factor scores offers a more robust alternative than scale scores, thereby improving the validity of conclusions drawn from the data. Moreover, the DIF approach is flexible in accommodating various types of covariate variables and latent profile factors, enabling more accurate estimation and comparison of models than approaches such as MG-CFA (Masyn, 2017; Tsaousis et al., 2020). This method provides potential nuances and different perspectives on a phenomenon by identifying sources of non-invariance that may impact the comparability of the person-centred model. It reflects robust data processing and analysis, ensuring the rigour of research by providing a more comprehensive understanding of the data and findings. Based on the work presented in this thesis, it is recommended that future research focused on children's learning and development should specifically address the advantages and limitations of the types of scores used for analysis and the specific analytical techniques

that accommodate the variables of interest.

The measurement non-invariance across ethnic groups highlights the need for both developing culturally responsive assessment tools and critically evaluating existing measurement tools that validly capture capabilities across diverse populations (Padilla & Medina, 2001; Thomas et al., 2010). The development of such tools and the evaluation of the existing tools should emphasise participants' identities and cultural backgrounds, recognising these as integral to the capabilities or constructs examined. This requires close collaboration with the ethnic communities to ensure the constructs being measured, the language used, and the overall design are meaningful and reflect culturally grounded values, norms, and perspectives. Additionally, adopting mixed methods may help address the limitations of existing tools. Such an approach allows for a better understanding of the socially and culturally situated capabilities that may relate to being viewed as competent within a given domain. In doing so, it helps mitigate potential sources of bias and provides a more comprehensive account of contextually grounded insights into children's capabilities. This perspective highlights the need for ongoing critique and refinement of existing measures.

7.2.3 Policy Implications

The findings of the current thesis have potential policy implications for how capabilities are understood and supported through policy initiatives and teaching and learning services. Government funding initiatives and programmes are designed to address health, social, educational, and economic challenges in ECE, to bridge gaps

between diverse populations and to improve outcomes. Evidence-based policy must address both technical and normative constraints in measuring outcomes, with particular attention to the quality of the measures used for policy purposes. Current education frameworks, such as the *Te Tūāpapa o He Pikorua* and *Te Whāriki*, consider various aspects of children's strengths and needs, reflecting a holistic approach to supporting children's learning and development. A focus on flexible, tailored and tiered support programmes in ECE may enhance lifelong outcomes by considering aspects of children's learning and development through a holistic lens by focusing on all domains of learning and development within a population-based model. Tailored support, including universal, targeted, and individualised support, identifies both similarities and differences in young children's capabilities. This is important for advancing modern policy development by incorporating and prioritising person-centred methods to examine children's learning and development holistically. Policies in designing initiatives should prioritise children's diverse capabilities and identities to support children's learning holistically and from a strengths-based perspective that addresses children's unique identity, language, and culture.

7.2.4 Theoretical Implications

The findings of homogeneous profiles—characterised by generally high, medium, or low levels across all domains—may suggest important theoretical considerations about the uniformity and interconnection of children's capabilities. Specifically, the homogeneous patterns suggest that children's capabilities are interconnected and tend

to co-develop in an integrated manner at 54 months, reflecting developmental theories that involve developmental mechanisms and environmental contexts. According to Piaget, children at this age are in the preoperational stage, during which social, emotional, language and cognitive abilities develop simultaneously, leading to parallel growth across domains (Piaget, 1983). Moreover, the interdependence of cognitive functions and active knowledge construction through environmental interaction reinforces the uniformity of capabilities. Similarly, the dynamic system theory views children's capabilities and development as an integrated process where multiple capabilities emerge together over time (Eynde & Turner, 2006; Fischer & Bidell, 1998; Lewis, 2000). Sameroff's transactional model of development explains the continuous and reciprocal interactions between the child and their environment (Sameroff, 2010). Early experiences in one domain can influence other domains, creating mutually reinforcing growth that contributes to the uniformity.

The homogenous patterns may also suggest that, in the early childhood stage, individual differences are more quantitative, with children's capabilities varying in overall level—generally high, medium, or low in each profile—rather than qualitative, where patterns of capabilities across domains would be distinguishable from others (e.g., high-low-high, or high-medium-medium within each profile). The identified overall differences in level resonated with the broader notion of stage-based development from Vygotsky's theories, which highlight that children tend to show similar levels of abilities across multiple domains at similar stages (Vygotsky, 1962).

Additionally, the domains selected and the measures used in research may shape the patterns observed in children's capabilities, influencing both the theoretical implications drawn from the findings and the guidance available for instructional support. In the thesis, the domains examined provide an overview of children's social, emotional, language and executive function capabilities, while an in-depth exploration of subdomains using appropriate measures may result in distinct patterns that inform targeted instructional strategies. For instance, within emotional capabilities, measures of emotional knowledge, emotional recognition, emotional expression, and emotional regulation can provide more targeted insights into children's emotional capabilities, enabling educators to plan instructional support that is both precise and holistic (Denham et al., 2012). Moreover, the inclusion of academic measures addresses both children's capabilities and exposure to learning experiences, underscoring the importance of considering both developmental and contextual factors when planning support (McLaughlin et al., 2024).

7.3 Limitations of the Research and Future Directions

The thesis advances knowledge of shared patterns of young children's social, emotional, language, and executive function capabilities; it is essential to acknowledge the overarching limitations associated with this research overall. In addition to the limitations specific to each study addressed in their respective chapters, this section offers a brief overview of the limitations. Building on the discussions and limitations, this section also provides a clear understanding of the challenges encountered throughout

the research process and suggests potential directions for future studies.

7.3.1 Summary of Limitations

The limitations discussed in the previous chapters primarily stem from the choices of the research design, measures, and the restricted sample. The following is a summary of the limitations identified in the current study. Some of these limitations are phase-specific and have been previously discussed in Chapters 5 and 6, while others are overarching in nature and identified across the phases of the research.

Although the research design and decisions—such as the selection of variables—were made thoughtfully to align with the research aims and the available data in the GUiNZs 54-month dataset, the lack of external validity and absence of longitudinal analysis represent key methodological limitations. The absence of validation regarding how the 3-group solution of children’s capabilities relates to subsequent outcome variables may limit the ability to assess the predictive utility and meaningfulness of the profiles. The cross-sectional design does not allow for the examination of developmental trajectories or causal relationships between the 3-group solution and outcome variables over time. With such limitations, it remains unclear whether the identified profiles can reliably inform the developmental trajectories of children’s capabilities or apply the conclusions to broader populations of young children. Together, these limitations highlight the need to incorporate longitudinal designs and validation against outcome variables to enhance the robustness of the findings.

The selected measures reflect a limited understanding of children’s capabilities. For

example, there are limitations in child observation tasks, such as the Affect Knowledge Test (AKT) and the Luria Hand Clap Task, which may not fully align with children's day-to-day experiences and cultural backgrounds. These tasks may be unfamiliar or less relevant to some children, which may impact their performance. Furthermore, the current research could not validate the raw scores of PPVT-III through CFA due to restricted data access. Similarly, the Hand Clap Task could not be validated as its scoring is based solely on count variables, limiting the precision of performance evaluation. While including single-ethnicity categories for analysis ensured adequate sample sizes for the analytical methods in both phases, this approach overlooked dual- or multiple-ethnicity groups, which limits the understanding of the diverse population. Moreover, the child's gender was measured by the birth sex, which may not be able to capture the contemporary identities of gender. The complexities and fluidity of children's identities can create challenges for the operational definitions of variables that have a sufficient sample size for analytic purposes. The ways these identities were measured and understood have important implications and limitations for assessing children's developmental outcomes and addressing inequities.

The current research did not report the profiles fully from a strengths-based lens, particularly in terms of emotional symptoms, which were framed as difficulties or problems in the SDQ subscales. Given the complexity of child development and assessment, it is challenging to use perfect data that captures all potential sources of information about child development in ways aligned with the current research.

Specifically, the nature and quality of measures used in New Zealand associated with children's capabilities are often deficit-focused and tend to be applied to the overall population rather than considering the diversity within the population and testing the invariance of the measure before use (Carr, 2001; Niles, 2015). The current study used what was available in the GUiNZs 54-month dataset to gather information that reflects children's strengths. In the current research, the strengths-based approach sees 'problems' or 'difficulties' as areas of potential growth and learning support in a broader context of child development rather than areas that need to be fixed and as characteristics of individual children, which is a more common perspective in a deficit-based approach (McMillen et al., 2004).

Although the secondary datasets the GUiNZs are broadly representative of the 54-month-old children in New Zealand, the sample was gathered in limited geographical regions and time, including the Auckland, Manukau, and Waikato District Health Boards, where children aged 54 months were born between 2009 and 2010. This means that the capability profiles generated in the current work may be limited to time and space, and may not fully reflect the reality of 5-year-olds today. Today's 5-year-olds grew up during the COVID-19 lockdowns, and their childhood experiences may display variations in the patterns of capabilities compared to children born and raised in the pre-COVID era. The COVID-19 pandemic has had significant indirect impacts on children's socialisation skills, physical and mental health, and educational attainment (Mulkey et al., 2023). Emerging research has shown that children who grew up during

the pandemic had reduced social interaction associated with children's early language and social development (Putnick et al., 2023), more opportunities for e-learning, which may relate to behavioural problems, lower vocabulary knowledge and delayed developmental milestones (McArthur et al., 2022). Parents, practitioners, and educators must consider the child's 'COVID-19 pandemic experience' as an important factor influencing their current development. With this in mind, children's development and their needs may differ from those of previous generations. Therefore, findings from the current work may not fully reflect children's capabilities today due to the population-specific nature of the data. It is important to contextualise findings with respective timeframes and ensure the most up-to-date information is used to inform policy decisions.

Finally, the data analytical procedures in the current research may present several methodological constraints that may impact the interpretation of the findings. Methodological decisions, including approaches to handling missing data, ethnicity prioritisation, and parameter specifications, were carefully made based on the researcher's knowledge and interpretations, informed by child development theories and statistical justifications for assessing model fit. Such decisions may inherently involve subjective elements that may result in variations in the interpretation of findings if approached differently by other researchers. For example, the interpretation of the 3-group model (*Emerging*, *Progressing*, and *Competent*) of young children's social, emotional, language, and executive function capabilities was guided by the

perspectives in early childhood education, especially *Te Whāriki*. Other researchers may interpret these group solutions in different ways, such as low, medium, and high achievement in these capabilities. Therefore, it is important to exercise caution in methodological decisions, as distinctions in interpretations may arise.

7.3.2 Future Directions

Future research may pursue alternative directions.

First, additional areas of development may be considered to broaden the scope of young children's capabilities as outlined by *Te Whāriki*, such as children's physical or motor development. Although physical and motor development are important aspects of early childhood, they were excluded from the present study, which prioritised domains aligned with the researcher's expertise and allowed for in-depth analysis of the key areas within the defined scope.

Second, future research may benefit from examining more diverse samples of young children and incorporating additional variables related to their social, emotional, language, and executive function capabilities. Notably, children with neurodivergence or disability were not explicitly identified in the current dataset and represent only a small proportion of the population. As such, they were not included in the latent profile analysis. However, these factors may influence both children's engagement with assessment tools and their representation across latent profiles. Accounting for the diverse ways in which neurodivergent or disabled children process information, communicate, and regulate behaviour could help mitigate measurement bias and more

accurately reflect their underlying capabilities. Furthermore, previous research suggests that disability classifications may not reliably capture children's functional abilities and may therefore have limited utility in the formation of latent profiles (McLaughlin et al., 2015). Further investigation into the intersection of young children's capabilities, disability classifications, and the utility of latent profiles is warranted.

Third, selecting statistical analysis techniques that align with the research objectives is essential for ensuring the transparency and rigour of the research design.

Researchers' methodological decisions significantly shape research outcomes and their broader implications. While no single method can be deemed universally superior, as each technique has its strengths and limitations, the selection process should be guided by the specific research aims, questions, and contextual factors of the study (Opoku et al., 2016). In the current work, compared to other clustering methods, mixture modelling techniques offer the advantage of rigorously generating profiles of young children's capabilities. These techniques allow researchers to examine differences in the population in two ways: qualitative, by capturing heterogeneous patterns of associations within a population, and quantitative, by identifying homogeneous patterns within a population. This approach provides a nuanced understanding of the diverse characteristics of the populations (Collins & Lanza, 2010).

Fourth, methodological refinements, such as the use of other advanced statistical techniques, may be needed. The current thesis employed a cross-sectional rather than a longitudinal design. Research investigating whether or how the shared patterns of

young children's social, emotional, language and executive function capabilities change over time may adopt a longitudinal design and use analytical techniques appropriate to such a design. One such technique is known as latent transition analysis (LTA), which analyses the changes in individual developmental trajectories over time by identifying distinct latent states and observing how individuals transition between these states (Collins & Lanza, 2010). Additionally, future research may consider alternative methodological approaches to examining MI at the construct or person-centred model level across diverse observed groups. For instance, the alignment method offers advantages for testing MI at the construct level, as it allows for a nuanced examination of MI through the selective alignment of parameters (Ioannis & Fathima, 2023). This method is especially effective when analysing MI across multiple groups, particularly when the number of groups being compared is substantial (Asparouhov & Muthén, 2014b). Further, researchers may incorporate additional observed grouping variables, such as socioeconomic status (SES) and different age groups, to explore whether and how these factors impact the latent profile factors.

7.4 Conclusion

Given the importance of young children's capabilities, it is important to unpack these capabilities holistically and ensure fairness in assessment across diverse subpopulations (i.e., gender and ethnic groups). The thesis was conducted in two sequential phases to explore latent subgroups with shared patterns of children's capabilities in the 54-month dataset in GUiNZs; and to examine the comparability of

these shared patterns of capabilities across each of the observed groups (i.e., the child's gender and ethnicity).

Phase 1 involved a latent profile analysis to generate up to six subgroups of possible latent profiles following a prior check of the validity of the selected measures through confirmatory factor analyses (CFA). The 3-group model was chosen as the optimal group model because it provided a more parsimonious yet meaningful representation of the data while capturing genuine heterogeneity within the population. These profiles—*Emerging* (9.1% below average), *Progressing* (45.1% average), and *Competent* (45.8% above average)—reflected overall differences in young children's capabilities based on the GUiNZs 54-month dataset. These findings suggest that only a small proportion of children may require additional support across the areas of development. Rather than distinct patterns, the observed overall differences in the 3-group model of children's capabilities were supported by developmental theories that addressed the interconnections between these capabilities. These findings from the current work offer a broader understanding of general trends in child development, thereby contributing to teaching and learning policies and practices in early childhood education in New Zealand.

Phase 2 included the MG-CFA of the selected measures to ensure the constructs of the available measures are understood equivalently across the child's gender and ethnic groups. Due to the data accessibility and count data nature of the Luria Hand Clap Task, the MG-CFA tests were limited to the available measures (prosocial

behaviour and emotional symptoms subscale in the SDQ, the AKT, and the PROL) in the current sample. This necessitates the tests of estimating the substantial bias in the associations between the covariates (i.e., the child's gender and ethnic groups) and the latent profile factors. The DIF approach through the MIMIC modelling was used to estimate the potential bias from the covariates to latent profile factors and to examine the comparability of latent subgroup solutions across gender and ethnic groups. Findings showed negligible or small uniform and nonuniform differential item functioning effects by gender, which supported the comparability of the 3-group model across gender groups. More substantial uniform and nonuniform differential item functioning effects by ethnicity were found, which did not support the comparability of the 3-group model across ethnic groups.

These analyses and findings contribute to fairness in the interpretation of child development data. Tests of measurement invariance for the constructs of the measures and person-centred model seek to ensure unbiased assessments in understanding the diversity and inclusivity of the population characteristics, which have implications for teaching and learning policies and practices in the early childhood context in New Zealand. When non-invariance is identified, appropriate cautions and restrictions on comparative statements can be made explicitly. This is important to advance research that is fair, equitable and rigorous in early childhood.

The current thesis makes an original and significant contribution to the holistic understanding of young children's social, emotional, language, and executive function

capabilities, emphasising strengths-based, holistic, and culturally relevant approaches to research about children's learning and development. Additionally, it highlights the importance of fairness in assessing and interpreting the diversity within the population of young children. The findings support the holistic development outlined in *Te Whāriki*, offering valuable insights for guiding early childhood teaching and learning policies and practices. The present study demonstrates research methods to describe and quantify children's capabilities with a focus on strengths-based, holistic, and culturally responsive approaches to inform early childhood programmes and services, ultimately fostering children's growth as competent and confident learners.

Appendices

8.1 Appendix 1. Full Ethics Approval



MASSEY
UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

19/10/2022

Dear: Ellen Zeng

Re: Ethics Application - NOR 22/57 - Exploring Patterns of Young Children's Social, Emotional, Behavioural and Cognitive Characteristics, and their Correlation with Internal and External Determinants and School Readiness

Thank you for the above application that was considered by the Massey University Human Ethics Committee:

Ohu Matatika 2 at their meeting held on **Thursday, 22 September 2022**

On behalf of the Committee I am pleased to advise you that the ethics of your application are approved.

Approval is for three years. If this project has not been completed within three years from the date of this letter, reapproval must be requested.

If the nature, content, location, procedures or personnel of your approved application change, please advise the Secretary of the Committee.

Yours sincerely

A handwritten signature in blue ink, appearing to read 'C Johnson', on a light-colored background.

Professor Craig Johnson
Chair, Human Ethics Chairs' Committee and Director (Research Ethics)

Research Ethics Office, Research and Enterprise
Massey University, Private Bag 11 222, Palmerston North, 4442, New Zealand T 06 951 6841; 06 95106840
E humanethics@massey.ac.nz; animaethics@massey.ac.nz; gtc@massey.ac.nz

8.2 Appendix 2. Ethnicity Prioritisation in Current Research

The use of the multi-faceted indicators of a child's ethnicity data in the current research, based on the GUiNZs 54-month dataset, requires a thoughtful examination of ethnicity classification and prioritisation. This appendix discusses disparities associated with ethnicity, issues in ethnicity classification and prioritisation, common ethnicity prioritisation methods and the prioritisation of ethnicity data in the current research before using the available ethnicity data for the examination of children's capabilities across ethnicities in the thesis.

8.2.1 Introduction

As noted in the 3.4.2.1 'Ethnicity' literature review section, ethnicity refers to a group of people who share a common ancestry, history, traditions, and cultural traits such as values, beliefs, language, food, music, and other aspects of life (Cokley, 2007; Phinney, 1990; Statistics New Zealand, 2005). Ethnicity disparities, shaped by social, political, and historical contexts (Carr et al., 2017; Cokley, 2007; Khawaja et al., 2007; Yao et al., 2021), remain strongly linked to health and educational outcomes in New Zealand (Boven et al., 2020; Cameron & Poot, 2019; The Education Review Office, 2023). The origins of some of these disparities associated with ethnicity can be traced back to New Zealand's history. Specifically, the literature highlighted historical barriers, including social, policy and education system issues, that have led to inequities in accessing health and education services and underachieved education outcomes for Māori (Bishop et al., 2009; Scott, 2014; Smith, 2000). The disparities continue to shape

the inequities observed today (e.g., Lim et al., 2017; O'Sullivan, 2012; The Education Review Office, 2023).

Moreover, the use of ethnicity data, including its measurement, classification and prioritisation, has the potential to help or hinder social inequities (Cokley, 2007).

Measuring and classifying ethnicity data is important for monitoring health and educational outcomes, assessing discrimination and disadvantage, understanding community burdens and interventions, and improving outcomes and equity (Carr et al., 2017). However, the approach to ethnicity classification and prioritisation requires careful consideration.

Ethnicity prioritisation is a classification method that assigns individuals identified with more than one ethnic group to a single, mutually exclusive group according to a pre-determined hierarchy (Cormack & Robson, 2010). From the work by Cormack and Robson (2010), common methods of classifying ethnicity are total response, prioritised responses, and single/combination. Total response is a classification method that assigns individuals to each ethnic group they reported, which may result in 'overlapping' responses. Prioritised responses assign individuals to one ethnic group using priority criteria, including administrative-prioritisation prioritised by researchers and self-prioritisation prioritised by participants. The single/combination approach classifies individuals into either a single group if they are identified with only one ethnic group or a combination ethnic group if they are identified with multiple ethnic groups. These methods offer advantages and disadvantages concerning statistical

appropriateness, ethical considerations, and general usability (Boven et al., 2020; Yao et al., 2021; Yao et al., 2022). For example, the total response captures individuals' ethnicity identification more comprehensively than other methods by counting all identified ethnic groups. However, this non-mutually exclusive classification method often results in a total count of ethnic groups that exceeds the actual sample size, which introduces challenges for conducting statistical analysis. Additionally, there is a lack of guidelines on appropriately incorporating the total response method into statistical modelling (Boven et al., 2020; Ministry of Health, 2017). In sum, the use of each ethnicity classification and prioritisation method requires careful consideration from the researchers.

The issues of classifying ethnic responses in ways that facilitate data description and data analysis have challenged quantitative researchers, that include the high frequency of multiple ethnic affiliations in the population and the implication of categorising ethnicity for equity of policies, funding, research, and services (Aspinall, 2018; Callister et al., 2007; Mays et al., 2003; Perez & Hirschman, 2009). Regarding access to education and other resources, the classification of ethnicity data needs to be accurate, ethical, and consistent with affirmative action and anti-discrimination agendas (Carr et al., 2017; Kukutai & Callister, 2009; Yao et al., 2021). Additionally, ethnicity categorisation may impact the interpretations of education and health outcomes associated with ethnicity (Boven et al., 2020; Pomeroy, 2020; Rutkowski et al., 2017). A study by Boven et al. (2020) assessed the impact of three main ethnic groups' categorisation methods

on health outcomes in New Zealand and suggests that prioritised ethnicity categorisation may not be appropriate for policy-relevant research concerned with children, and a single/combination ethnic categorisation may be considered of greater use when describing and interpreting associations between ethnicity and outcomes.

Among mutually exclusive categorisations of ethnicity, administrative-prioritisation and single/combination ethnic classification are most commonly used in research (Education Counts, 2021; Ministry of Health, 2017; Yao et al., 2021). Administrative-prioritisation assigns individuals to mutually exclusive ethnic groups based on a prioritised hierarchy framework in the six ethnicity categories in Level 1 category level outlined by the Department of Statistics (1993): Māori, followed by Pacific Peoples, Asian, MELAA (Middle Eastern / Latin American / African) and Other ethnic groups except for European, and European. This approach is commonly used in education and health sections, likely due to the recognition of Māori rights as Treaty partners and the need to reduce groups with small sample sizes and large numbers of ethnicity categories (Yao et al., 2021). However, this approach may not provide the best indicator of respondents' self-identity where individuals are asked to select a single-ethnicity group or to self-prioritise their ethnic affiliation (Leather, 2009). A single/combination approach to ethnic classification may provide a more nuanced representation of overlapping ethnic groups, while still generating mutually exclusive groups suitable for statistical models (Cormack & Robson, 2010; Didham, 2005). According to Cormack and Robson (2010), the single/combination approach classifies respondents with one

ethnicity under that ethnic group, and those with multiple ethnicities into a combination group. It proves valuable for analysing the associations between ethnicity and outcomes. However, using a single/combination ethnicity approach with caution is advised due to potential limitations in sample size for certain combination ethnicities, leading to less informative outcomes, particularly with smaller datasets.

The ethnicity categorisation approach employed in research should be aligned with the research question and sociocultural context to ensure the accuracy, relevance, and authenticity of research outcomes (Carr et al., 2017). This is evident in the GUiNZs research. For example, the study from Jeffreys et al. (2022) did not strictly follow the three conventional ethnicity prioritisation methods but prioritised the child's ethnicity as Māori, Pacific Peoples and NZ European/Other (including New Zealanders) to align with the research purposes. This approach emphasised inequities in access to primary health care for children among different ethnic groups while ensuring Māori children, as the indigenous people in New Zealand, were meaningfully represented in the prioritisation. Other GUiNZs research has employed administrative-prioritisation of ethnicities to classify fathers', mothers', and mother-prioritised children's ethnicity by Statistic New Zealand prioritised-ethnicity guidelines (Jan et al., 2021; Neumann et al., 2021b; Reese et al., 2016; Rusten et al., 2019). Overall, researchers choose a tailored approach to prioritise ethnicities for the specific purposes of the research.

8.2.2 Current Research

In the current research, indicators of child ethnicity are used as an observed

grouping variable to examine the measurement invariance of indicators of social, emotional, language, and executive function capabilities at both the individual construct level and in developing a model of children's capabilities in which the interrelationship of these constructs is identified and described. To ensure due diligence in the research, it is important to assess whether the measures at the construct level and the model across ethnicities are equitably understood and reflect the population characteristics. By addressing the measurement invariance across ethnicity, the research can better account for fairness in assessment across ethnicities, which further helps inform equitable early childhood educational services in New Zealand.

In the current research, I propose to use administrative-prioritisation to re-classify mother-prioritised child's main ethnicities into broad ethnic groups, using the Statistics New Zealand (2005, updated 2023) Level 1 six ethnic categories. Notably, the selected MELAA with low responses ($n= 69$, less than 5% of the sample) observed are re-classified into 'Other ethnicities' in line with ethnicity classification references (e.g., Ministry of Health, 2017; Statistics New Zealand, 2005). I will then explore the potential to use single-ethnicity and single/combination categorisation of ethnic identities in the current analyses with reference to the available data. The question 'Which is the main ethnic group that the child identifies with?' in the GUiNZs 54-month data collection wave asked mothers to prioritise one to two of their child's main ethnicities from a list of thirty-three ethnicity options, corresponding to the 33 categories of ethnicities in Statistics New Zealand 'Level 3' ethnic level categories (Statistics New Zealand, 2005).

Following the re-classification of responses into the Stats NZ Level 1 categories, total responses were described, and both administrative-prioritised and single/combination classifications were applied.

Ethnicity prioritisation methods and associated cell counts for categories as observed in the 54-month data wave are reported in Appendix 2 Table 1. Unlike the administrative-prioritised method, the single/combination approach generates a 'dual ethnicities' group (i.e., comprising any child whose mother nominated the child as having two main ethnic groups that were classified into separate Statistics New Zealand Level 1 ethnicity classifications) and five single-ethnicity groups: Māori, followed by Pacific Peoples, then Asian, then Other ethnic groups besides European, and followed by European. The current research includes the group 'Other ethnicities' in measurement invariance tests at the construct- or model-levels, as it represents a group of ethnicities, enabling a broader representation of ethnic diversity without necessitating detailed analysis for each ethnic group. Additionally, the group 'dual ethnicities' is reported as a single group due to low numbers of observations of unique dual ethnicities combinations, with the largest unique group being children with dual Māori and European identities ($n = 246$). The combination of dual ethnicities is important for understanding New Zealand's bicultural society; however, it is unlikely to represent meaningful variables due to the lack of data for the detailed analysis. Therefore, the dual ethnicity group will not be included in tests of measurement invariance at the construct- or model-levels, and its omission may limit the

interpretation of the bicultural or multicultural backgrounds.

Appendix 2 Table 1

Ethnicity Categorisation Based on Mother-Prioritised Child's Ethnicity in the Current Research (n)

Classification methods	Statistic New Zealand Level 1 ethnic groups	Frequency	Sum	
Total responses	Māori	810	6625*	
	Pacific Peoples	866		
	Asian	746		
	Other ethnicities	764		
	European	3338		
Administrative-prioritisation	Māori	810	6156*	
	Pacific Peoples	796		
	Asian	732		
	Other ethnicities	706		
	European	3008		
Single/combination	Dual ethnicities	466	6156*	
	Single-ethnicity	Māori only		518
		Pacific Peoples only		729
		Asian only		670
		Other ethnicities only		661
		European only		3008

Note. *with 104 cases with missing values, including system missing, user missing, and residual categories (DK and References).

Other ethnicities include MELAA, and Other ethnicities, including 'New Zealander'. As mother-prioritised ethnicity allowed up to two responses, Total responses represent a count of Level 1 ethnic groups that exceed the respondent sample size. Administrative-prioritisation assigns individuals into one ethnic group following a prioritised sequence: Māori > Pacific Peoples > Asian > Other ethnic groups (including MELAA) except for European > European; the outcomes are mutually exclusive, and so the summed count aligns with the respondent sample size. The single/combination approach involves constructing mutually exclusive groups, with each category comprising either a single-ethnicity group label (such as Māori only) or a combination of dual ethnicities; the outcomes are mutually exclusive, and so the summed count aligns with the respondent sample size.

The approach to specifying the child's ethnicity for the research has been selected because of the potential utility for education policies, the research's ethical considerations, and the requirements of statistical analysis, including sample size. The advantage of this approach is that categorisation will facilitate statistical modelling. This approach may also support the accuracy of comparisons generated with national and cohort study datasets, which include ethnicity variables. However, the key limitation of

not employing dual ethnicities for statistical analysis may be overlooking the diversity that arises from dual ethnicities' affiliation. The impacts of this approach on the interpretation of results will be explored and discussed throughout the research process.

8.3 Appendix 3. Assessing Young Children's Social and Emotional Characteristics using the Prosocial Behaviour and Emotional Symptoms Subscales of Strengths and Difficulties Questionnaire

8.3.1 Introduction

Goodman (1997) designed the Strengths and Difficulties Questionnaire (SDQ) to support the investigation of mental, behavioural, and social functioning and development in childhood. Today, this measure is widely used as a behavioural screening tool among children 4-17 years of age (Goodman & Goodman, 2009; Stone et al., 2010). Versions of the SDQ vary with the age of children to be assessed, including parent/teacher-reported forms for children aged 4-17 and self-reported forms for children aged 11-17 (Goodman, 1997; Goodman et al., 2003). Importantly, in New Zealand, the parent-reported SDQ is part of the Before School Check (B4SC), with item-weighted total and subscale scores used to indicate children's emotional and behavioural strengths and difficulties (Richards et al., 2019). In addition, SDQ total and subscale scores have been applied to predict behavioural and psychosocial outcomes of general and clinical cohorts in New Zealand contexts (Clark et al., 2014; Cornforth et al., 2012; D'Souza et al., 2019b; Lambie & Krynen, 2017; Thompson et al., 2021).

The 25-item SDQ consists of five subscales, with four assessing behavioural Difficulties (i.e., conduct problems, hyperactivity, emotional symptoms, and peer problems) and one assessing behavioural Strengths (prosocial behaviour). Each SDQ item is endorsed by the respondent on a 3-point ordinal scale ('not true', 'somewhat

true', 'certainly true'), scored 0-2, with five positively-worded items among the Difficulties subscales reverse-scored such that higher scores indicate more significant behavioural difficulties. Items of each of the five subscales are summed, with total scores ranging from 0-10. The four Difficulties subscales may be combined to create a 'Difficulties' total score ranging from 0-40. Receiver Operating Characteristic (ROC) curves, assessing the sensitivity and specificity of parent-rated and teacher-rated Difficulties total scores, have indicated that SDQ scores may effectively discriminate between psychiatric and nonpsychiatric samples (Goodman, 1997).

While the SDQ is widely used for research and administrative purposes, its construct and predictive validity have long been debated. At a measurement level, concerns about the SDQ's construct validity arise from observations of weaker reliability and validity in its subscales compared to the total score, its unstable factor structure between samples, and a low structural consistency indicated by loading of 'peer problems' subscale items with both 'emotional symptoms' and 'prosocial' subscale items (Santiago, 2021; Stone et al., 2010). Given these concerns, evaluating the construct validity of these factors is important. Additionally, the labelling of 'Difficulties' can be problematic, as it may stigmatise children who score higher on these domains, potentially impacting their social functioning. Because the screening efficiency of SDQ varies across diagnostic groups, research has suggested that diagnostic conclusions based on SDQ should be avoided (Goodman et al., 2003; Goodman et al., 2000; Sanne et al., 2009). Moreover, it is essential to establish and use age-appropriate cut-off

values, as the SDQ cut-off scores established by Goodman (1997) for school-aged children disproportionately led to high numbers of preschoolers classified as in the borderline category, categorised by the total difficulties scores of 14-16 (Downs et al., 2012; Koglen et al., 2007). Hence, the current study addresses the key measurement issue of evaluating the appropriate factor structure of the parent-report SDQ among the current sample of children prior to school entry.

8.3.2 Validity of the Five-Factor Structure of the SDQ

Recent research into the factor structure of the 25-item SDQ has supported the five-factor model of the SDQ, which comprises the originally hypothesised/designed factors reflecting emotional symptoms, conduct problems, hyperactivity, peer problems and prosocial behaviours (Karlsson et al., 2022; Van Leeuwen et al., 2006; Van Roy et al., 2008). Van Leeuwen et al. (2006) supported five-factor solutions for parent- and teacher-reported SDQ in a community sample of children aged 4-8 years, although some studies show that confirmatory models additionally indicate the presence of method effects associated with the ten positively worded items (the five indicators of the prosocial behaviour subscale, and the five positively worded Difficulties items) (D'Souza et al., 2019c; Ellis et al., 2014; Karlsson et al., 2022; Kersten et al., 2014; McAloney-Kocaman & McPherson, 2017; Van Roy et al., 2008; Vugteveen et al., 2020). As noted by Kersten et al. (2018), the five subscales do not fit the Rasch model (a confirmatory approach to examining whether items belong to the subscales under investigation) among 51,251 preschoolers from the 2011 B4SC SDQ dataset in New

Zealand. Therefore, it is necessary to confirm the fit of individual factors and the overall factor structure of the SDQ.

8.3.3 Validation of the Parent-Reported SDQ in the GUiNZs

The SDQ is a key measure of children's behaviour in the Growing Up in New Zealand study (GUiNZs), with the parent-reported version administered at both 24- and 54-month assessment waves (D'Souza et al., 2019b). Using data from the 24-month assessment, the five-factor model of the SDQ was tested using confirmatory factor analysis (CFA) with robust diagonal weighted least squares estimation to account for the ordinal nature of indicators. A modified five-factor model (D'Souza et al., 2017; D'Souza et al., 2019c), incorporating a positive construal method effect (allowing all five positively-worded items in the Difficulties subscales to cross-load onto the Strengths/Prosocial subscale), was assessed for improvements in model fit. Results indicated that the modified five-factor model displayed an acceptable fit to the data (D'Souza et al., 2019b; Van de Looij-Jansen et al., 2011). Van de Looij-Jansen et al. (2011) also observed evidence for a 'positive construal' effect via cross-loading positively worded Difficulties items onto the Prosocial factor. This finding supports the validity of a modified five-factor model of the SDQ. However, it has yet to be established whether these results are also borne out in the 54-month dataset.

Within a CFA framework, the method effect indicated by D'Souza et al. (2019a) can be modelled in two additional ways. Researchers may acknowledge the systematic error associated with method effects by allowing correlation of the error terms associated

with affected indicators (i.e., positively worded items loading on Difficulties subscales). In this approach, error terms reflect measurement and systematic method-related error variance. Alternatively, where a greater understanding of these effects and the reliability of items as indicators are desired, researchers may obtain a more explicit assessment of method effects and better represent the co-variance structure by specifying a separate latent residual methods factor (Eid, 2000) as influencing all positively-worded indicator items and uncorrelated with any of the five theorised SDQ subscales. Specifying this latent methods factor allows researchers to explicitly model the residual variance associated with positively worded items that are variously explained by the subscale of interest, the methods factor, and residual error. This allows researchers to estimate the reliability of items as indicators of SDQ subscales, separate the error and method variance components, and calculate coefficients that refer to different proportions of explained variance.

8.3.4 Current Research

The current research aims to examine young children's social-emotional competence through a strengths-based lens to support a holistic view of child development prior to school entry in New Zealand. To support this aim, this report uses confirmatory models to assess the factor structure and measurement invariance of SDQ subscales among the sample of children participating in the GUiNZs 54-month assessment wave. Due to GUiNZs data availability (i.e., timing related to verifying reverse-scored items with the GUiNZs team), the present analyses focused on two

subscales of the SDQ that do not include positively worded Difficulties items. The current research focused on the prosocial behaviours and emotional symptoms subscales. Specifically, the factor structure of the prosocial behaviours and emotional symptoms subscales in SDQ in the current sample were confirmed to 1) examine the fit of the one-factor models of the prosocial behaviour and emotional symptoms subscales, 2) derive a more precise assessment of young children's social and emotional characteristics in the current sample rather than those implied by the measure's summed item-weighted scores (i.e., tau-equivalent models, in which all responses and items are considered equal indicators of a factor), and 3) enable assessment of measurement invariance across gender and ethnic subgroups. As prior evidence supports the discriminant validity of the prosocial behaviours and emotional symptoms factors (e.g., weak factor covariance, lack of cross-loading items, no shared methods effects), and the SDQ is typically scored as summed (item-weighted) scores for each subscale, the current study utilised one-factor congeneric models to test construct validity and derive factor scores for the prosocial behaviours and emotional symptoms subscales of the SDQ in the GUiNZs 54-month datasets. This also enables the assessment of the multi-group invariance of each of the two factors using congeneric models (i.e., the fit of all items as indicators of their hypothesised latent factor in isolation from other scale items or factors).

8.3.5 Method

8.3.5.1 Sample and Inclusion Criteria.

The 'standard' parent-reported SDQ (Goodman, 1997) was administered at the 54-month mother interviews. Parents were asked to report on the child's behaviour over the last six months. Children whose mothers participated in the parent-reported SDQ and responded to at least one item from the test were included in the current analysis.

8.3.5.2 Analysis Plan.

Item wording and number of responses endorsing each response option for the prosocial behaviours and emotional symptoms subscale item are presented in Appendix 3 Table 1. Descriptive analyses were conducted using SPSS version 28.0.1.1, and inferential models were generated using Mplus version 8.10. Initial descriptive statistics and checks were conducted to confirm that responses displayed adequate variability across response options for each item in the analysis sample. A minimum of $n \geq 5$ cases endorsing each response option was required for a response option to be included in the analysis. Where $n < 5$ cases endorsing a given item response were observed, cases endorsing this option were excluded from analysis, and the item was treated as dichotomous among the remaining cases. For brevity, item-level statistics are presented for the overall sample, with conclusions regarding response coverage for gender and ethnic groups reported in the text. Missing data in inferential models was handled using Maximum Likelihood Estimation.

Confirmatory Factor Analyses (CFA) were used to assess one-factor congeneric models of the prosocial behaviours and emotional symptoms subscales of the SDQ. Models were estimated using ordered categorical variable methodology, weighted least squares means-variance (WLSMV) estimation, and the Theta parameterisation as implemented in Mplus version 8.10. Acceptable fit was suggested by the combination rules with SRMR values less than or equal to 0.08 and CFI close to or greater than 0.95, or with an RMSEA less than or equal to 0.06 (Hu & Bentler, 1999).

Once an optimal factor structure for each subscale was identified in the sample overall, CFA was fitted for each group, and multi-group confirmatory factor analyses (MG-CFA) were used to assess the measurement invariance of these models across gender and ethnic groups. The current assessment of invariance tests includes configural and threshold invariance for ordered categorical variables, and the metric invariance could not be assessed due to issues of model identification with categorical indicators (Millsap, 2011). Multi-group invariance tests assessed the relative fit of nested models with increasingly restrictive equality constraints on the factor structure (configural/weak invariance) and on the factor loadings and item thresholds (threshold/strong invariance) across groups. Decrements in model fit were assessed with reference to changes (Δ) in fit indices with increasingly restrictive models, with critical values of Δ CFI $|0.01|$, SRMR ($\Delta < 0.02$ for configural vs. metric; $\Delta < 0.01$ metric vs. threshold), and RMSEA ($\Delta < 0.015$) used to evaluate whether model fit was significantly reduced as factor structure and item thresholds were constrained to be equal across

the gender and ethnic groups (Chen, 2007; Wang & Wang, 2019). Noticeably, these values are based on models with continuous indicators and associated estimators and are used due to limited criteria for assessing invariance for models employing ordinal indicators and the WLSMV estimator. Due to its sensitivity to sample size, changes in chi-square for increasingly restricted models, assessed using the DIFFTEST function, are reported but not considered in the current evaluation of model fit.

The configural invariance of each factor was assessed by constraining the factor structure to be consistent across groups, while the factor loadings and thresholds were freely estimated across groups. For identification, residual variances were fixed to one in the reference groups for gender and ethnic model comparisons (respectively, 'boys' and 'European'), with factor means fixed to zero across groups and factor variances freely estimated. The threshold invariance model was estimated by constraining factor loadings and thresholds to be equal across groups. For model identification, residual variances were fixed at one, and factor means were fixed at zero in the reference group, while factor means were freely estimated in the remaining groups. Where model evaluation indicated a lack of threshold invariance, nested partial invariance models were assessed. In this process, a backward approach to partial exact invariance testing was used by sequentially releasing parameters and re-testing model fit. Partial invariance was identified if acceptable levels of decline in model fit could be achieved where 25% of parameters (i.e., of the total number of thresholds/intercepts and factor

loadings) to be invariant across groups overall for reliably unbiased estimates to be obtained were considered invariant across groups (Asparouhov & Muthén, 2014a).

8.3.6 Results

6156 children whose mothers participated in the parent-reported SDQ and responded to at least one item from the test were included in the current analysis. Proportions of observed data by item and response option among the included sample are presented in Appendix 3 Table 1. There were very low proportions of missing data for any item (range 0.0-1.2% missing), likely reflecting the interview-based mode of data collection. The lowest endorsement for any response option within scale items was 1.5% (n = 89) of respondents who indicated it was ‘not true’ that their child was generally kind to younger children (SDQ17). Based on this observation, it was determined that there was adequate data to conduct CFA in the overall sample, modelling all items and response options.

Appendix 3 Table 1

Items of the Prosocial and Emotional Subscales in the SDQ With the Number of Response Endorsements by Item (n = 6156)

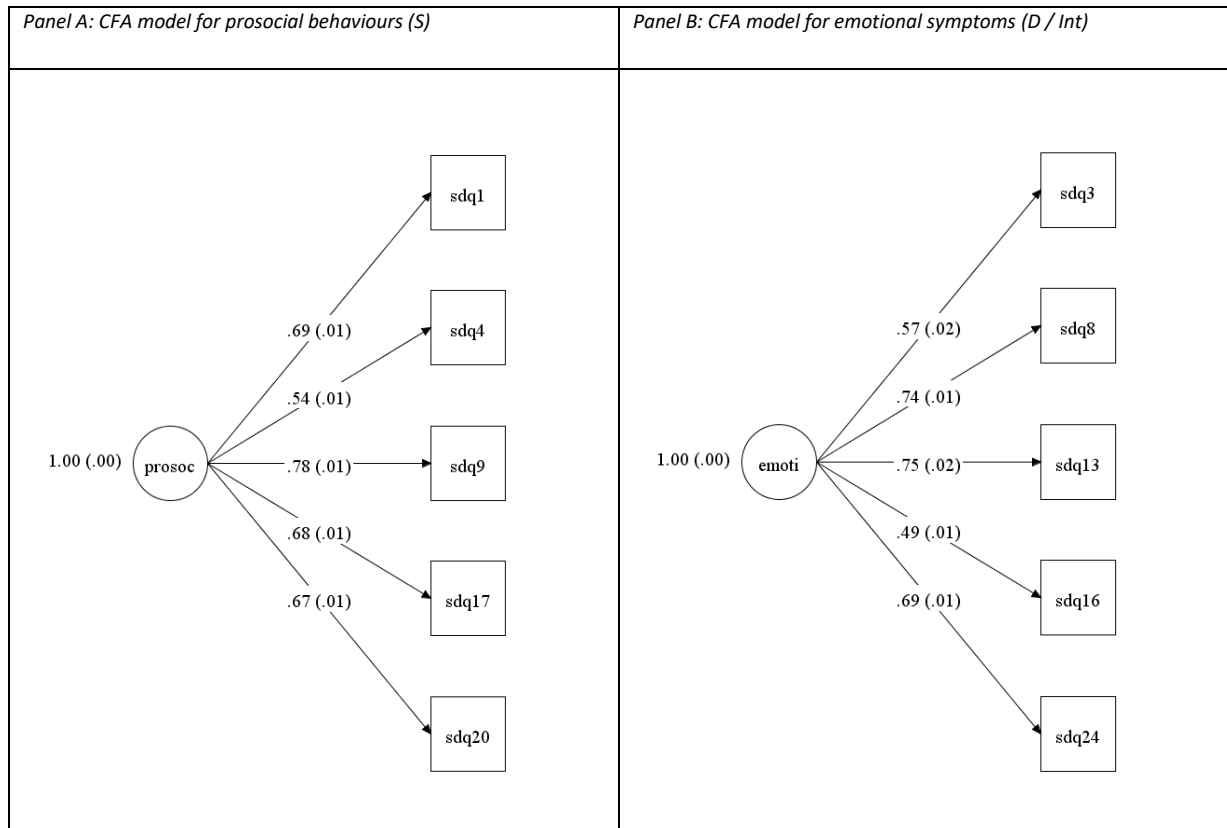
SDQ subscales	item#	Item stem	Missing	0 = Not true	1 = Somewhat true	2 = Certainly true
Prosocial behaviour	SDQ1	Considerate of other people's feelings	12	173	2633	3338
	SDQ4	Shares readily with other children (treats, toys, pencils, etc.)	<10	223	3438	2486
	SDQ9	Helpful if someone is hurt, upset, or feeling ill	15	142	1880	4119
	SDQ17	Kind to younger children	22	89	1638	4407
	SDQ20	Often volunteers to help others (parents, teachers, other children)	24	225	2484	3423
Emotional problems	SDQ3	Often complains of headaches, stomach-aches, or sickness	<10	4555	1274	319
	SDQ8	Many worries or often seems worried	13	4732	1233	178
	SDQ13	Often unhappy, down-hearted or tearful	<10	5246	791	110
	SDQ16	Nervous or clingy in new situations, easily loses confidence	18	2046	3059	1033
	SDQ24	Many fears, easily scared	12	3915	1928	301

Note. Items are administered in ascending order of SDQ item number.

Appendix 3 Figure 1 presents measurement models for the one-factor prosocial behaviours and emotional symptoms subscales in the sample overall. Appendix 3 Table 3 summarises model fit and associated modification indices.

Appendix 3 Figure 1

Measurement Models for Parent-Report Prosocial and Emotional Symptoms SDQ Items in the GUiNZs 54-Month Dataset



Note. Panel A represents the standardised estimates in the CFA model of the five-item prosocial behaviours subscale. Panel B represents the standardised estimates in the CFA model of the five-item emotional symptoms subscale in the SDQ.

Appendix 3 Table 2 presents the unstandardised coefficients for the factor models of the prosocial behaviours and emotional symptoms subscales. SDQ 1 ‘Considerate of other people’s feelings’ and SDQ 3 ‘Often complains of headaches, stomach-aches, or sickness’ were used as the reference items in the prosocial behaviours and emotional symptoms subscales, respectively.

Appendix 3 Table 2

Unstandardised Coefficients and Items of the Five Items One-Factor Models of the SDQ Prosocial Behaviour and Emotional Symptoms Subscales

		Model 1	Model 1 Gender		Model 1 Ethnic groups				
		Overall	Boys	Girls	European	Māori	Pacific	Asian	Others
n		6150	3161	2984	3007	518	729	669	661
Variables	Items								
Prosocial Behaviour (s)		B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)
SDQ1	Considerate of other people's feelings	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)
SDQ4	Shares readily with other children (treats, toys, pencils, etc.)	0.795(0.025)	0.665(0.045)	0.697(0.053)	0.682(0.048)	0.816(0.149)	0.821(0.130)	0.687(0.107)	0.697(0.111)
SDQ9	Helpful if someone is hurt, upset, or feeling ill	1.139(0.027)	1.354(0.099)	1.252(0.096)	1.366(0.103)	1.596(0.296)	1.419(0.209)	1.096(0.166)	1.196(0.202)
SDQ17	Kind to younger children	0.989(0.026)	0.882(0.060)	1.075(0.083)	0.915(0.064)	1.076(0.199)	1.440(0.222)	0.949(0.142)	0.925(0.144)
SDQ20	Often volunteers to help others (parents, teachers, other children)	0.976(0.025)	0.879(0.055)	1.011(0.075)	0.937(0.063)	1.196(0.202)	1.339(0.207)	0.797(0.116)	0.928(0.147)
Emotional Symptoms (D / Int)									
SDQ3	Often complains of headaches, stomach-aches, or sickness	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)
SDQ8	Many worries or often seems worried	1.289(0.045)	1.491(0.129)	1.673(0.145)	2.473(0.263)	1.298(0.246)	1.369(0.204)	1.266(0.270)	1.628(0.349)
SDQ13	Often unhappy, down-hearted or tearful	1.301(0.046)	1.503(0.135)	1.711(0.159)	1.785(0.196)	2.304(0.515)	1.585(0.258)	2.817(0.951)	1.045(0.213)
SDQ16	Nervous or clingy in new situations, easily loses confidence	0.847(0.036)	0.755(0.063)	0.831(0.072)	1.241(0.123)	0.935(0.176)	0.566(0.091)	0.600(0.158)	0.566(0.116)
SDQ24	Many fears, easily scared	1.209(0.043)	1.327(0.109)	1.422(0.118)	2.285(0.243)	1.482(0.275)	0.809(0.113)	1.004(0.206)	1.080(0.201)

Note. B represented the unstandardised beta coefficient, and se represented standard error. Model 1 confirms the original factor structure of the prosocial behaviour and emotional symptoms subscales.

Appendix 3 Table 3

Model Fit Indices for CFA and MG-CFA of the SDQ Prosocial Behaviour Subscale

Analysis	Model	Chi-Square	Free parameters	CFI	SRMR	RMSEA (90% C.I.), p	DIFFTEST
CFA by gender	CFA Overall (n =6150)	96.289(5), p< .001	15	0.988	0.023	0.054 (0.045, 0.064), p= .204	-
	Boys (n =3161)	71.591(5), p< .001	15	0.984	0.026	0.065 (0.052, 0.079), p< .05	-
	Girls (n =2984)	40.362(5), p< .001	15	0.988	0.024	0.049 (0.035, 0.063), p= .531	-
MG-CFA by gender	Configural	110.128(10), p< .001	30	0.986	0.025	0.057 (0.048, 0.067), p= .104	-
	Scalar	102.510(18), p< .001	22	0.988	0.026	0.039 (0.032, 0.047), p= .992	14.131(8), p= .078
CFA by ethnicity	European (n =3007)	47.196(5), p< .001	15	0.991	0.022	0.053 (0.040, 0.067), p= .334	-
	Māori (n =518)	13.548(5), p< .05	15	0.984	0.033	0.057(0.021, 0.095), p= .316	-
	Pacific (n =729)	36.800(5), p< .001	15	0.944	0.041	0.093(0.067, 0.123), p< .01	-
	Asian (n =669)	13.872(5), p< .05	15	0.986	0.029	0.052(0.020, 0.085), p= .414	-
	Other (n =661)	15.951(5), p< .01	15	0.987	0.030	0.058(0.027, 0.090), p= .300	-
MG-CFA by ethnicity	Configural	122.758(25), p< .001	75	0.987	0.028	0.059(0.049, 0.070), p= .069	-
	Scalar	229.181(57), p< .001	43	0.977	0.035	0.052 (0.045, 0.059), p= .308	118.199(32), p< .001
	¹ Partial scalar	182.762(55), p< .001	45	0.983	0.033	0.046(0.038, 0.053), p= .831	77.305(30), p< .001

Note. Confirmatory Factor Analyses (CFA) and Multi-group confirmatory factor analyses (MG-CFA) were conducted using ordered categorical variable methodology using the Theta parameterisation and weighted least squares means-variance (WLSMV) estimation as implemented in Mplus v8.10; SRMR = standardised root mean squared residual; RMSEA = root mean square error of approximation; CFI = comparative fit index; DIFFTEST = Chi-Square difference test for nested models. ¹ represents freeing the 2nd threshold in 'SDQ20 Often volunteers to help others (parents, teachers, other children)' for both European and Pacific People's groups.

Appendix 3 Table 4

Model Fit Indices for CFA and MG-CFA of the SDQ Emotional Symptoms Subscale

Analysis	Model	Chi-Square	Free parameters	CFI	SRMR	RMSEA (90% C.I.), p	DIFFTEST
	CFA Overall (n =6150)	143.421(5), p< .001	15	0.974	0.032	0.067(0.058, 0.077), p< .01	-
CFA by gender	Boys (n =3161)	58.915(5), p< .001	15	0.980	0.029	0.058(0.046, 0.072), p= .135	-
	Girls (n =2984)	90.990(5), p< .001	15	0.967	0.036	0.076(0.063, 0.090), p< .01	-
MG-CFA by gender	Configural	149.766(10), p< .001	30	0.974	0.032	0.067(0.058, 0.077), p< .01	-
	Scalar	136.170(18), p< .001	22	0.978	0.033	0.046(0.039, 0.054), p= .793	12.998(8), p= .11
CFA by ethnicity	European (n =3007)	60.031(5), p<.001	15	0.971	0.036	0.060(0.047, 0.075), p= .092	-
	Māori (n =518)	6.688(5), p= .24	15	0.997	0.022	0.026(0.000, 0.070), p= .775	-
	Pacific (n =729)	29.039(5), p< .001	15	0.973	0.035	0.081(0.054, 0.111), p< .05	-
	Asian (n =669)	27.479(5), p<.001	15	0.926	0.048	0.082(0.054, 0.113), p< .05	-
	Other (n =661)	7.748(5), p= .170	15	0.993	0.026	0.029(0.000, 0.066), p= .792	-
MG-CFA by ethnicity	Configural	131.817(25), p< .001	75	0.974	0.036	0.062(0.052, 0.072), p< .05	-
	Scalar	324.142(57), p< .001	43	0.935	0.054	0.065(0.058, 0.072), p< .001	194.170(32), p< .001
	¹ Partial scalar	271.152(57), p< .001	43	0.948	0.051	0.058(0.051, 0.065), p< .05	147.623(32), p< .001
	² Partial scalar	243.686(56), p< .001	44	0.954	0.049	0.055(0.048, 0.062), p= .126	122.930(31), p< .001
	³ Partial scalar	204.183(55), p< .001	45	0.964	0.044	0.049(0.042, 0.057), p= .551	88.020(30), p< .001
	⁴ Partial scalar	184.525(54), p< .001	46	0.968	0.042	0.047(0.039, 0.054), p= .772	69.921(29), p< .01

Note. Confirmatory Factor Analyses (CFA) and Multi-group confirmatory factor analyses (MG-CFA) were conducted using ordered categorical variable methodology using the Theta parameterisation and weighted least squares means-variance (WLSMV) estimation as implemented in Mplus v8.10; SRMR = standardised root mean squared residual; RMSEA = root mean square error of approximation; CFI = comparative fit index; DIFFTEST = Chi-Square difference test for nested models. 1 represents freeing factor loading of SDQ3 for the Pacific group. 2 represents freeing factor loading of SDQ3 and SDQ13 for the Pacific group. 3 represents the freeing factor loading of SDQ3 for the Pacific group, and SDQ13 for both the Pacific and Asian groups. 4 represents the freeing factor loading of SDQ3 for the Pacific group, SDQ13 for both the Pacific and Asian groups and SDQ8 for the Māori group.

8.3.6.1 Prosocial Behaviours.

As indicated in Appendix 3 Table 3, the one-factor CFA model of the *Prosocial Behaviours* provided a good fit to the data for the sample overall, as well as the groups by gender. When conducting MG-CFA, the configural model similarly provided a good fit. Compared to the configural invariance model, constraining thresholds to be equal across groups did not significantly decrease model fit on indicators of CFI, SRMR or RMSEA [$\Delta\text{CFI} = 0.002$; $\Delta\text{SRMR} = 0.001$]. Overall, indicators suggest the congeneric Prosocial factor displayed strong invariance across gender groups. Examining CFAs by ethnicity, results in Appendix 3 Table 3 also show that the one-factor model fit the data well for five single-ethnicity groups, but less than optimal model fit for children identified as being of Pacific ethnicity in the 54-month dataset, for whom values of SRMR, and CFI were acceptable; however, RMSEA was above the cut-off. When assessing the MG-CFA of the one-factor model across all single-ethnicity groups, the initial configural model provided an acceptable fit (CFI = 0.974, SRMR = 0.036, RMSEA=0.062). Compared to the configural invariance model, full-threshold invariance of the prosocial behaviour factor by child ethnicity was not supported. Investigating partial invariance models using a backward stepwise strategy, results indicated that freeing the 2nd threshold in 'SDQ20 Often volunteers to help others (parents, teachers, other children)' for both European and Pacific People's groups, resulted in a nested partial threshold invariance model, which did not significantly decrease indicators of

model fit compared to the configural model: $\Delta CFI = -0.004$; $\Delta SRMR = 0.005$; $\Delta RMSEA = -0.013$.

8.3.6.2 Emotional Symptoms.

As indicated in Appendix 3 Table 4, the one-factor model for the *Emotional Symptoms* provided a good fit for the data overall. Inspection of CFA models by gender group similarly indicated good model fit. Assessing measurement invariance models, the configural model provided a good fit across gender groups. Further, compared to the configural invariance model, constraining thresholds to be equal across gender groups did not significantly decrease model fit on indicators of CFI, SRMR or RMSEA: $\Delta CFI = 0.004$; $\Delta SRMR = 0.001$]. Overall, indicators suggest the congeneric emotional symptoms factor displayed strong invariance across gender groups.

Model fit indicators in Appendix 3 Table 4 for CFA models by child ethnicity suggest acceptable model fit indices for four of the single-ethnicity groups except Asian children, with $CFI = 0.926$, lower than 0.95 and $RMSEA = 0.082$, above 0.06. Model modification indices (M.I.) indicate specification of the correlation errors between SDQ24 with SDQ16 =17.176 may improve the model fit. However, an MG-CFA assessing a model of configural invariance provided an acceptable fit to the data. The scalar model did not provide a good fit for the data ($CFI=0.935$, $SRMR=0.054$, $RMSEA=0.065$). Decrements in model fit indices showed that $\Delta CFI=-0.029$, $\Delta SRMR=0.010$, and $\Delta RMSEA=0.000$, indicating a partial threshold invariance across

children's single-ethnicity groups. A partial threshold invariance test was conducted based on the acceptable fit of the configural model. Assessing models of partial threshold invariance indicated 4 steps specifications based on M.I. by first assessing the partial threshold invariance model freeing factor loading of SDQ3 for the Pacific group, second model freeing factor loading of SDQ3 and SDQ13 for the Pacific group, third model freeing factor loading of SDQ3 for the Pacific group, and SDQ13 for both the Pacific and Asian groups, and last model specification was freeing factor loading of SDQ3 for the Pacific group, SDQ13 for both the Pacific and Asian groups and SDQ8 for the Māori group. Compared to the configural invariance model, the partial invariance model did not significantly decrease model fit on indicators of CFI, SRMR or RMSEA [$\Delta\text{CFI} = -0.006$; $\Delta\text{SRMR} = 0.006$; $\Delta\text{RMSEA} = -0.015$]. Overall, given that 4% of model parameters were observed non-invariant within all ethnic groups in this model, we concluded that the emotional symptoms factor displayed partial threshold invariance across five single-ethnicity groups.

8.3.7 Conclusion

Results suggest partial invariance of the prosocial behaviours and emotional symptoms factors of the SDQ across the child's gender and ethnic groups at the 54-month assessment wave. The high reliability of items within each latent subscale factor score may individually provide interpretable representations of child strengths and difficulties. MG-CFA of the two one-factor models evaluated whether items of the SDQ and associated item-weighted or factor-weighted scores may equally represent the

same underlying constructs across gender and ethnic groups in this population. The use of factor scores provides a more robust alternative to scale scores by giving more weight to items with lower measurement errors, thus offering partial implicit control for measurement errors (Morin et al., 2016; Morin & Marsh, 2015). Furthermore, the limitations of the MG-CFA approach, such as the detection of potentially trivial differences and impact on power when considering more than one intersection group, have driven the use and development of approximation methods, such as the Alignment method. Methods, such as the covariate approach, specifically support the examination of MI for model indicators at the model level, which helps assess the impact of measurement non-invariance on model indicators.

8.4 Appendix 4. Factor Structure of the Affect Knowledge Test (AKT): Application to Young Children's Emotion Knowledge and Emotional Expression in the New Zealand Context

8.4.1 Introduction

The Affect Knowledge Test (AKT), designed to assess preschoolers' expressive and receptive emotional knowledge (Denham, 1986), is widely used as an age-appropriate measurement tool to assess 2.5- to 5-year-old children's recognition of emotions (i.e., happy, sad, angry, and afraid) and understanding of situational emotion knowledge (Bassett et al., 2012; Sette et al., 2015). The AKT includes both a shortened version (AKT-S) with 4-item expressive/verbal emotion recognition tasks (happy, sad, angry, and afraid) and a full version assessing 8-item emotion recognition tasks and 20-vignette situational emotional knowledge (8 vignettes for stereotypical and 12 vignettes for non-stereotypical) (Denham, 1986; Denham et al., 2015). Strengths of the AKT include its acknowledgement of young children's emotional knowledge and its ability to predict social-emotional learning and relevant skills (Domitrovich et al., 2007; Shields et al., 2001). Additionally, the test is relatively easy and quick to administer, compared to other tests of emotion recognition, including the Pearlman Emotional Knowledge Task and the Test of Emotion Comprehension (Pons et al., 2004) and the Emotion Matching Task (Morgan et al., 2010), due to little verbalisation and its play-based characteristics.

The full AKT consists of two subtests: *emotion recognition* and *situational emotional knowledge*. The 8-item tasks testing recognition of four emotions (happy, sad, angry and afraid) — verbally naming these four facial expressions (expressive knowledge) and then pointing to the four emotions non-verbally (receptive knowledge) – aims to measure young children’s understanding of emotions by nonverbally identifying facial expressions depicted on puppets and verbally naming the emotion (Denham, 2006; Machado et al., 2012). The expressive emotion recognition subscale is administered in-person only so that children’s verbal reactions to expressive stimuli can be recorded (Denham et al., 2020). Situational emotional knowledge is examined through vignettes presented in two subtests: stereotypical and non-stereotypical situation knowledge. The stereotypical situation knowledge test consists of 8 vignettes illustrating standard facial and vocal expressions of emotions, such as feeling afraid during a nightmare, in which the child is asked to place a facial expression on the puppet to reflect the puppet’s feeling (Denham et al., 1994). The non-stereotypical situational knowledge test comprises 12 vignettes – divided into two sets of 6 vignettes each. One set focuses on positive versus negative emotions (e.g., happy, or sad when attending preschool), while the other set deals with both negative emotions (e.g., anger at or fear of a sibling’s aggression). This test evaluates children’s situational emotional knowledge in identifying the feelings of others that may differ from their own (Denham, 1986, 2006). Children received two points for correctly identifying emotion in any

section of the measure and one for identifying the correct valence but not the correct emotion (e.g., sad for afraid).

Previous studies have shown acceptable internal consistency, test-retest validity, and concurrent and predictive reliabilities of the AKT (Cutting & Dunn, 2002; Denham et al., 2003; Denham et al., 2002; Denham & Couchoud, 1990a). For example, a study by Cutting and Dunn (2002) indicated the sum scores of the stereotypical situation knowledge with Cronbach's alpha equal to .73, representing moderate to good internal consistency of the AKT among 141 preschoolers, with children scoring higher on the AKT being more sensitive to teacher criticism. Shields et al. (2001) used the AKT to assess emotional understanding in 49 children who attended a Head Start school in New England in a short-term longitudinal study. The study found that children's emotional competence, measured by emotion regulation early in the school year (Time 1) and emotion understanding during the winter months (Time 2), predicted school adjustment at the end of the school year (Time 3) after controlling for the potential confounds of behaviour problems and verbal abilities.

Emotions are sometimes classified as positive or negative related to desired states of being, although this type of classification is not universally accepted. Nonetheless, previous research has shown that children understand positive emotions like happiness earlier than negative emotions like fear and surprise (Izard, 1994). Analyses of responses to the six emotions (happy, sad, angry, fear, surprise, and disgust) revealed that the frequency of incorrect identification of both positive (e.g., happiness) and

negative emotions exhibited consistent patterns (Widen & Russell, 2003). This reflects lower task difficulty associated with discriminating between positive emotions (which comprise relatively distinct feature configurations) and negative emotions (which share more features), seen in all ages, from children to adults (Lawrence et al., 2015; Serrat et al., 2020).

8.4.2 Construct Validity of the AKT

Despite emotion knowledge encompassing two distinct dimensions (i.e., *emotion recognition* and *situational emotional knowledge*) and the variation of emotion knowledge development at different ages, research commonly focuses on one domain or a unidimensional construct of emotion knowledge (Izard et al., 2001; Miller et al., 2005). Research using the AKT as a unidimensional construct typically aggregates emotion recognition and situational knowledge subscale scores to create an overall index of total emotion knowledge.

To examine preschoolers' emotional recognition at 54-month assessment, the Growing Up in New Zealand study (GUiNZs) adapted the shortened version of AKT (AKT-S) to assess young children's emotional knowledge. To avoid previously observed ceiling effects among children aged 54 months Denham (2006), additional face stimuli expressing 'negative' emotions of disgust and surprise were added to the emotion recognition task. It is important to note that some researchers have asserted that surprise encompasses both positive and negative aspects and is differentiated by the degree of alignment between the surprising event and its conduciveness to one's goals

(Noordewier & Breugelmans, 2013). In the adapted AKT-S in the GUiNZs, the six facial expressions of emotions (happy, surprise, sad, angry, scared, and disgusted/yucky) were adapted from the Ekman et al. (1969) stimulus set. Responses were scored concerning accuracy and emotional valence, with '0' indicating the child's response was an incorrect emotion with the opposite emotional valence from the correct response, '1' an incorrect emotion within the same emotional valence, and '2' the correct emotion or acceptable synonym. A total score from the task responses was calculated as a summed score representing the child's performance on the modified AKT-S.

8.4.3 Current Research

The current study aims to assess the factor structure of the adapted/modified AKT-S in the GUiNZs 54-month dataset. It aims to validate the scoring methods employed, enhance the interpretive value of the results, and explore whether the factor model of emotion knowledge in preschoolers remains consistent across gender and major ethnic groups in the New Zealand population. Initial Confirmatory Factor Analyses were conducted using the AKT data to investigate the one-factor model of emotion recognition in the sample overall and by observed sub-group (i.e., gender, ethnicity). Multi-Group Confirmatory Factor Analysis (MG-CFA) was employed to assess the measurement invariance of the optimal model across the observed groups.

8.4.4 Method

8.4.4.1 Sample and Inclusion Criteria.

The AKT was administered at the 54-month mother interviews. Children's responses to the randomly presented face cards (happy, sad, scared, angry, surprised, and disgusted) are based on their emotional knowledge. The children who participated in the adopted version of AKT and responded to at least one item from the test were included in the current analysis.

8.4.4.2 Analysis Plan.

Confirmatory factor analyses and measurement invariance analyses were conducted using ordered categorical variable methodology and weighted least squares means-variance (WLSMV) estimation as implemented in Mplus version 8.10. A one-factor CFA model based on theory and prior research was evaluated with reference to indicators of model fit, and to modification indices, where model fit was deemed inadequate. An acceptable fit was suggested with reference to Hu & Bentler (1999) combination rules with SRMR values less than or equal to 0.08 and CFI close to or greater than 0.95 or with an RMSEA less than or equal to 0.06.

Once an acceptable model of the observed data was identified, nested MG-CFA were used to assess whether model fit was significantly reduced as the factor structure and the factor loadings (λ) and item thresholds (τ) were constrained to be equal across the child's gender and ethnic groups (Chen, 2007; Wang & Wang, 2019; Wu & Estabrook, 2016). Decrements in model fit were assessed regarding changes (Δ) in fit

indices with increasingly restrictive models. Changes in model fit indices indicating measurement non-invariance from studies using Maximum Likelihood (ML) estimators and continuous indicators for unequal samples with group $N > 300$ suggest a decline in CFI greater than -0.010 , accompanied by $\Delta RMSEA \geq 0.015$ or $\Delta SRMR \geq 0.030$, indicating non-invariance of measures (Chen, 2007). Noticeably, these values are based on models with continuous indicators and associated estimators and are used due to limited criteria for assessing invariance for models employing ordinal indicators and the WLSMV estimator. Due to its sensitivity to sample size (Bollen, 1989; Tucker & Lewis, 1973), changes in chi-square with increasingly restricted models, assessed using the DIFFTEST function, are reported but not considered in the current evaluation of model fit.

8.4.5 Results

8.4.5.1 Item Endorsement.

5605 children who participated in the AKT and responded to at least one item from the test were included in the current analysis. Proportions of the observed data by item and response option among the included sample are presented in Appendix 4 Table 1. The proportion of missing data for any item was low, with only 2 missing data points observed among the included sample. The lowest observed response score for any AKT item was displayed by 3.9% ($n=221$) of respondents, who responded with an incorrect emotion of the opposite emotional valence when identifying the 'happy' face. Based on

these observations, a CFA model of all items and response options among participants who responded to the AKT was conducted.

Appendix 4 Table 1

Items and the Number of Responses Endorsements by Item in the Included Sample (n = 5605)

item#	Emotion displayed	Missing	(0 = incorrect emotion with the opposite emotional valence)	(1 = incorrect emotion within the same emotional valence)	(2= correct emotions or acceptable synonyms)
AKT2	Happy	<10	221	304	5080
AKT3	Surprised	<10	2159	2460	986
AKT4	Sad	<10	401	427	4777
AKT5	Angry	<10	241	842	4522
AKT6	Scared	<10	1764	2953	886
AKT7	Yucky/disgusted	<10	2143	2086	1376

8.4.5.2 Confirmatory Factor Analysis.

Appendix 4 Table 2 presents unstandardised parameter estimates of factor loadings and standard errors (M1 to M3) and M3 for each gender and single-ethnicity group for the AKT. Model fit indices for the CFA results are summarised in Appendix 4 Table 3. The model fit indicators for the one-factor model (M1) did not indicate a good fit to the data for the sample overall, as the RMSEA was greater than 0.06, and the CFI fell below 0.95. Additionally, M1 did not demonstrate a good fit for subgroups, including girls and single-ethnicity groups such as Pacific and European. Inspections of modification indices suggested that allowing a residual covariance between two items, AKT7 'Yucky/Disgusted' and AKT6 'Scared', may improve model fit for the overall sample (M.I.= 86.739) and specific subgroups, including girls (M.I. = 62.323), Pacific (M.I.=17.796) and European (M.I.= 20.943). Re-estimation of the revised model (M2)

improved model fit, whereas M2 did not provide a good fit to the Māori and Asian children, with the RMSEA greater than 0.06 and CFI below 0.95. Allowing an additional residual covariance between two items, AKT6 'Scared' and AKT3 'Surprised' for Māori (M.I.= 24.694) and Asian (M.I.= 32.356), may provide a better fit to these two subgroups. Estimation of the revised model (M3), accounting for the residual covariances between items AKT7 'Yucky/Disgusted' and AKT6 'Scared', AKT6 'Scared' and AKT3 'Surprised', indicated a better fit to models for the overall sample and all subgroups, especially Māori and Asian. Hence, Model 3 was selected as the baseline model for testing MG-CFA across the children's gender and single-ethnicity groups.

8.4.5.3 Multi-Group Confirmatory Factor Analysis.

Appendix 4 Table 4 demonstrates the measurement invariance results of M3 across the children's gender and single-ethnicity groups.

Gender. Results indicated that the configural invariance model provided a good fit to the data when held constant across gender groups, and the scalar invariance model also provided a good fit. Compared to the configural invariance model, constraining factor loadings and thresholds to be equal across groups did not significantly decrease model fit on any indicator [Δ CFI = -0.002; Δ SRMR = 0.001; Δ RMSEA = -0.01]. Overall, indicators suggest this model (M3) displayed strong invariance across children's gender groups.

Single-ethnicity groups. Model fit indices indicated that the configural model provided a good fit to the data for comparisons by children's single-ethnicity groups. The scalar model also provided a good fit of the data, whereas decrements in model fit indices showed that $\Delta CFI = -0.029$, $\Delta SRMR = 0.010$ and $\Delta RMSEA = 0.000$, indicating a partial threshold invariance across children's single-ethnicity groups. To assess whether a partial threshold invariance of categorical AKT indicators, a forward estimation approach was adopted, with each of the six-factor indicators fixed to be equal by using the European as a reference group and testing 'one-item at a time' and the associated decline in model fit assessed against the configural model. Based on the inspection of model fit indices, scalar model fit indices improved by freeing the first threshold for AKT5 in Pacific children (M.I. = 11.322). Assessment of decrements in the partial threshold invariance model [$\Delta CFI = -0.014$; $\Delta SRMR = 0.008$; $\Delta RMSEA = -0.07$] indicated a significant decline in model fit compared to the configural model by children's single-ethnicity groups.

Appendix 4 Table 2

Items and Factor Loadings of the One-Factor CFA Models of the 6 Interviewer-Observed AKT Items in the GUIiNZs 54-Month Dataset

Model		M1	M2	M3	M3 Gender		M3 Ethnic Groups				
		Overall	Overall	Overall	Boys	Girls	European	Māori	Pacific	Asian	Others
n		5605	5605	5605	2851	2751	2849	461	650	567	555
item#	Emotion displayed	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)
AKT2	Happy	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)
AKT3	Surprised	0.58(0.02)	0.58(0.02)	0.51(0.02)	0.31(0.04)	0.24(0.05)	0.45(0.09)	0.29(0.08)	0.27(0.06)	0.38(0.08)	0.43(0.16)
AKT4	Sad	0.88(0.03)	0.88(0.03)	0.88(0.03)	0.74(0.11)	0.58(0.13)	0.71(0.16)	0.82(0.23)	0.58(0.14)	0.94(0.23)	0.88(0.40)
AKT5	Angry	0.72(0.03)	0.72(0.03)	0.73(0.03)	0.52(0.07)	0.39(0.08)	0.45(0.09)	0.73(0.22)	0.70(0.19)	0.73(0.15)	0.45(0.17)
AKT6	Scared	0.58(0.02)	0.51(0.02)	0.42(0.02)	0.26(0.04)	0.18(0.04)	0.26(0.06)	0.26(0.08)	0.27(0.06)	0.35(0.08)	0.40(0.15)
AKT7	Yucky/disgusted	0.56(0.02)	0.49(0.02)	0.48(0.02)	0.28(0.04)	0.23(0.04)	0.40(0.07)	0.35(0.09)	0.25(0.06)	0.32(0.07)	0.29(0.12)
	AKT6<-> AKT7	-	0.14(0.02)	0.17(0.02)	0.16(0.03)	0.24(0.03)	0.18(0.03)	0.26(0.06)	0.23(0.05)	0.18(0.06)	0.15(0.06)
	AKT3<->AKT6	-	-	0.15(0.02)	0.18(0.03)	0.18(0.03)	0.17(0.03)	0.29(0.06)	0.10(0.06)	0.33(0.05)	0.01(0.07)

Note. The overall sample includes children with at least one item in the AKT observation task. B represented the unstandardised beta coefficient, and se represented the standard error. M1 CFA of one-factor model; M2 one-factor model indicates additional covariance between AKT7 Yucky/Disgusted and AKT6 Scared; M3 one-factor model indicates additional covariance between AKT6 Scared face and AKT3 Surprised based on M2.

Appendix 4 Table 3

Model Fit Indices for One-Factor Models of the Six-Item AKT Administered in the GUIiNZs 54-Month Dataset

Model		Chi-Square	N of free parameters	CFI	SRMR	RMSEA
M1	Overall (n = 5605)	276.243 (9), p < .001	18	0.945	0.047	0.073 (0.066, 0.08), p < .001
	Boys (n = 2851)	118.723 (9), p < .001	18	0.958	0.042	0.065 (0.055, 0.076), p < .01
	Girls (n = 2751)	162.487 (9), p < .001	18	0.930	0.056	0.079 (0.068, 0.090), p < .001
	European (2849)	75.167 (9), p < .001	18	0.918	0.049	0.051 (0.041, 0.062), p = .428
	Māori (461)	50.397 (9), p < .001	18	0.914	0.069	0.10 (0.074, 0.128), p = .001
	Pacific (650)	59.061 (9), p < .001	18	0.965	0.053	0.093 (0.071, 0.116), p = .001
	Asian (567)	71.491 (9), p < .001	18	0.929	0.068	0.111 (0.088, 0.135), p < .001
	Other (555)	14.697(9), p= .100	18	0.981	0.038	0.034(0.000,0.064), p= .787
M2	Overall (n = 5605)	193.871 (8), p < .001	19	0.962	0.038	0.064 (0.057, 0.072), p < .001
	Boys (n = 2851)	94.93 (8), p < .001	19	0.967	0.036	0.062 (0.051, 0.073), p < .01
	Girls (n = 2751)	103.497(8), p < .001	19	0.956	0.042	0.066 (0.055, 0.077), p < .001
	European (2849)	55.08 (8), p < .001	19	0.941	0.04	0.045 (0.035, 0.057), p = .724
	Māori (461)	41.37 (8), p < .001	19	0.93	0.06	0.095 (0.068, 0.125), p = .004
	Pacific (650)	43.12 (8), p < .001	19	0.975	0.044	0.082 (0.059, 0.107), p = .012
	Asian (567)	67.89 (8), p < .001	19	0.932	0.066	0.115 (0.091, 0.141), p < .001
	Other (555)	8.959(8), p= .345	19	0.997	0.030	0.015(0.000,0.053), p= .928
M3	Overall (n = 5605)	100.713(7), p < .001	20	0.981	0.026	0.049 (0.041,0.058), p= .567
	Boys (n = 2851)	47.598 (7), p < .001	20	0.985	0.025	0.045 (0.033, 0.058), p = .724

Model		Chi-Square	N of free parameters	CFI	SRMR	RMSEA
M3	Girls (n = 2751)	57.763 (7), p < .001	20	0.977	0.029	0.051 (0.040, 0.064), p = .402
	European (2849)	25.374 (7), p < .001	20	0.977	0.023	0.030 (0.018, 0.043), p = .994
	Māori (461)	18.09 (7), p < .012	20	0.977	0.038	0.059 (0.026, 0.093), p = .291
	Pacific (650)	41.688 (7), p < .001	20	0.976	0.041	0.087 (0.063, 0.114), p = .007
	Asian (567)	36.787 (7), p < .001	20	0.966	0.048	0.087 (0.060, 0.115), p = .013
	Other (555)	8.977 (7), p = .254	20	0.993	0.030	0.023(0.000, 0.06), p = .868

Note. The overall sample includes children with one or more items in the AKT observation task. CFA = Confirmatory Factor Analysis; SRMR = Standardised root mean squared residual; RMSEA = root mean square error of approximation; CFI = comparative fit index. M1 CFA of one-factor model; M2 one-factor model indicates additional covariance between AKT7 Yucky/Disgusted and AKT6 Scared would improve model fit (M.I. =86.739; revised model fit: SRMR=0.038, CFI=0.962, RMSEA= 0.064 (0.057, 0.072), p< .001); M3 one-factor model indicates additional covariance between AKT6 Scared face and AKT3 Surprised based on M2 (M.I. =96.144; revised model fit: SRMR=0.026, CFI=0.981, RMSEA= 0.049 (0.041, 0.058), p= .567).

Appendix 4 Table 4

Model Fit Indices for the One-factor AKT (M3) with Multi-Group Invariance Comparison by Children's Gender and Ethnicity

Subgroup	Model	Chi-Square	Free parameters	CFI	SRMR	RMSEA (90% CI)	Chi-Square DIFFTEST
Gender	Configural	105.472 (14), $p < .001$	40	0.981	0.027	0.048 (0.040, 0.057), $p = 0.609$	-
	Scalar	123.088 (24), $p < .001$	30	0.979	0.028	0.038 (0.032, 0.045), $p = 0.998$	25.263(10), $p = .005$
Single-ethnicity groups	Configural	130.477 (35), $p < .001$	100	0.975	0.032	0.052(0.042, 0.061), $p = 0.360$	-
	<i>Metric*</i>	<i>168.377(55), $p < .001$</i>	<i>80</i>	<i>0.971</i>	<i>0.037</i>	<i>0.045(0.037, 0.053), $p = 0.847$</i>	<i>43.047 (20), $p = .002$</i>
	Scalar	283.574 (75), $p < .001$	60	0.946	0.042	0.052 (0.046, 0.059), $p = 0.69$	158.274 (40), $p < .001$
	Partial scalar (AKT5 1 st thresholds free for European only)	226.814(74), $p < .001$	61	0.961	0.040	0.045 (0.038, 0.052), $p = 0.883$	105.631(39), $p < .001$
	Partial scalar (AKT5 1 st thresholds free for European, Pacific only)	223.646(73), $p < .001$	62	0.961	0.040	0.045 (0.038, 0.052), $p = 0.882$	102.554 (38), $p < .001$

Note. Multi-group confirmatory factor analyses (MG-CFA) were conducted using ordered categorical variable methodology using the Theta parameterisation and weighted least squares means-variance (WLSMV) estimation as implemented in Mplus v8.10; SRMR = standardised root mean squared residual; RMSEA = root mean square error of approximation; CFI = comparative fit index; DIFFTEST = Chi-Square difference test for nested models; M3 one-factor AKT model with residual covariance for AKT7 Yucky/Disgusted and AKT6 Scared, AKT6 Scared face and AKT3 Surprised.

8.4.6 Conclusion

Findings reassured the AKT as a reliable tool for measuring young children's emotional knowledge. CFA results indicate that Model 3 (M3), which accounts for the residual covariances between items AKT7 'Yucky/Disgusted' and AKT6 'Scared', and between AKT6 'Scared' and AKT3 'Surprised', provided a good fit for the overall sample data, each gender and single-ethnicity group. MG-CFA of M3 across children's gender and single-ethnicity groups demonstrated full MI by the children's gender and partial MI by single-ethnicity groups in the 54-month dataset. Therefore, M3 was chosen as the key construct of young children's emotional knowledge in the current research. Confusion among ambiguous negative faces was observed from the covariance between AKT7 'Yucky/Disgusted' and AKT6 'Scared', and AKT6 'Scared' and AKT3 'Surprised'. The results align with previous studies indicating fear and disgust, heavily reliant on shape-oriented characteristics, share many similarities, like negative valence, high arousal levels, and association with withdrawal-related emotions (Woody & Teachman, 2000), making their identification less precise (Martinez & Du, 2012). Additionally, the visual configuration and muscle movements with surprise also complicate recognition of fear, leading to lower accuracy compared to surprise (Chamberland et al., 2017; Roy-Charland et al., 2014). Gagnon et al. (2010) found that children aged 5 and 10 struggled more with discerning fear-surprise than other emotional pairs involving fear.

Based on these results, factor scores generated for the overall sample from the CFA model 3 of the 6-item AKT are used in the LPA and testing MI at the person-centred model level. It is important to note that factor scores were estimated by weighted least squares, which did not explicitly control for measurement errors like latent variables (Kam et al., 2016; Morin et al., 2016; Skrondal & Laake, 2001). Item covariances included in factor scores were given more weight to items presenting lower levels of measurement errors (Skrondal & Laake, 2001). This provides a partial implicit control for measurement errors, making them a stronger alternative than scale scores.

8.5 Appendix 5. The Parent Rating of Oral Language (PROL): Application to Young Children's Expressive Communication

8.5.1 Introduction

Within the context of the Growing Up in New Zealand study (GUiNZs), the Parent Rating of Oral Language (*PROL*) is an assessment tool derived from the *Teacher Rating of Oral Language and Literacy* (*TROLL*: Dickinson & McCabe, 2001; Dickinson et al., 2003), which has been internationally validated as an evaluation of early language and literacy development in classroom settings. Scoring of the *TROLL* assigns equal weight to each item, with higher scores given to responses indicating greater language skills, and scores are aggregated to generate an overall oral language and literacy proficiency score. Unlike its classroom-based counterpart, the *PROL* relies on parental reports of the child's proficiency in communicating in their primary language. The measure comprises parent ratings of five items on a four-point ordinal scale tailored for each item, as presented in Appendix 5 Table 1.

Although the *PROL* has not undergone extensive validation or widespread adoption, its items have been examined regarding their frequency of endorsement within the GUiNZs 54-month dataset (i.e., Morton et al., 2017) and, consistent with the scoring of the *TROLL*, conceptualised as reflecting a single factor representing a child's oral language abilities (Meissel et al., 2019). Consequently, the current research aims to confirm the single-factor model of the *PROL* using Confirmatory Factor Analysis (CFA) followed by an examination of measurement invariance (MI) across child gender and

ethnicity for this tool.

8.5.2 Method

8.5.2.1 Sample and Inclusion Criteria.

The PROL was administered at the 54-month mother interviews. Parents were asked to respond to items based on the child's overall oral language and communication proficiency in the child's main language. Children whose mothers participated in the parent-reported PROL and responded to at least one item from the test were included in the current analysis.

8.5.2.2 Analysis Plan.

Confirmatory Factor Analyses (CFA) of the entire sample data and in each gender and ethnicity group were obtained from 54-month mother interviews in the GUiNZs. Consistent with previous conceptualisation and scoring methods utilised in the *TROLL* and the *PROL*, a one-factor model was assessed for model fit. Analyses were performed using Mean- and Variance-adjusted Weighted Least Squares (WLSMV) estimation in Mplus version 8.10. Model fit was evaluated using the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardised root mean squared residual (SRMR).

Multi-group analyses of MI for the one-factor model of the PROL were conducted as an initial assessment of MI by gender and ethnicity. MI of the one-factor model of the PROL was assessed by the relative fit of nested models with increasingly restrictive constraints on the factor structure and the factor loadings (λ) and item thresholds (τ).

Decrements in model fit were assessed regarding changes (Δ) in fit indices with increasingly restrictive models. Changes in model fit indices indicating measurement non-invariance from studies using Maximum Likelihood (ML) estimators and continuous indicators for unequal samples with group $N > 300$ suggest a decline in CFI greater than $-.010$, accompanied by $\Delta RMSEA \geq .015$ or $\Delta SRMR \geq .030$, indicating non-invariance of measures (Chen, 2007). Sources of model non-invariance were tested with reference to estimates in the configural model and model modification indices. Noticeably, these values are based on models with continuous indicators and associated estimators and are used due to limited criteria for assessing invariance for models employing ordinal indicators and the WLSMV estimator. Due to its sensitivity to sample size (Bollen, 1989; Tucker & Lewis, 1973), changes in chi-square with increasingly restricted models, assessed using the DIFFTEST function, are reported but not considered in the current evaluation of model fit. Tests for full and partial MI of the PROL are presented in Appendix 5 Table 2.

The invariance of the configural model was initially examined, evaluating whether a one-factor structural model held an equally good fit across groups. Subsequently, the threshold model was assessed, wherein all factor loadings and thresholds were constrained to be equal across groups (i.e., full MI). In cases where the tests for scalar invariance were not supported, a backward approach to partial invariance testing was adopted, with parameters sequentially released and model fit reassessed.

8.5.3 Results

8.5.3.1 Item Endorsement.

Appendix 5 Table 1

Items and Response Options for the Parent Rating of Oral Language (PROL) With the Proportion and Number of Endorsements by Item Response Option in the Included Sample (n = 6133)

Label	Item wording	Response options	Endorsement (n)
LD35	Which of the following best describes [NAME]'s pattern of asking questions?	a) never or rarely asks adults questions	76
		b) occasionally asks adults questions	501
		c) sometimes asks adults interesting questions	1536
		d) often asks adults long questions	4015
		missing	< 10
LD36	How often does [NAME] try out new words?	a) never	< 10
		b) rarely	117
		c) sometimes	1853
		d) often	4141
		missing	14
LD34	Which of the following best describes [NAME]'s ability to communicate personal experiences in a clear and logical way? Think of a time if [NAME] was telling you about something that happened when you were not present.	a) is very tentative, only offers a few words and requires you to ask questions	210
		b) offers some information, but needs you to prompt for key parts of the story	1376
		c) offers information and includes the necessary information to really understand the event	2197
		d) offers information and tells experiences in a way that is nearly always complete, logical, and understandable	2332
		missing	18
LD37	How often is [NAME] understandable when speaking to adults other than you or other family members?	a) never	< 10
		b) rarely	83
		c) sometimes	1129
		d) often	4910
		missing	< 10
LD33	Which of the following best describes [NAME]'s ability to communicate when [HE/SHE] is not first understood?	a) never continues to try to be understood	89
		b) often gives up trying to be understood	345
		c) often keeps trying to be understood	2819
		d) will work hard to be understood	2867
		missing	13

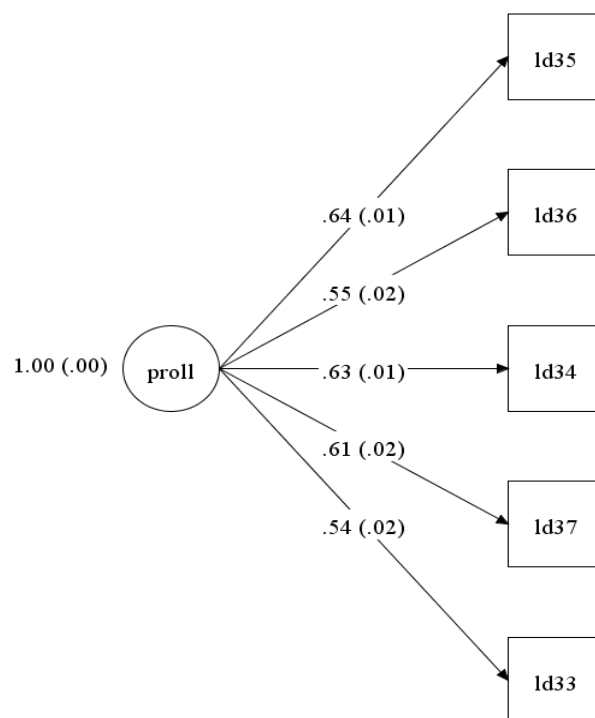
Note. The numbering of item labels refers to the order in which analogous items appear in the Teacher Rating of Oral Language and Literacy. Low-response items where cell size < 10 were recoded for the following analyses, as noted in the footnotes in Tables 2 and 3. Cases with missing values were excluded from the following CFA and MG-CFA analyses.

8.5.3.2 Confirmatory Factor Analyses (CFA).

Before conducting CFA, response options with low-frequency (cell counts <5) were merged with the adjacent response option. All items displayed moderate positive loadings on the latent PROL construct for the entire sample (Appendix 5 Figure 1). Appendix 5 Table 2 showed items and factor loadings of the five items one-factor PROL. Indicators of model fit in Appendix 5 Table 3 suggested that the one-factor model provided a good fit to the data for the overall sample. Appendix 5 Table 3 also validated the structure of the overall model across specific demographic cohorts. CFA was conducted separately for boys and girls, and children in each ethnic group. Results suggested that the one-factor CFA model fit the data for gender and ethnic groups.

Appendix 5 Figure 1

Structure and Standardised Factor Loadings for A One-Factor Model of the PROL in the Growing Up in New Zealand Study 54-Month Dataset



Note. n = 6133. This measurement model predicts children's expressive language and communication using five items included in

the PROL. Standardised scores of factor loadings (standard errors) are in the middle.

8.5.3.3 Multi-Group Confirmatory Factor Analyses (MG-CFA).

Appendix 5 Table 3 provides MI results across the child's gender and ethnicity.

Gender. Results indicated that the configural invariance model provided a good fit to the data when held constant across gender groups, and the threshold invariance model also provided a good fit. Assessment of decrements in model fit indices between these models showed that $\Delta CFI=0.006$, $\Delta SRMR=0.005$ and $\Delta RMSEA=0.002$.

suggesting that the PROL demonstrated full MI by gender.

Ethnicity. Results indicated that the configural model provided a good fit to the data for comparisons by ethnicity. The scalar model provided a good fit for the data; however, assessment of decrements in model fit indices between these models showed that $\Delta CFI=0.045$, $\Delta SRMR=0.012$ and $\Delta RMSEA=0.021$, indicating measurement non-invariance of the PROL by ethnicity. Inspection of configural model estimates and model modification indices suggested that model fit may be improved by freeing the 2nd threshold of LD37 (MI = 33.920) for Europeans, then freeing the 2nd threshold of LD37 (MI = 6.528) for the Pacific group. Based on inspection of model fit indices, threshold specifications for freeing the 2nd thresholds of LD34 for European (MI = 7.226) and Pacific children (MI = 3.861) and the 3rd threshold for item LD33 (MI = 6.788) in the Asian group sequentially increase model fit. The partial invariance model resulted in decrements in model fit indices between these models, showing that $\Delta CFI=0.007$, $\Delta SRMR=0.005$ and $\Delta RMSEA=0.023$ compared to the configural model by ethnic groups. Results suggest the PROL demonstrated partial MI by the child's ethnicity.

Appendix 5 Table 2

Items and Factor Loadings of the Five Items One-Factor Model of the PROL

		Model 1	Model 1 Gender		Model 1 Ethnic Groups				
		Overall	¹ Boys	¹ Girls	¹ European	² Māori	¹ Pacific	³ Asian	Others
		<i>n</i> = 6133	(<i>n</i> = 3149)	(<i>n</i> = 2979)	(<i>n</i> = 3002)	(<i>n</i> = 517)	(<i>n</i> = 729)	(<i>n</i> = 666)	(<i>n</i> = 659)
Label	Items	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)	B(se)
LD35	Which of the following best describes [NAME]'s pattern of asking questions?	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)
LD36	How often does [NAME] try out new words?	0.857(0.034)	0.662(0.055)	0.978(0.089)	0.886(0.054)	0.659(0.086)	0.846(0.098)	0.956(0.136)	1.050(0.108)
LD34	Which of the following best describes [NAME]'s ability to communicate personal experiences in a clear and logical way? Think of a time if [NAME] was telling you about something that happened when you were not present.	0.976(0.033)	0.821(0.064)	1.132(0.098)	1.018(0.051)	0.858(0.087)	1.097(0.114)	0.994(0.141)	1.127(0.099)
LD37	How often is [NAME] understandable when speaking to adults other than you or other family members?	0.946(0.035)	0.834(0.068)	0.997(0.094)	0.995(0.059)	0.688(0.086)	0.962(0.103)	1.039(0.144)	0.970(0.108)
LD33	Which of the following best describes [NAME]'s ability to communicate when [HE/SHE] is not first understood?	0.834(0.032)	0.668(0.053)	0.858(0.075)	0.871(0.048)	0.569(0.079)	0.785(0.096)	0.930(0.132)	0.879(0.094)

Note. B represented the unstandardised beta coefficient, and se represented standard error. ¹ Low-frequency observations for LD36 and LD37 response 'never' (*n* < 5) for group(s) meant that responses of 0 ('never') were re-coded to be equal to 1 (i.e., combined responses with 'Rarely'); ² Low-frequency observations for LD35 response 'never or rarely asks adults questions' (*n* < 5) for group(s) meant that responses of 1 ('never or rarely asks adults questions') were re-coded to be equal to 2 (i.e., combined responses with 'Occasionally'); ³ Low-frequency observations for LD36 response for group(s) meant that responses of 0 ('Never') were re-coded to be equal to 1 (i.e., combined responses with 'Rarely').

Appendix 5 Table 3

Model Fit Indices for the One-Factor Model of the PROL Items Administered in the GUINZs 54-Month Maternal-Report Assessment

Analysis	Model	Chi-Square	Free parameters	CFI	SRMR	RMSEA (90% CI)	DIFFTEST
CFA by gender	Overall (n = 6133)	22.82(5), p < .001	20	0.996	0.012	0.024 (0.015, 0.035), p = 1.000	-
	¹ Boys (n = 3149)	15.66(5), p = .008	18	0.995	0.014	0.026 (0.012, 0.041), p = .997	-
	¹ Girls (n = 2979)	9.07(5), p = .106	18	0.998	0.012	0.017 (0.000, 0.033), p = 1.000	-
¹ MG-CFA by gender	Configural	24.06(10), p = .006	36	0.997	0.013	0.022 (0.011, 0.033), p = 1.000	-
	Scalar	58.01(21), p < .001	25	0.991	0.018	0.024 (0.017, 0.031), p = 1.000	33.01(11), p = .001
CFA by ethnicity	¹ European (n = 3002)	22.51(5), p < .001	18	0.991	0.019	0.034 (0.021, 0.049), p = .960	
	² Māori (n = 517)	9.45(5), p = .077	17	0.988	0.028	0.044 (0.000, 0.084), p = .542	
	¹ Pacific (n = 729)	6.16(5), p = .291	18	0.997	0.017	0.018 (0.000, 0.057), p = .900	
	³ Asian (n = 666)	7.45(5), p = .189	18	0.991	0.022	0.027 (0.000, 0.065), p = .811	
	Other (n = 659)	5.61(5), p = .346	18	0.999	0.020	0.014 (0.000, 0.057), p = .901	
⁴ MG-CFA by ethnicity	Configural	50.79(25), p < .01	85	0.993	0.021	0.030 (0.018, 0.042), p = .997	
	Scalar	252.96(65), p < .001	45	0.948	0.033	0.051 (0.044, 0.058), p = .396	191.98(40), p < .001
	Partial scalar (LD34 2 nd + LD37 2 nd thresholds free for European and Pacific, and LD33 3 rd free for Asian)	111.730 (60), p < .001	50	0.986	0.028	0.028 (0.020, 0.036), p = 1.000	63.01 (35), p < .01

Note. Confirmatory Factor Analyses (CFA) and Multi-group confirmatory factor analyses (MG-CFA) were conducted using ordered categorical variable methodology using the Theta parameterization and weighted least squares means-variance (WLSMV) estimation as implemented in Mplus v8.10; SRMR = standardised root mean squared residual; RMSEA = root mean square error of approximation; CFI = comparative fit index; DIFFTEST = Chi-Square difference test for nested models; ¹ Low-frequency observations for LD36 and LD37 response 'never' (n < 5) for group(s) meant that responses of 0 ('never') were re-coded to be equal to 1 (i.e., combined responses with 'Rarely'); ² Low-frequency observations for LD35 response 'never or rarely asks adults questions' (n < 5) for group(s) meant that responses of 1 ('never or rarely asks adults questions') were re-coded to be equal to 2 (i.e., combined responses with 'Occasionally'); ³ Low-frequency observations for LD36 response for group(s) meant that responses of 0 ('Never') were re-coded to be equal to 1 (i.e., combined responses with 'Rarely'); ⁴ Low-frequency observations for LD35 'never or rarely asks adults questions' were re-coded to be equal to 2 (i.e., combined responses with 'Occasionally', LD36 and LD37 'Never' (n < 5) responses of 0 ('never') were re-coded to be equal to 1 (i.e., combined responses with 'Rarely') across groups ; gender missing for n = 5 case

8.5.4 Conclusion

Results in the current research confirmed that the one-factor model of the PROL provided a good fit to the data overall, as well as within each gender and ethnicity group. These findings provide clearer insight into the validity and reliability of the PROL as administered in the GUiNZs, reassuring its use as a reliable tool for measuring young children's expressive language. Invariance testing results further showed that the PROL functions consistently across children's gender groups, and are partially equivalent across their ethnic groups, demonstrating suitability for a diverse young children's sample and supporting its generalisability in capturing the intended construct across different demographic subgroups. When assessing the invariance of its model across groups, results of MCGFA indicate that the PROL demonstrated full MI by the child's gender and partial MI by ethnicity in the 54-month GUiNZs dataset. These full and partial invariance models may be used to generate unbiased factor estimates for oral language as assessed by the PROL. Using these factors, researchers can proceed with substantive analyses combining PROL scores within the population or comparing group means.

8.6 Appendix 6. The adapted Peabody Picture Vocabulary Test-Third Edition (PPVT-III): Application to Young Children's Receptive Vocabulary and Communication in the New Zealand Context

8.6.1 Introduction

The Peabody Picture Vocabulary Test (PPVT) was developed to evaluate receptive vocabulary and verbal ability in standard American English among individuals aged 2 to 90+ years (Dunn, 2021; Dunn, 1959; Dunn & Dunn, 2007; Dunn & Dunn, 1981; Dunn & Dunn, 1997). Since its inception in 1959, the PPVT has undergone several revisions, with the latest edition available being the PPVT-5th (Dunn, 2021; Dunn, 1959). Each version differs in terms of the stimulus words, illustrations and the samples used to obtain normative scores. This assessment offers several advantageous features: 1) ease of administration; 2) no requirement for reading or writing; 3) appropriate for a wide range of ages, and 4) its broad applicability. The PPVT-III test administration involves presenting examinees with a series of black-and-white line drawings and prompting them to select the drawing that best corresponds to a stimulus word spoken by the examiner. Examinees are not required to read or write during the test, making it accessible to individuals with various literacy levels. As such, the test is suitable for administration across a broad spectrum of ages and diverse educational backgrounds, enhancing its versatility in various testing environments. However, the PPVT-III is not suitable for use with individuals who have significant visual or auditory impairments. In light of these qualities, the PPVT-III was adapted for administration as part of the

Longitudinal Study of Australian Children (LSAC) and GUiNZs (Ly et al., 2020; Rothman, 2005). This current report views the content as evidence for the validity of the PPVT-III and describes its adaptation for the research purposes of the LSAC and GUiNZs.

8.6.2 Format and scoring of the PPVT-III

The PPVT-III assessment is a stand-alone test designed to measure receptive vocabulary, comprising four practice items and a set of 204 test items. These 204 test items are organised into 17 sets, each containing 12 items of increasing difficulty. Four simple black-and-white illustrations (plates) are presented in each test item. During the assessment, the examiner says a word, and the examinee's task is to select the picture that they believe best represents the meaning of the word (Dunn & Dunn, 1997). For each plate, the interviewer verbally presented a word and asked the child to point to the corresponding image or to indicate "don't know". The total raw PPVT score is computed by tallying the number of correct responses on items with the highest discriminatory power for individuals of similar ability. These raw scores can then be transformed into standard scores with a mean of 100 and a standard deviation of 15. Additionally, age/mental age equivalents, stanines, and percentiles can be derived from the raw scores to comprehensively evaluate the examinee's performance (McKinlay, 2011).

8.6.3 Validity of the PPVT-III as an Assessment of Vocabulary Among Children

Previous research has explored the PPVT-III's validity, consistency and reliabilities in evaluations of children's English vocabulary skills, establishing its utility as both a tool

for evaluating English vocabulary acquisition, and a screening instrument for identifying potential underachievement (Campbell et al., 2001; Dollaghan & Campbell, 2009; Gerde & Powell, 2009; McKinlay, 2011; Phillips et al., 2004). Previous research by Culbert et al. (1989) indicated a one-factor structure of the PPVT first version in a psychiatric sample aged from 6 to 16 through a principal axis factor analysis using varimax rotation. The subsequent PPVT-III has also been identified as an evidence-based assessment instrument for use with paediatric populations and typically developing children (Campbell et al., 2008). Research from the US shows the PPVT-III is a culturally valid test for use with African American children, as well as children in ethnic minorities (Dunn & Dunn, 1997). For example, the study by Campbell et al. (2001) examining the concurrent validity of the PPVT-III found that a receiver operating characteristic analysis supported the PPVT-III as a valid intellectual and achievement screener for children. In the research, the area under the curve (AUC) was used as the index of the ROC to interpret the probability of a screening test score correctly classifying a pair of participants. 0.7 to 0.9 denotes moderate test accuracy and 0.9 to 1.0 high test accuracy (Swets, 1988, 1992). Results of the ROC analysis indicated that using standard scores, the PPVT-III yielded an AUC of 0.909, with a standard error of 0.051 used as an intellectual screener, and an AUC of 0.823 with a standard error of 0.031 as an achievement screener. Longitudinal investigations in New Zealand have found that antenatal and perinatal risk factors in mothers, such as antenatal anxiety, pre-pregnancy smoking, or lack of folic acid intake during pregnancy, are associated with

below-average receptive English language skills in preschool children at 54 months, as assessed by PPVT-III (Galvin et al., 2020; Neumann et al., 2019).

8.6.4 Cultural Biases on PPVT-III

While research findings support the validity and reliability of the PPVT-III for the populations for which it was originally developed, caution is needed when applying this assessment to individuals from different cultural and linguistic backgrounds (Haitana et al., 2010; Stockman, 2000). Stockman (2000) raised concerns regarding the equitable nature of standardised vocabulary measures and the need to re-evaluate potential test biases, especially considering that children from culturally, linguistically, and socioeconomically diverse backgrounds may have varying understandings of English word meanings compared to monolingual children due to their distinct cultural and linguistic experiences. Indeed, a subsequent study by Champion et al. (2003) indicated that the PPVT-III tended to assign lower scores disproportionately to African American children from low-income backgrounds. Children's receptive language performance on the measure does not reflect receptive language development per se but is confounded by environmental exposures to the English language. For instance, research suggests that the support for children's access to vocabulary learning resources available at home and within the community can have positive effects on children's proficiency in their ethnic language vocabulary (Dixon et al., 2012). However, due to prior research supporting the fairness of the PPVT-III for children from diverse cultural, linguistic and socioeconomic backgrounds, the PPVT-III version appears more

suitable for individuals from diverse racial and income groups compared to earlier versions (Wood & Schatschneider, 2019).

Even though research indicated minimal biases in the PPVT test (Argulewicz & Abel, 1984; Halpin et al., 1990), there remains relatively less attention investigating potentially culturally biased items of the PPVT-III in the New Zealand context (Haitana et al., 2010). For example, a controversy arose over whether lower maternal education shows greater predictive validity of the NZ Communicative Development Inventories: Words and Sentences in relation to the PPVT-III, compared to those with higher maternal education in New Zealand contexts (Reese & Read, 2000). Possible reasons for the inconsistency of such findings may relate to differences in maternal reporting strategies. Additionally, variances in the methodologies used or true differences in vocabulary could contribute to the observed ethnic group disparities in PPVT-III scores. Research by Haitana et al. (2010) within the New Zealand context, suggests that the PPVT-III can serve as an appropriate assessment of English vocabulary for Māori children aged 5-11, but it may not be suitable for measuring overall language ability. The research assured the PPVT-III to be a relatively unbiased test for Māori children, and culture can impact the ways people respond to standardised assessments.

Concepts and ideas valued by one culture may not be interpreted equally by another. Ensuring the fairness of a test requires the incorporation of culturally appropriate knowledge and testing methods that are suitable for individuals from diverse cultural backgrounds (Palmer, 2004). Further, ECE education and interventions

that address the social, emotional, language and cognitive needs of young children are requested to be culturally sensitive. Therefore, the following sections present an examination of the validity and predictability of the adapted PPVT-III used as an indicator of young children's receptive language in the current research.

8.6.5 The Adapted PPVT-III in the GUiNZs

To assess receptive vocabulary and verbal ability among 54-month preschoolers, the Growing Up in New Zealand study (GUINZs) employed an adapted version of the original PPVT-III comprising a selection of 40 plates/items. This approach aligns with the methodology utilised in the Longitudinal Study of Australian Children (LSAC), noted by Rothman (2005), in which the adapted PPVT-III overall score was created through the Rasch Modelling technique within the LSAC dataset (Australian Bureau of Statistics, 2020). These stimuli were selected from the full PPVT-III battery based on the position of items in the original PPVT-III, and the administration purpose of the test, especially regarding the number of correct items required for the basal/ceiling sets (Rothman, 2005). The adapted version of the PPVT-III provided similar results to the full version, with a robust correlation between the adapted and full test scores of 0.93 for all children and 0.91 for children aged 4.

The adapted version of the PPVT-III has predominantly been used in Australia and New Zealand (Christensen et al., 2017; Feeney et al., 2015; Walker & Berthelsen, 2017). Research by Hammersley et al. (2022) using LSAC data found that a healthier dietary intake measured by the frequency of children's consumption of seven types of food at

age 2-3 was longitudinally associated with better verbal cognitive skills at ages 4-5, as measured by the adapted PPVT-III. Longitudinal research from Chong et al. (2019) found that temperament at 2-3 years of age is associated with verbal cognitive skills at 6-7 years of age. Significant differences shown in the adapted PPVT-III were observed between children with multilingual backgrounds living in Australia compared to children with monolingual backgrounds (McLeod et al., 2016). In the New Zealand context, lower PPVT-III receptive language scores among children participating in the 54-month assessment wave from the GUiNZs are associated with delayed executive function. Further analysis of the GUiNZs data indicated that maternal smoking before pregnancy, folate intake during the first trimester and antenatal anxiety were significantly highly associated with children's receptive language measured by the adapted PPVT-III at 54 months (Neumann et al., 2021a). Overall, the adapted PPVT-III has supported the research regarding children's vocabulary or receptive language development in Australia and the New Zealand context.

8.6.5.1 Administration of the Adapted PPVT-III.

The administration of the adapted PPVT-III begins by presenting the child with two training items/plates, each displaying four distinct images, and asking them to point to the image that matches the given word or respond with 'don't know'. The actual test phase begins if the child responds correctly to both items/plates. The test phase begins with the presentation of **Set 1 (Core Set)**, comprising a sequence of 20 plates/items, each displaying four distinct images. If the child made 15-20 errors in the Core set, they

were presented with **Set 2 (Basal set)**, comprising 10 additional items/plates. If a child made 7-14 errors on the Core set, they were assumed to have answered all 10 items in the Basal set correctly, and the test was concluded. If the child made 0-6 errors in the Core set, they were assumed to have answered all 10 items in the Basal set correctly and were presented with **Set 3 (Ceiling set)**, comprising 10 new items. If the child declined to continue at any point during the test, the test was concluded, and no total score was calculated.

8.6.5.2 Scoring the Adapted PPVT-III.

In the GUINZs PPVT-III technical report Ly et al. (2020), three potential approaches to scoring responses to the adapted PPVT-III were evaluated, based on similar work from the LSCA (Rothman, 2005). The first was equal weighting raw score summation, in which scores were calculated by summing the number of correct answers given/assumed, with each item carrying an equal weighting of '1'. This resulted in a tri-modal distribution of scores for the sample, reflecting the range of total scores achievable with completion of Set 1 and Set 2; Set 1 and with all 10 Set 2 items assumed correct; Set 1 and with all 10 Set 2 items assumed correct + Set 3.

The second method reflected a differential weighting score summation. In this method, an index of the relative difficulty of each item relative to the average inaccuracy displayed within each set (i.e., % item inaccuracy ÷ average % item inaccuracy in the set) was calculated based on observed responses in the sample overall. As such, items that were identified with higher inaccuracy relative to the average inaccuracy

within the set were weighted greater than 1, and items identified with lower inaccuracy relative to the average inaccuracy within the set were weighted less than 1. As seen in the raw score summation, a score calculated as the sum of correct/assumed difficulty-weighted items for participants resulted in a tri-model distribution of scores for the sample. To normalise this distribution, GUiNZs researchers built a model to estimate the missing Set 2 and/or Set 3 scores for those who did not complete these Sets. Linear regression models predicting Set 2 or Set 3 weighted scores from Set 1 weighted scores were estimated among those who completed Set 1 + Set 2 or Set 1 + Set 3. For those who had not completed Set 2 and/or Set 3, coefficients from the linear models were used to estimate a Set 2 and/or Set 3 difficulty-weighted score (range 0-10) based on transformed summed Set 1 difficulty-weighted scores. The sample's adjusted difficulty-weighted scores displayed a normal distribution.

The third scoring method was a latent ability assessment conducted using Item Response Theory (IRT) models. In these models, latent ability levels were determined by anchoring within the Core Set (Set 1) items to allow for proportional scaling of both the Basal and Ceiling Sets, rather than applying equal increments solely to the Basal Set. Ultimately, the GUiNZs determined that a 3-parameter model (modelling item difficulty, discrimination, and guessing) was preferred.

8.6.6 Current Research

Current research will employ the variable derived from the 3-parameter IRT model analysis conducted by the GUiNZs to assess receptive vocabulary performance in

children from the GUiNZs 54-month cohort. The derived scores range from -4 to 3 with a mean score of 0 and display a positively skewed distribution.

Prior research has categorised scores that were one or more standard deviations below the mean as indicating 'below average' receptive language performance, with all others considered to indicate typical receptive language performance (Neumann et al., 2019). The current work will treat the variable as a continuous indicator of receptive language performance. While the adapted PPVT-III score derived from the 3-parameter IRT model offers advantages over 1- and 2-parameter models by accounting for item difficulty, discrimination and guessing, which allows more precise calculation of the latent ability, it is important to note its limitations as an indicator of development in the current research. First, it is not possible to assess the measurement invariance of the adapted PPVT-III at the construct level by ethnicity or gender. Further, it is important to consider the measure's validity for children with diverse cultural or linguistic backgrounds (Dixon, 2011; Dixon et al., 2012). In Appendix 6 Table 1, most children with the adapted PPVT-III scores in the GUiNZs research were identified as English speakers at the 54-month assessment (99.7% of children spoke English). 490 children with the PPVT-III scores did not speak English as their primary language at home. The adapted PPVT-III was completed by 5606 (90.82%) of the 6156 children participating in the Child Observation component of the GUiNZs 54-month assessment wave. Appendix 6 Table 1 describes this group of children by gender, Mother-

prioritised children’s single-combined ethnicity and language ability assessed by the adapted PPVT-III.

Appendix 6 Table 1

Frequency of the PPVT-III Across Children’s Gender and Mother-Prioritised Single-Combined Ethnicity

Grouping variables	Categories	Number of participants engaged in the PPVT-III test (n)
Children’s gender	Boy	2854
	Girl	2749
	missing	< 10
Mother-prioritised children’s single-combined ethnicity	New Zealand Europeans only	2842
	Māori only	456
	Pacific people only	658
	Asian only	575
	Other ethnicities only	551
	Dual	436
	missing	88
Language speaking	English	5591
	Not speaking	11
	missing	< 10
Language mostly speaking at home with PPVT-III scores	English	1340
	Language other than English	490
	Missing	3770
	Don’t know’	< 10
Total		5606

Note. Other ethnicities include MELAA, Other and New Zealander. Low-responses item ‘Don’t know’ where cell size < 10 were treated as missing values. Cases with missing values were excluded in the following analyses.

The current research uses children’s gender and ethnic groups as potential indicators for understanding the diverse young children’s sample. Of those children with adapted PPVT-III scores, gender groups made up a balanced sample, with each gender group accounting for approximately 50% of the overall sample; for children identified with a single-ethnicity group, 436 were identified with two or more ethnic

groups. For those who were best identified with a single-ethnicity group, around half were NZ European, 8.13% Māori, 11.74% Pacific Peoples, 10.26% Asian, and 9.83% other ethnicities including MELAA (Middle Eastern/Latin American/African) and New Zealanders, a detailed description can be found in **Appendix 2** Ethnicity Prioritisation in the current research. Previous research using the GUiNZs dataset highlights distinct patterns in the distribution of latent receptive language skills among 54-month-old children, adding insights into the diverse sample. The study by Ly et al. (2020) highlights the following key findings:

(1) There are gender differences in adapted PPVT-III scores. 54-month-old girls scored higher ($m = 0.04$, $sd = 0.89$) than boys ($m = -0.04$, $sd = 0.92$). The distributions of latent receptive language abilities were approximately normally distributed for both gender groups, with mean scores closely aligned with the median scores.

(2) There are differences in adapted PPVT-III scores between ethnic groups. Ly et al. (2020) applied the administered prioritisation of mother-prioritised child's ethnicity which assigned individuals to one ethnic group according to the prioritised order as listed in the Department of Statistics (1993): Māori > Pacific Peoples > Asian > Other ethnic groups except for European > European, and examined the latent receptive language abilities for each of the ethnic groups. Results indicated that European children exhibited higher scores ($m = 0.32$, $sd = 0.78$), compared to children from other ethnic groups, including Māori ($m = -0.27$, $sd = 0.87$), Pacific ($m = -0.71$, $sd = 0.881$), Asian ($m = -0.49$, $sd = 0.79$), and 'Other' ethnicities that include MELAA and other

ethnicities not reported within the European, Māori, Pacific, or Asian groupings ($m = -0.453$, $sd = 0.938$). Examination of the distribution of latent receptive language abilities scores revealed some distinctive patterns. There was no clear distribution pattern for the 'Other' ethnicities group. The distribution for NZ European was skewed to the right, with the majority of children scoring above the sample mean within the group indicating a high level of performance on the adapted PPVT-III task within this group, whereas the distributions for Māori, Pacific, and Asian children are skewed to the left, indicating fewer children score higher than the sample means within each ethnic group and performance disparities within these ethnic groups.

In light of existing analyses of the adapted PPVT-III score and theoretical models suggesting cultural- and language-associated biases associated with childhood assessments of receptive language ability, these observed trends can not necessarily be viewed as either group differences or indicative of measurement non-invariance of the adapted PPVT-III as these cannot be assessed at the construct level from the single indicator (the latent variable scores) available. The current research intended to use these scores as an indicator of receptive language ability within a finite mixture model. However, we cannot assess measurement invariance at the construct level, and the formation and comparison of finite mixture models across gender and ethnic groups may be affected by the potential non-invariance of adapted PPVT-III scores. To address these issues, potential measurement non-invariance at the person-centred model level, accounting for measurement non-invariance in the model indicators, will be assessed

across children's gender and ethnic groups using the covariate approach (Bettencourt et al., 2022; Masyn, 2017).

8.7 Appendix 7. The Adapted Pencil Tap Task-Hand Clap Task: Application to Young Children's Executive Function in the New Zealand Context

8.7.1 Introduction

The Hand Clap task, as adapted from the Pencil Tap task in the GUiNZs, primarily measures a child's performance in inhibitory control alongside other components of executive function such as working memory (Buckley et al., 2019). The original Pencil Tap task, derived from the Luria-Nebraska Neuropsychological battery (Bialystok et al., 2010; Golden et al., 1980; Plaisted et al., 1983), has been utilised in longitudinal studies such as the Preschool Self-Regulation Assessment (Smith-Donald et al., 2007) and has shown sensitivity to inhibitory control (Rhoades et al., 2009). The task requires the capacity to simultaneously consider two things while demonstrating inhibitory control over one's behaviour (Diamond & Taylor, 1996). Specifically, children are instructed to tap once when an assessor taps twice and twice when an assessor taps once. Performance on the task is assessed using the percentage of correct responses, serving as an indicator of inhibitory control. The task is well-suited for assessing young children's inhibitory control as it aligns with the developmental progression observed during the preschool years when significant improvements in their ability to regulate their behaviour (Diamond & Taylor, 1996). In the GUiNZs modification, children perform hand clapping instead of pencil tapping for the response. The task retained the original task's reverse imitation, requiring inhibitory control as children perform actions that are the opposite to those of an assessor's across 16 trials. Task outcomes provide

the count of correct responses, scored using a standardised scoring technique resulting in a total scale score ranging from 0 to 16, such that higher scores indicate greater inhibitory control. A previous study from Buckley et al. (2019) in the GUiNZs using data from this task from the 54-month child observation assessment wave indicates that the hand clap task had good internal reliability, with trial performance demonstrating a Cronbach's alpha of 0.87.

The Hand Clap task has been used across various GUiNZ studies as an indicator of 54-month-old children's inhibitory control, EF, and, in concert with other assessments, as an indicator of general cognitive outcomes among young children (Buckley et al., 2019; Corkin et al., 2021; Neumann et al., 2021a; Slykerman et al., 2023). Early assessment of executive function is key to children's overall development, which informs strategies to improve EF (Beauchamp & Anderson, 2010; Londono et al., 2024). Research in the GUiNZs found that concurrent and persistent behavioural difficulties, as noted as abnormal at 54 months, at 2 years and 54 months, are linked with delayed executive control during early childhood (D'Souza et al., 2020). High inhibitory control in young children as assessed by the Hand Clap Task and categorised into low, intermediate, and high groups, is associated with maternal relationship status, maternal education, first-trimester folate supplementation, maternal BMI and smoking before pregnancy (Buckley et al., 2019). In short, the Hand Clap task can be used as an effective tool for assessing EF in young children.

8.7.2 Current Research

The current thesis uses total scores from the Luria Hand Clap task (count scores) as an indicator of children's executive function. The distribution of scores for the Hand Clap Task (Appendix 7 Figure 1) exhibited a negative and a small ceiling effect, representing a high proportion of children making few errors in task performance. As noted in Appendix 7 Table 1, the Hand Clap Tasks score displayed a mean score of 11.14 with a standard deviation of 5.03, with a median score of 13. The previous score categorisation and their respective proportions from the GUiNZs 54-month assessment employed an arbitrary categorisation to divide participants into three performance groups: low, intermediate, and high groups (Buckley et al., 2019). This categorisation was designed to approximate the 25th and 50th percentiles (8 and 13, respectively). Within this categorisation, 1473 children scored in the low-performance category, 1387 fell into the intermediate category, and 2588 were considered high-performance in the task.

Score distributions of the Luria Hand Clap Task by gender and ethnicity are presented in Appendix 7, Figures 2 and 3. Regarding gender, mean and median scores were slightly higher in girls (mean of 11.52 and median of 14) than in boys (mean of 10.77 and median of 13). In terms of ethnicity, European children had the highest mean (11.60) compared to children of other ethnicities. Pacific children had the lowest mean (9.46) and median (11) scores. The distributions of the raw scores also show that scores from children in the Māori and Pacific groups displayed a lesser ceiling effect, with

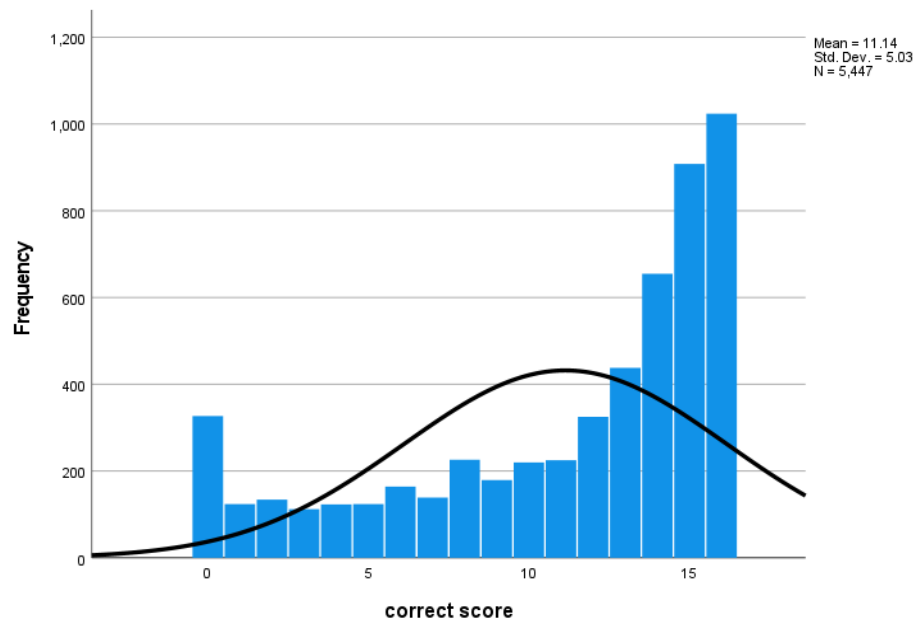
higher proportions of scores in the low-to-intermediate range. In contrast, other groups share similar skewness and kurtosis in distribution. It is important to note that a negative binomial model was chosen because it accounts for overdispersion. The model accommodates count data with non-negative integers and includes a dispersion parameter to handle variability in the distribution. However, binning data into large ordinal categories may reduce statistical power and precision, potentially losing important information. Properly assessing the distribution's shape and dispersion is critical to model fitting and interpretation.

There are three key considerations for using the Luria Hand Clap task in the present research. First, the nature of the count data did not allow for testing measurement invariance at the construct level using MG-CFA in the current research. Measurement invariance test using MG-CFA often requires the data to be continuous or normally distributed, whereas the count data violate these distribution assumptions. Second, count data may exhibit overdispersion, in which the variance exceeds the mean, requiring the use of appropriate models to account for the overdispersion or skewness (Cameron & Trivedi, 2013). Third, the current research used median scores of the Luria Hand Clap task to compare observed group differences in executive function, as the median provides a robust measure of central tendency of discrete data, providing a more robust summary of the measure (Wilcox, 2017). Despite these methodological considerations, the Luria Hand Clap task in the GUiNZs 54-month dataset remains a

valuable indicator of executive function, offering insights into children's learning and development.

Appendix 7 Figure 1

Distribution of the Luria Hand Clap Task Count Scores in the GUINZs 54-month Dataset



Note. Missing data is not included here.

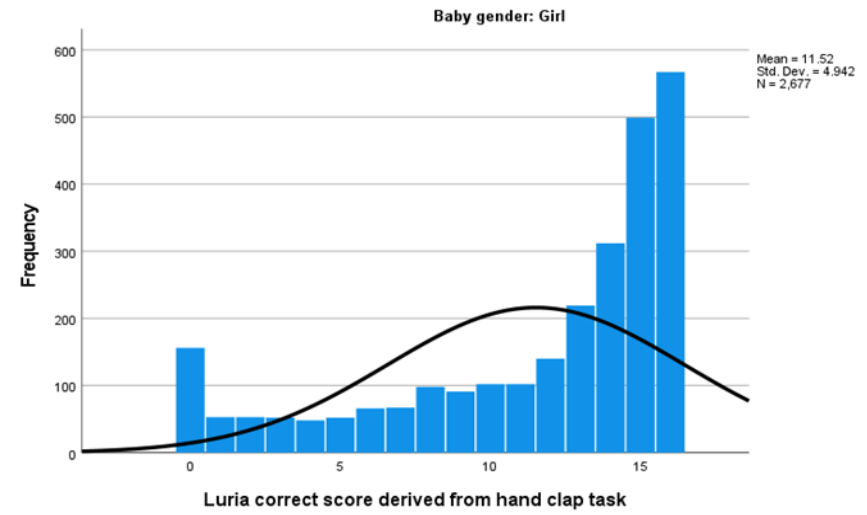
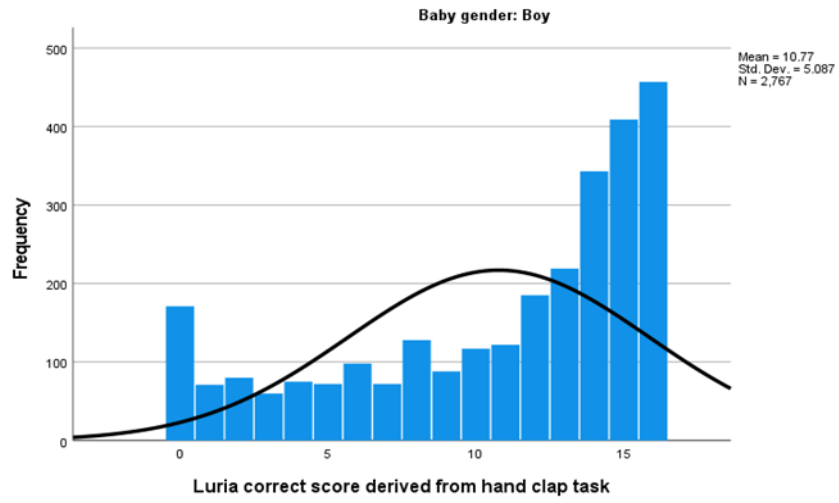
Appendix 7 Table 1

Summary Statistics for the Raw Scores of the Adapted Luria Hand Clap Task

	N	Missing	Mean	1st	Median	3rd	SD	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis
Total	5447	709	11.14	8	13	13	5.03	-0.966	0.033	-0.341	0.066
Boy	2767	398	10.77	7	13	13	5.087	-0.843	0.047	-0.585	0.093
Girl	2677	309	11.52	9	14	14	4.942	-1.109	0.047	-0.008	0.095
European only	2781	227	11.6	9	14	14	4.713	-1.129	0.046	0.139	0.093
Māori only	443	75	10.25	6	12	12	5.394	-0.695	0.116	-0.927	0.231
Pacific people only	628	101	9.46	4.25	11	11	5.492	-0.421	0.098	-1.229	0.195
Asian only	554	116	11.46	8	14	14	5.185	-1.103	0.104	-0.154	0.207
Other only - Other, MELAA, NZer	537	124	11.19	8	13	13	4.875	-1.02	0.105	-0.115	0.21
Dual	418	48	11.23	8	13	13	5.128	-0.988	0.119	-0.342	0.238

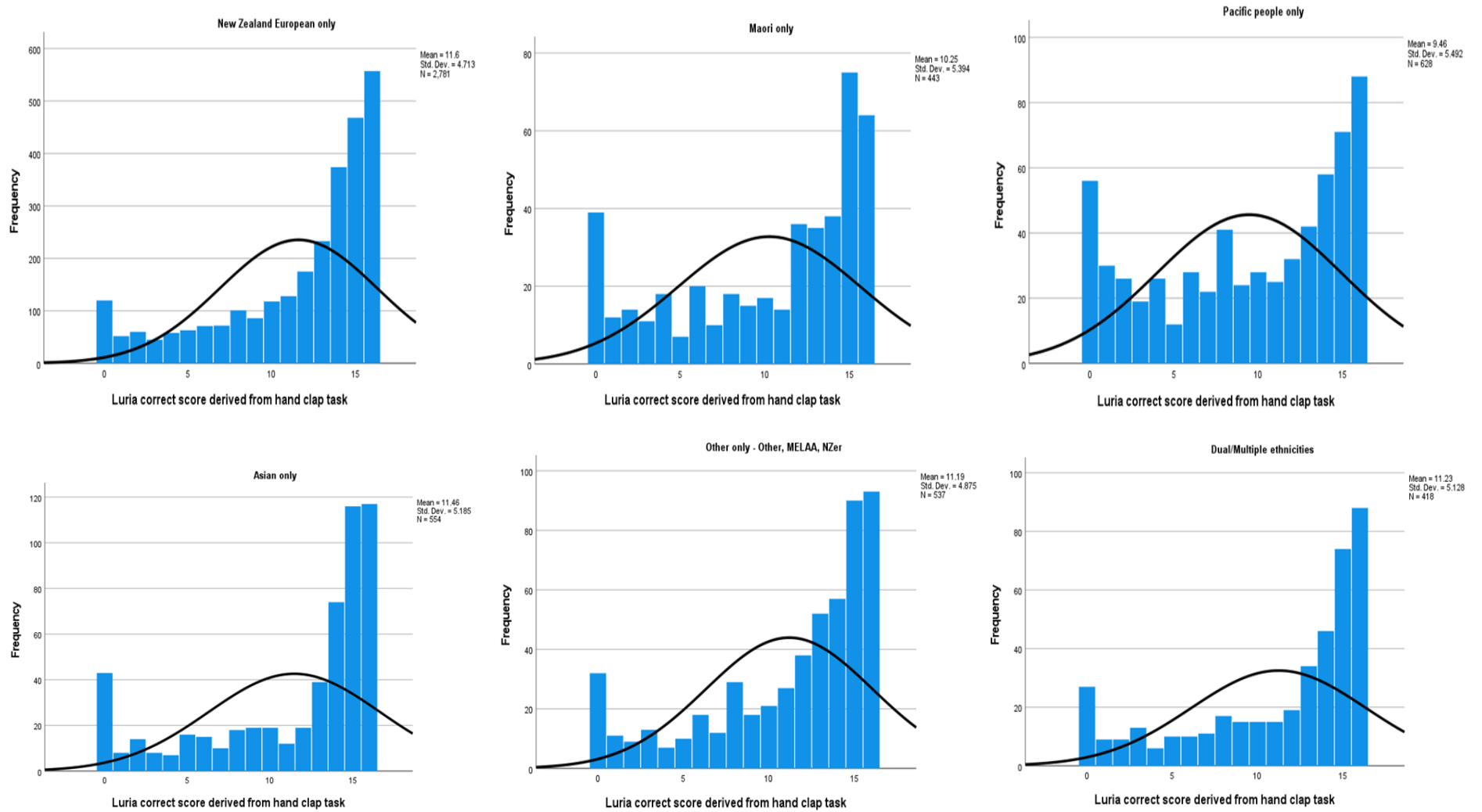
Appendix 7 Figure 2

Distribution of the Luria Hand Clap Task Count Scores by the Child's Gender



Appendix 7 Figure 3

Distribution of the Luria Hand Clap Task Count Scores by Mother-Prioritised the Child's Ethnicity



8.8 Appendix 8. Profile Membership Transitions in Chapter 5

8.8.1 Profile Membership Transitions Between Profiles in Variances Fully Constrained

Models

Appendix 8 Table 1

Changes in Numbers of Participants in 3-Group Versus 4-Group Variances Fully-Constrained Models (n)

		4-group model			
		Profile 1	Profile 2	Profile 3	Profile 4
3-group model	Profile 1	< 10	< 10	360	2249
	Profile 2	440	76	< 10	< 10
	Profile 3	171	641	1443	321

Note. The number of participants in the middle sections suggested changes in participant counts between the 3-group and 4-group variances fully constrained models. Low counts ($n < 10$) do not indicate low item responses but rather reflect small changes in participant numbers between profiles in 3-group and 4-group variances fully constrained models. Darker 25% grey represented the most significant changes/stabilities in the numbers of participants between profiles in the 3-group and 4-group variances fully constrained models; darker 15% grey represented the second largest changes/stabilities; and darker 5% grey represented the third largest changes/stabilities.

Appendix 8 Table 2

Changes in Numbers of Participants in 4-Group Versus 5-Group Variances Fully-Constrained Models (n)

		5-group model				
		Profile 1	Profile 2	Profile 3	Profile 4	Profile 5
4-group model	Profile 1	41	< 10	477	93	< 10
	Profile 2	628	78	< 10	< 10	14
	Profile 3	28	21	< 10	1071	684
	Profile 4	16	2077	< 10	< 10	478

Note. The number of participants in the middle sections suggested changes in participant counts between the 4-group and 5-group variances fully constrained models. Low participant counts ($n < 10$) do not indicate low item responses but rather reflect small changes in participant numbers between profiles in 4-group and 5-group variances fully constrained models. Darker 25% grey represented the most significant changes/stabilities in the numbers of participants between profiles in the 4-group and 5-group variances fully constrained models; darker 15% grey represented the second largest changes/stabilities; and darker 5% grey represented the third largest changes/stabilities.

8.8.2 Profile Membership Transitions Between Profiles in Variances Freed Models

Appendix 8 Table 3

Changes in Numbers of Participants in 3-Group Model Versus 4-Group Model With Variances Freed (n)

		4-group model			
		Profile 1	Profile 2	Profile 3	Profile 4
3-group model	Profile 1	912	67	38	< 10
	Profile 2	< 10	< 10	1680	307
	Profile 3	< 10	2300	139	262

Note. The number of participants in the middle sections suggested changes in participant counts between the 3-group and 4-group variances freed models. Low participant counts ($n < 10$) do not indicate low item responses but rather reflect small changes in participant numbers between profiles in 3-group and 4-group variances freed models. Darker 25% grey represents the most significant changes/stabilities in the numbers of participants between profiles in the 3-group and 4-group variances freed models; darker 15% grey represents the second largest changes/stabilities; darker 5% grey represents the third largest changes/stabilities.

Appendix 8 Table 4

Changes in Numbers of Participants in 4-Group Model Versus 5-Group Model With Variances Freed (n)

		5-group model				
		Profile 1	Profile 2	Profile 3	Profile 4	Profile 5
4-group model	Profile 1	87	532	295	< 10	< 10
	Profile 2	868	< 10	334	11	1154
	Profile 3	74	< 10	193	1562	28
	Profile 4	< 10	< 10	< 10	281	280

Note. The number of participants in the middle sections suggested changes in participant counts between the 4-group and 5-group variances freed models. Low participant counts ($n < 10$) do not indicate low item responses but rather reflect small changes in participant numbers between profiles in 4-group and 5-group variances freed models. Darker 25% grey represents the most significant changes/stabilities in the numbers of participants between profiles in the 4-group and 5-group variances freed models; darker 15% grey represents the second largest changes/stabilities; darker 5% grey represents the third largest changes/stabilities.

References

- Abenavoli, R. M., Greenberg, M. T., & Bierman, K. L. (2017). Identification and validation of school readiness profiles among high-risk kindergartners. *Early Childhood Research Quarterly, 38*, 33-43. <https://doi.org/10.1016/j.ecresq.2016.09.001>
- Acker, J. (1990). Hierarchies, jobs, bodies: A theory of gendered organisations. *Gender & Society, 4*(2), 139-158. <https://doi.org/10.1177/089124390004002002>
- Ackerman, D. J., & Friedman-Krauss, A. H. (2017). Preschoolers' executive function: Importance, contributors, research needs and assessment options. *ETS Research Report Series, 2017*(1), 1-24. <https://doi.org/10.1002/ets2.12148>
- Adams, K., Bull, R., & Maynes, M.-L. (2016). Early childhood spirituality in education: Towards an understanding of the distinctive features of young children's spirituality. *European Early Childhood Education Research Journal, 24*(5), 760-774. <https://doi.org/10.1080/1350293X.2014.996425>
- Adani, S., & Capanec, M. (2019). Sex differences in early communication development: Behavioral and neurobiological indicators of more vulnerable communication system development in boys. *Croatian medical journal, 60*(2), 141-149. <https://doi.org/10.3325/cmj.2019.60.141>
- Ahmad, S., Elizabeth, R. P., Karen, E. W., & Susan, M. B. M. (2019). Development of an index of socio-emotional competence for preschool children in the Growing Up in New Zealand study. *Frontiers in Education, 4*. <https://doi.org/10.3389/feduc.2019.00002>

Ahmad, S., Waldie, K. E., Morton, S. M. B., & Peterson, E. R. (2021). Do patterns of levels of socio-emotional competence during early childhood predict executive function at 4.5 years? *Child Psychiatry & Human Development*, 1.

<https://doi.org/10.1007/s10578-021-01128-3>

Ainsworth, M. D. S., Blehar, M.C., Waters, E., & Wall, S. (1979). *Patterns of attachment: A psychological study of the strange situation* (1st ed.). Psychology Press.

<https://doi.org/10.4324/9781315802428>

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, 18(3), 91-93. <https://doi.org/10.1016/j.tjem.2018.08.001>

Al-Shawaf, L., & Lewis, D. M. G. (2020). Evolutionary psychology and the emotions. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1452-1461). Springer International Publishing.

https://doi.org/10.1007/978-3-319-24612-3_516

Albritton, K., Stein, R., & Cruz, K. (2023). Embracing the promise and potential of preschool-age black boys: Strength-based opportunities for early childhood school psychologists. *School Psychology Review*, 52(3), 343-356.

<https://doi.org/10.1080/2372966X.2021.1977586>

Allen, B. (2023). *The science and clinical practice of attachment theory: A guide from infancy to adulthood*. American Psychological Association.

<https://doi.org/10.1037/0000333-000>

Anderson, P. (2002). Assessment and development of executive function (EF) during

childhood. *Child Neuropsychology*, 8(2), 71-82.

<https://doi.org/10.1076/chin.8.2.71.8724>

Anderson, S., & Messick, S. (1974). Social competency in young children.

Developmental psychology, 10(2), 282-293. <https://doi.org/10.1037/h0035988>

Anderson, V. (2001). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Pediatric Rehabilitation*, 4(3), 119-136.

<https://doi.org/10.1080/13638490110091347>

APA Dictionary of Psychology. (2007). American Psychological Association.

Argulewicz, E. N., & Abel, R. R. (1984). Internal evidence of bias in the PPVT-R for

Anglo-American and Mexican-American children. *Journal of School Psychology*,

22(3), 299-303-303. [https://doi.org/10.1016/0022-4405\(84\)90010-4](https://doi.org/10.1016/0022-4405(84)90010-4)

Asparouhov, T., & Muthén, B. (2014a). Auxiliary variables in mixture modeling: Three-

step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary*

Journal, 21(3), 329-341. <https://doi.org/10.1080/10705511.2014.915181>

Asparouhov, T., & Muthén, B. (2014b). Multiple-group factor analysis Alignment.

Structural Equation Modeling: A Multidisciplinary Journal, 21(4), 495-508.

<https://doi.org/10.1080/10705511.2014.919210>

Aspinall, P. J. (2018). What kind of mixed race/ethnicity data is needed for the 2020/21

global population census round: The cases of the UK, USA, and Canada? *Ethnic and Racial Studies*, 41(11), 1990-2008.

<https://doi.org/10.1080/01419870.2017.1346267>

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89-195). Academic Press.
- [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Australian Bureau of Statistics. (2020). *Growing up in Australia: The longitudinal study of Australian children—Data issues waves 1 to 8*.
- Aviles, A. M., Anderson, T. R., & Davila, E. R. (2006). Child and adolescent social-emotional development within the context of school. *Child & Adolescent Mental Health, 11*(1), 32-39. <https://doi.org/10.1111/j.1475-3588.2005.00365.x>
- Baddeley, A. (1986). *Working memory*. Clarendon Press/Oxford University Press.
- Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485-493. <https://doi.org/10.1037/0894-4105.8.4.485>
- Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education, 10*(1), 10-33.
- <https://doi.org/10.1111/mbe.12100>
- Bakk, Z. (2024). Latent class analysis with measurement invariance testing: Simulation study to compare overall likelihood ratio vs residual fit statistics based model selection. *Structural Equation Modeling: A Multidisciplinary Journal, 31*(2), 253-264. <https://doi.org/10.1080/10705511.2023.2233115>
- Bando, R., Lopez-Boo, F., Fernald, L., Reynolds, S., & Gertler, P. (2024). Gender differences in early child development: Evidence from large-scale studies of very

young children in nine countries. *Journal of Economics, Race, and Policy*.

<https://doi.org/10.1007/s41996-023-00131-1>

Bandura, A. (1978). Social learning theory. *Journal of School Psychology, 16*, 282-282.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*.
Prentice-Hall, Inc.

Barbu, S., Nardy, A., Chevrot, J. P., Guellai, B., Glas, L., Juhel, J., & Lemasson, A. (2015).

Sex differences in language across early childhood: Family socioeconomic status does not impact boys and girls equally. *Frontiers in Psychology, 6*, 1874.

<https://doi.org/10.3389/fpsyg.2015.01874>

Barkley, R. A. (1997). Behavioural inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65-94. <https://doi.org/10.1037/0033-2909.121.1.65>

Barkley, R. A. (2001). The executive functions and self-regulation: An evolutionary neuropsychological perspective. *Neuropsychology Review, 11*(1), 1-29.
<https://doi.org/10.1023/A:1009085417776>

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved*. The Guilford Press.

Barkley, R. A. (2021). *Executive functioning: Critical issues for understanding and managing deficits*. <http://www.continuingcourses.net/active/courses/course069.php>

Barnett, M. A., Gustafsson, H., Deng, M., Mills-Koonce, W. R., & Cox, M. (2012).

Bidirectional associations among sensitive parenting, language development, and social competence. *Infant & Child Development*, 21(4), 374-393.

<https://doi.org/10.1002/icd.1750>

Bassett, H. H., Denham, S., Mincic, M., & Graling, K. (2012). The structure of preschoolers' emotion knowledge: Model equivalence and validity using a structural equation modeling approach. *Early Education and Development*, 23(3), 259-279. <https://doi.org/10.1080/10409289.2012.630825>

Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (Vol. 21, pp. 255-283). Lawrence Erlbaum Associates Publishers.

Bauer, J. (2022). A primer to latent profile and latent class analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for Researching Professional Learning and Development: Challenges, Applications and Empirical Illustrations* (pp. 243-268). Springer International Publishing. https://doi.org/10.1007/978-3-031-08518-5_11

Beauchamp, M. H., & Anderson, V. (2010). SOCIAL: An integrative framework for the development of social skills. *Psychological Bulletin*, 136(1), 39-64.

<https://doi.org/10.1037/a0017768>

Bellanti, C. J., & Bierman, K. L. (2000). Disentangling the impact of low cognitive ability and inattention on social behavior and peer relationships. *Journal of Clinical*

Child Psychology, 66-75. https://doi.org/10.1207/S15374424jccp2901_7

Belmont, J. M., & Butterfield, E. C. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology*, 2(4), 411-420.

[https://doi.org/10.1016/0010-0285\(71\)90024-7](https://doi.org/10.1016/0010-0285(71)90024-7)

Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, 9(2), 291-319. <https://doi.org/10.1017/s095457949700206x>

Bergman, L. R., Magnusson, D., & El-Khoury, B. M. (2003). *Studying individual development in an interindividual context: A person-oriented approach*.

Lawrence Erlbaum Associates Publishers.

Bergman, L. R., & Trost, K. (2006). The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Quarterly*, 52(3), 601-632.

<https://doi.org/10.1353/mpq.2006.0023>

Bergman, L. R., Von Eye, A., & Magnusson, D. (2015). Person-oriented research strategies in developmental psychopathology. In *Developmental Psychopathology* (pp. 850-888). <https://doi.org/10.1002/9780470939383.ch21>

Berk, L. E. (2013). *Child development* (9th ed.). Pearson.

Berk, L. E. (2016). *Infants and children: prenatal through middle childhood* (7th ed.). Pearson.

Berlin, K. S., Williams, N. A., & Parra, G. R. (2013). An introduction to Latent Variable

Mixture Modeling (Part 1): Overview and cross-sectional Latent Class and Latent Profile Analyses. *Journal of Pediatric Psychology*, 39(2), 174-187.

<https://doi.org/10.1093/jpepsy/jst084>

Berndt, T. J. (2002). Friendship quality and social development. *Current Directions in Psychological Science*, 11(1), 7-10. <https://doi.org/10.1111/1467-8721.00157>

Bettencourt, A. F., Musci, R. J., Masyn, K. E., & Farrell, A. D. (2022). Latent classes of aggression and peer victimization: Measurement invariance and differential item functioning across sex, race-ethnicity, cohort, and study site. *Child Development*, 93(2), e117-e134. <https://doi.org/10.1111/cdev.13691>

Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, 11(4), 485-508.

<https://doi.org/10.1080/15248372.2010.516420>

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., Morgan, C. L., & Faraone, S. V. (2004). Impact of executive function deficits and Attention-Deficit/Hyperactivity Disorder (ADHD) on academic outcomes in children. *Journal of Consulting & Clinical Psychology*, 72(5), 757-766.

<https://doi.org/10.1037/0022-006X.72.5.757>

Bierman, K. L. (1988). The clinical implications of children's conceptions of social relationships. In S. R. Shirk (Ed.), *Cognitive Development and Child Psychotherapy* (pp. 247-272). Springer US. <https://doi.org/10.1007/978-1-4899->

- Bierman, K. L., & Erath, S. A. (2006). Promoting social competence in early childhood: Classroom curricula and social skills coaching programs. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development*. (pp. 595-615). Blackwell Publishing. <https://doi.org/10.1002/9780470757703.ch29>
- Birch, S. H., & Ladd, G. W. (1998). Children's interpersonal behaviors and the teacher-child relationship. *Developmental psychology*, *34*(5), 934-946. <https://doi.org/10.1037/0012-1649.34.5.934>
- Bird, A., Reese, E., Schaughency, E., Waldie, K., Atatoa-Carr, P., Morton, S., & Grant, C. (2024). Talking, praising and teaching: How parent interaction during a learning task relates to children's early learning. *Early Childhood Research Quarterly*, *66*, 255-268. <https://doi.org/10.1016/j.ecresq.2023.10.001>
- Bird, A. L., Carr, P. E. A., Reese, E., & Morton, S. M. B. (2016). Policy translation for early childhood education and care: the Growing Up in New Zealand approach. *International Journal of Child Care and Education Policy*, *10*(1), 5. <https://doi.org/10.1186/s40723-016-0021-7>
- Bishop, R., Berryman, M., Cavanagh, T., & Teddy, L. (2009). Te Kotahitanga: Addressing educational disparities facing Māori students in New Zealand. *Teaching and Teacher Education*, *25*(5), 734-742. <https://doi.org/10.1016/j.tate.2009.01.009>
- Black-Hawkins, K. (2017). Understanding inclusive pedagogy. In V. Plows & B. Whitburn (Eds.), *Inclusive Education: Making Sense of Everyday Practice* (pp. 13-28).

SensePublishers. https://doi.org/10.1007/978-94-6300-866-2_2

- Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *The American psychologist*, 57(2), 111-127. <https://doi.org/10.1037//0003-066x.57.2.111>
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78(2), 647-663. <https://doi.org/10.1111/j.1467-8624.2007.01019.x>
- Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. Wiley.
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-634.
<https://doi.org/10.1146/annurev.psych.53.100901.135239>
- Bosacki, S. L., & Moore, C. (2004). Preschoolers' understanding of simple and complex emotions: Links with gender and language. *Sex Roles: A Journal of Research*, 50(9-10), 659-675. <https://doi.org/10.1023/b:sers.0000027568.26966.27>
- Boven, N., Exeter, D., Sporle, A., & Shackleton, N. (2020). The implications of different ethnicity categorisation methods for understanding outcomes and developing policy in New Zealand. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 15(1), 123-139. <https://doi.org/10.1080/1177083X.2019.1657912>
- Bowen, N. K., & Masa, R. D. (2015). Conducting measurement invariance tests with

- ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research*, 6(2), 229-249. <https://doi.org/10.1086/681607>
- Bowlby, J. (1980). *Attachment and loss*. Basic Books.
- Bowlby, J. (1982). Attachment and loss: Retrospect and prospect. *American Journal of Orthopsychiatry*, 52(4), 664-678. <https://doi.org/10.1111/j.1939-0025.1982.tb01456.x>
- Brandao, N., & Theodotou, E. (2020). The Reggio Emilia and the Mosaic approach: Opponents or allies in multimodal teaching and learning? A discussion of their contribution to multimodal learning in early years education. *Journal of Global Education and Research*, 4, 1-13. <https://doi.org/10.5038/2577-509X.4.1.1047>
- Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning-related behaviours, and engagement in kindergarten. *Early Childhood Research Quarterly*, 24(3), 337-349. <https://doi.org/10.1016/j.ecresq.2009.06.001>
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental psychology*, 22(6), 723-742. <https://doi.org/10.1037/0012-1649.22.6.723>
- Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualised in

- developmental perspective: a bioecological model. *Psychological Review*, *101*(4), 568-586. <https://doi.org/10.1037/0033-295x.101.4.568>
- Brooks, B., McBee, M., Pack, R., & Alamian, A. (2017). The effects of rurality on substance use disorder diagnosis: A multiple-groups latent class analysis. *Addictive Behaviors*, *68*, 24-29. <https://doi.org/10.1016/j.addbeh.2017.01.019>
- Brown, G. T. L., Harris, L. R., O'Quin, C., & Lane, K. E. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. *International Journal of Research & Method in Education*, *40*(1), 66-90. <https://doi.org/10.1080/1743727X.2015.1070823>
- Brownell, C. A., Svetlova, M., & Nichols, S. (2009). To share or not to share: When do toddlers respond to another's needs? *Infancy: the official journal of the International Society on Infant Studies*, *14*(1), 117-130. <https://doi.org/10.1080/15250000802569868>
- Bruce, M., & Bell, M. A. (2022). Vocabulary and executive functioning: A scoping review of the unidirectional and bidirectional associations across early childhood. *Human Development*, *66*(3), 167-187. <https://doi.org/10.1159/000524964>
- Bruner, J. S. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy*, *1*(1), 111-114.
- Bruner, J. S. (1990). *Acts of meaning: Four lectures on mind and culture* (Vol. 3). Harvard University Press.
- Brusco, M. J., Shireman, E., & Steinley, D. (2017). A comparison of latent class, K-means,

and K-median methods for clustering dichotomous data. *Psychological methods*, 22(3), 563-580. <https://doi.org/10.1037/met0000095>

Buckley, J., Peterson, E., Underwood, L., D'Souza, S., Morton, S., & Waldie, K. (2019).

Socio-demographic and maternal health indicators of inhibitory control in preschool age children: evidence from Growing Up in New Zealand.

Longitudinal and Life Course Studies, 11.

<https://doi.org/10.1332/175795919X15746664055477>

Bukowski, W. M. (2003). Peer relationships. In M. H. Bornstein, L. Davidson, C. L. M.

Keyes, & K. A. Moore (Eds.), *Well-being: Positive development across the life course*. (pp. 221-233). Lawrence Erlbaum Associates Publishers.

<https://doi.org/10.4324/9781410607171>

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and

executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, 33(3), 205-228.

<https://doi.org/10.1080/87565640801982312>

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child*

Development Perspectives, 8(1), 36-41. <https://doi.org/10.1111/cdep.12059>

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor

covariance and mean structures: The issue of partial measurement invariance.

Psychological Bulletin, 105(3), 456-466. <https://doi.org/10.1037/0033->

[2909.105.3.456](https://doi.org/10.1037/0033-2909.105.3.456)

- Cacioppo, J. T. (2002). Social neuroscience: Understanding the pieces fosters understanding the whole and vice versa. *American Psychologist*, *57*(11), 819-831. <https://doi.org/10.1037/0003-066x.57.11.819>
- Cairns, R. B. (1979). *Social development: The origins and plasticity of interchanges*. W. H. Freeman.
- Callister, P., Didham, R., Potter, D., & Blakely, T. (2007). Measuring ethnicity in New Zealand: Developing tools for health outcomes analysis. *Ethnicity & Health*, *12*(4), 299-320. <https://doi.org/10.1080/13557850701300699>
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (2 ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013567>
- Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child Development*, *83*(4), 1229-1244. <https://doi.org/10.1111/j.1467-8624.2012.01768.x>
- Cameron, M. P., & Poot, J. (2019). Towards superdiverse Aotearoa: Dimensions of past and future ethnic diversity in New Zealand and its regions. *New Zealand Population Review*, *45*, 18-45.
- Campbell, J. M., Bell, S. K., & Keith, L. K. (2001). Concurrent validity of the Peabody Picture Vocabulary Test-third edition as an intelligence and achievement screener for low SES African American children. *Assessment*, *8*(1), 85-94. <https://doi.org/10.1177/107319110100800108>

- Campbell, J. M., Brown, R. T., Cavanagh, S. E., Vess, S. F., & Segall, M. J. (2008). Evidence-based assessment of cognitive functioning in pediatric psychology. *Journal of Pediatric Psychology, 33*(9), 999-1014. <https://doi.org/10.1093/jpepsy/jsm138>
- Campos, J. J., Mumme, D. L., Kermoian, R., & Campos, R. G. (1994). A functionalist perspective on the nature of emotion. *Monographs of the Society for Research in Child Development, 59*(2/3), 284-303. <https://doi.org/10.2307/1166150>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47. <https://doi.org/10.1093/applin/l.1.1>
- Caporaso, J. S., Boseovski, J. J., & Marcovitch, S. (2019). The individual contributions of three executive function components to preschool social competence. *Infant and Child Development, 28*(4). <https://doi.org/10.1002/icd.2132>
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595-616. https://doi.org/10.1207/s15326942dn2802_3
- Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*(4), 299-319. <https://doi.org/10.1016/j.jecp.2004.01.002>
- Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive function. In P. D. Zelazo (Ed.),

Oxford Handbook of developmental psychology (Vol. 1: Body and mind, pp. 706–743). Oxford University Press.

Carr, M. (2001). *Assessment in early childhood settings: Learning stories*. Paul Chapman Publishing.

Carr, P. A., Bandara, D., Berry, S., Kingi, T. K., Grant, C., & Morton, S. (2017). Ethnic identification complexity across generations: Evidence from Growing Up in New Zealand. *New Zealand Population Review*, 43, 35-61.

Castellanos, F. X., Sonuga-Barke, E. J., Milham, M. P., & Tannock, R. (2006). Characterizing cognition in ADHD: Beyond executive dysfunction. *Trends in Cognitive Sciences* 10(3), 117-123. <https://doi.org/10.1016/j.tics.2006.01.011>

Cazden, C. B. (2011). Dell Hymes's construct of "Communicative Competence". *Anthropology & Education Quarterly*, 42(4), 364-369. <http://www.jstor.org/stable/41410139>

Chamberland, J., Roy-Charland, A., Perron, M., & Dickinson, J. (2017). Distinction between fear and surprise: An interpretation-independent test of the perceptual-attentional limitation hypothesis. *Social Neuroscience*, 12(6), 751-768. <https://doi.org/10.1080/17470919.2016.1251964>

Champion, T. B., Hyter, Y. D., McCabe, A., & Bland-Stewart, L. M. (2003). "A Matter of Vocabulary": Performances of low-income African American Head Start children on the Peabody Picture Vocabulary Test—III. *Communication Disorders Quarterly*, 24(3), 121-127. <https://doi.org/10.1177/15257401030240030301>

Chaplin, T. M., Cole, P. M., & Zahn-Waxler, C. (2005). Parental socialization of emotion expression: Gender differences and relations to child adjustment. *Emotion, 5*(1), 80-88. <https://doi.org/10.1037/1528-3542.5.1.80>

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464-504. <https://doi.org/10.1080/10705510701301834>

Cheung, R. W., Hartley, C., & Monaghan, P. (2022). Receptive and expressive language ability differentially support symbolic understanding over time: Picture comprehension in late talking and typically developing children. *Journal of Experimental Child Psychology, 214*, 105305. <https://doi.org/10.1016/j.jecp.2021.105305>

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology 38*, 967-978. <https://doi.org/10.1037/0012-1649.38.6.967>

Chomsky, N. (1976). *Reflections on language*. Temple Smith

Chomsky, N. (2002). *On nature and language*. Cambridge University Press.

Chong, S. Y., Chittleborough, C. R., Gregory, T., Lynch, J., Mittinty, M., & Smithers, L. G. (2019). The controlled direct effect of temperament at 2-3 years on cognitive and academic outcomes at 6-7 years. *PLOS ONE, 14*(6), 1-14. <https://doi.org/10.1371/journal.pone.0204189>

Christensen, D., Taylor, C. L., Hancock, K. J., & Zubrick, S. R. (2020). School readiness is

more than the child: a latent class analysis of child, family, school and community aspects of school readiness. *Australian Journal of Social Issues*, 57, 125-143. <https://doi.org/10.1002/ajs4.138>

Christensen, D., Taylor, C. L., & Zubrick, S. R. (2017). Patterns of multiple risk exposures for low receptive vocabulary growth 4-8 years in the Longitudinal Study of Australian Children. *PLOS ONE*, 12(1), e0168804. <https://doi.org/10.1371/journal.pone.0168804>

Chung-Hall, J., & Chen, X. (2010). Aggressive and prosocial peer group functioning: Effects on children's social, school, and psychological adjustment. *Social development*, 19(4), 659-680. <https://doi.org/10.1111/j.1467-9507.2009.00556.x>

Chung, H., & Anthony, J. C. (2013). A Bayesian approach to a multiple-group latent class-profile analysis: The timing of drinking onset and subsequent drinking behaviors among U.S. adolescents. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4), 658-680. <https://doi.org/10.1080/10705511.2013.824783>

Cieciuch, J., Davidov, E., Schmidt, P., & Algesheimer, R. (2019). How to obtain comparable measures for cross-national comparisons. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 71(1), 157-186. <https://doi.org/10.1007/s11577-019-00598-7>

Clark, C., Prior, M., & Kinsella, G. (2002). The relationship between executive function abilities, adaptive behaviour, and academic achievement in children with

- externalising behaviour problems. *Journal of child psychology and psychiatry, and allied disciplines*, 43(6), 785-796. <https://doi.org/10.1111/1469-7610.00084>
- Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental psychology*, 46(5), 1176-1191. <https://doi.org/10.1037/a0019672>
- Clark, T. C., Johnson, E. A., Kekus, M., Newman, J., Patel, P. S., Fleming, T., & Robinson, E. (2014). Facilitating access to effective and appropriate care for youth with mild to moderate mental health concerns in New Zealand. *Journal of Child & Adolescent Psychiatric Nursing*, 27(4), 190-200. <https://doi.org/10.1111/jcap.12095>
- Clogg, C. C., & Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79(388), 762-771. <https://doi.org/10.1080/01621459.1984.10477093>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Cokley, K. (2007). Critical issues in the measurement of ethnic and racial identity: A referendum on the state of the field. *Journal of Counseling Psychology*, 54(3), 224-234. <https://doi.org/10.1037/0022-0167.54.3.224>
- Cole, P. M., Armstrong, L. M., & Pemberton, C. K. (2010). The role of language in the development of emotion regulation. In S. D. Calkins & M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition*. (pp. 59-77).

American Psychological Association. <https://doi.org/10.1037/12059-004>

Cole, V. T. (2017). Adapting latent profile analysis to take into account measurement noninvariance. *Multivariate Behavioral Research*, *52*(1), 115-115.

<https://doi.org/10.1080/00273171.2016.1264288>

Collie, R. J., Martin, A. J., Nassar, N., & Roberts, C. L. (2019). Social and emotional behavioural profiles in kindergarten: A population-based latent profile analysis of links to socio-educational characteristics and later achievement. *Journal of educational psychology*, *111*(1), 170-187. <https://doi.org/10.1037/edu0000262>

Collie, R. J., Martin, A. J., Roberts, C. L., & Nassar, N. (2018). The roles of anxious and prosocial behavior in early academic performance: A population-based study examining unique and moderated effects. *Learning and Individual Differences*, *62*, 141-152. <https://doi.org/10.1016/j.lindif.2018.02.004>

Collins, L., & Lanza, S. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470567333>

Connors, M. E. (2011). Attachment theory: A "secure base" for psychotherapy integration. *Journal of Psychotherapy Integration*, *21*, 348-362.

<https://doi.org/10.1037/a0025460>

Conti-Ramsden, G., & Durkin, K. (2012). Language development and assessment in the preschool period. *Neuropsychology Review*, *22*(4), 384-401.

<https://doi.org/10.1007/s11065-012-9208-z>

Cook, C. J., Howard, S. J., Scerif, G., Twine, R., Kahn, K., Norris, S. A., & Draper, C. E.

(2019). Associations of physical activity and gross motor skills with executive function in preschool children from low-income South African settings.

Developmental Science, 22(5), e12820. <https://doi.org/10.1111/desc.12820>

Corkin, M. T., Meissel, K., Peterson, E. R., Lee, K., Giacaman, N., Janicot, S., & Morton, S.

M. B. (2022). Are distinct modes of using digital technologies evident by age eight? Implications for digital divides. *Computers & Education*, 191, 104642.

<https://doi.org/10.1016/j.compedu.2022.104642>

Corkin, M. T., Peterson, E. R., Henderson, A. M. E., Waldie, K. E., Reese, E., & Morton, S.

M. B. (2021). Preschool screen media exposure, executive functions and symptoms of inattention/hyperactivity. *Journal of Applied Developmental Psychology*, 73, 101237.

<https://doi.org/10.1016/j.appdev.2020.101237>

Cormack, D., & Robson, C. (2010). *Classification and output of multiple ethnicities:*

considerations for monitoring Māori health. Te Rōpū Rangahau Hauora a Eru Pōmare.

Cornell, S. E., & Hartmann, D. (1998). *Ethnicity and race : making identities in a*

changing world. Pine Forge Press.

Cornforth, C. M., Thompson, J. M. D., Robinson, E., Waldie, K. E., Pryor, J. E., Clark, P.,

Becroft, D. M. O., Sonuga-Barke, E. J. S., & Mitchell, E. A. (2012). Children born small for gestational age are not at special risk for preschool emotion and

behaviour problems. *Early Human Development*, 88(7), 479-485.

<https://doi.org/10.1016/j.earlhumdev.2011.12.001>

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of emotions*, 2(2), 91-115.

Cross, M. (1998). Undetected communication problems in children with behavioural problems. *International Journal of Language & Communication Disorders*, 33, 509. <https://doi.org/10.3109/13682829809179477>

Cross, M. (2011). *Children with social, emotional and behavioural difficulties and communication problems: There is always a reason* (2nd ed.). Jessica Kingsley Publishers.

Culbert, J. P., Hamer, R., & Klinge, V. (1989). Factor structure of the Wechsler Intelligence Scale for Children-Revised, Peabody Picture Vocabulary Test, and Peabody Individual Achievement Test in a psychiatric sample. *Psychology in the Schools*, 26(4), 331-336-336. [https://doi.org/10.1002/1520-6807\(198910\)26:4<331::AID-PITS2310260402>3.0.CO;2-U](https://doi.org/10.1002/1520-6807(198910)26:4<331::AID-PITS2310260402>3.0.CO;2-U)

Cumming, M. M., Poling, D. V., Patwardhan, I., & Ozenbaugh, I. C. (2022). Executive function in kindergarten and development of behavioural competence: The moderating role of positive parenting practices. *Early Childhood Research Quarterly*, 60, 161-172. <https://doi.org/10.1016/j.ecresq.2022.01.008>

Cutting, A. L., & Dunn, J. (2002). The cost of understanding other people: Social cognition predicts young children's sensitivity to criticism. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 43(7), 849-860.

<https://doi.org/10.1111/1469-7610.t01-1-00047>

D'Souza, S., Underwood, L., Peterson, E. R., Buckley, J., Morton, S. M. B., & Waldie, K. E. (2019). Determinants of persistence and change in early childhood behavioural problems: The roles of parenting and maternal mental health. *Journal of Child and Family Studies*, *28*(7), 1826-1842. <https://doi.org/10.1007/s10826-019-01404-w>

D'Souza, S., Underwood, L., Peterson, E. R., Morton, S. M. B., & Waldie, K. E. (2020). The association between persistence and change in early childhood behavioural problems and preschool cognitive outcomes. *Child psychiatry and human development*, *51*(3), 416-426. <https://doi.org/10.1007/s10578-019-00953-x>

D'Souza, S., Underwood, L., Peterson, E. R., Morton, S. M. B., & Waldie, K. E. . (2019b). Persistence and change in behavioural problems during early childhood. *BMC Pediatrics*, *19*(1). <https://doi.org/10.1186/s12887-019-1631-3>

D'Souza, S., Waldie, K. E., Peterson, E. R., Underwood, L., & Morton, S. M. (2017). Psychometric properties and normative data for the preschool Strengths and Difficulties Questionnaire in two-year-old children. *Journal of Abnormal Child Psychology*, *45*(2), 345-357. <https://doi.org/10.1007/s10802-016-0176-2>

D'Souza, S., Waldie, K. E., Peterson, E. R., Underwood, L., & Morton, S. M. B. (2019c). The Strengths and Difficulties Questionnaire: Factor structure of the father-report and parent agreement in 2-year-old children. *Assessment*, *26*(6), 1059-1069. <https://doi.org/10.1177/1073191117698757>

- D'Souza, S., Waldie, K. E., Peterson, E. R., Underwood, L., & Morton, S. M. B. (2019a). Antenatal and postnatal determinants of behavioural difficulties in early childhood: Evidence from Growing Up in New Zealand. *Child Psychiatry & Human Development*, *50*(1), 45. <https://doi.org/10.1007/s10578-018-0816-6>
- Dancey, C. (2007). *Statistics without maths for psychology*. Prentice Hall.
- Daniel, J. R., Santos, A. J., Peceguina, I., & Vaughn, B. E. (2015). Affiliative structures and social competence in Portuguese preschool children. *Developmental psychology*, *51*(7), 905-912. <https://doi.org/10.1037/a0039425>
- Daniel, S. N. (2005). *Group-based modeling of development*. Harvard University Press.
- Deak, A. (2011). Brain and emotion: Cognitive neuroscience of emotions. *Review of Psychology*, *18*, 71-80.
- Delaunay-El Allam, M., Guidetti, M., Chaix, Y., & Reilly, J. (2011). Facial emotion labeling in language impaired children. *Applied Psycholinguistics*, *32*(4), 781-798. <https://doi.org/10.1017/S0142716411000063>
- Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: Contextual validation. *Child Development*, *57*(1), 194-201. <https://doi.org/10.2307/1130651>
- Denham, S. A. (1998). *Emotional development in young children*. Guilford Press.
- Denham, S. A. (2001). Dealing with feelings: Foundations and consequences of young children's emotional competence. *Early Education and Development*, *12*(1), 5-10. https://doi.org/10.1207/s15566935eed1201_1

- Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education & Development, 17*(1), 57-89. https://doi.org/10.1207/s15566935eed1701_4
- Denham, S. A. (2018). Implications of Carolyn Saarni's work for preschoolers' emotional competence. *European Journal of Developmental Psychology, 15*(6), 643-657. <https://doi.org/10.1080/17405629.2018.1479250>
- Denham, S. A., & Bassett, H. H. (2018). Implications of preschoolers' emotional competence in the classroom. In *Emotional intelligence in education: Integrating research with practice*. (pp. 135-171). Springer International Publishing/Springer Nature. https://doi.org/10.1007/978-3-319-90633-1_6
- Denham, S. A., Bassett, H. H., Brown, C., Way, E., & Steed, J. (2015). "I Know How You Feel": Preschoolers' emotion knowledge contributes to early school success. *Journal of Early Childhood Research, 13*(3), 252-262. <https://doi.org/10.1177/1476718x13497354>
- Denham, S. A., Bassett, H. H., Mincic, M., Kalb, S., Way, E., Wyatt, T., & Segal, Y. (2012). Social-emotional learning profiles of preschoolers' early school success: A person-centered approach. *Learning and Individual Differences, 22*(2), 178-189. <https://doi.org/10.1016/j.lindif.2011.05.001>
- Denham, S. A., Bassett, H. H., Zinsser, K. M., Bradburn, I. S., Bailey, C. S., Shewark, E. A., Ferrier, D. E., Liverette, K. H., Steed, J., Karalus, S. P., & Kianpour, S. (2020). Computerized social-emotional assessment measures for early childhood

settings. *Early Childhood Research Quarterly*, 51, 55-66.

<https://doi.org/10.1016/j.ecresq.2019.07.002>

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence? *Child Development*, 74(1), 238-256. <https://doi.org/10.1111/1467-8624.00533>

Denham, S. A., Bouril, B., & Belouad, F. (1994). Preschoolers' affect and cognition about challenging peer situations. *Child Study Journal*, 24(1), 1-21.

Denham, S. A., & Burton, R. (1996). A social-emotional intervention for at-risk 4-year-olds. *Journal of School Psychology*, 34(3), 225-245.

[https://doi.org/10.1016/0022-4405\(96\)00013-1](https://doi.org/10.1016/0022-4405(96)00013-1)

Denham, S. A., Caverly, S., & Schmidt, M. (2002). Preschool understanding of emotions: Contributions to classroom anger and aggression [Article]. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 43(7), 901-916.

<https://doi.org/10.1111/1469-7610.00139>

Denham, S. A., & Couchoud, E. A. (1990a). Young preschoolers' ability to identify emotions in equivocal situations. *Child Study Journal*, 20(3), 153-169.

Denham, S. A., & Couchoud, E. A. (1990b). Young preschoolers' understanding of emotions. *Child Study Journal*, 20(3), 171-192.

Denham, S. A., Ferrier, D. E., Howarth, G. Z., Herndon, K. J., & Bassett, H. H. (2016). Key considerations in assessing young children's emotional competence. *Cambridge*

Journal of Education, 46(3), 399-317.

<https://doi.org/10.1080/0305764X.2016.1146659>

Denham, S. A., & Grout, L. (1993). Socialisation of emotion: Pathway to preschoolers' emotional and social competence. *Journal of Nonverbal Behavior*, 17(3), 205-227. <https://doi.org/10.1007/BF00986120>

Denham, S. A., Kalb, S., Way, E., Warren-Khot, H., Rhoades, B. L., & Bassett, H. H. (2013). Social and emotional information processing in preschoolers: Indicator of early school success? *Early Child Development & Care*, 183(5), 667-688.

<https://doi.org/10.1080/03004430.2012.682728>

Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of preschool peer ratings. *Child Development*, 61(4), 1145-1152. <https://doi.org/10.2307/1130882>

Denham, S. A., Zahn-Waxler, C., Cummings, E. M., & Iannotti, R. J. (1991). Social competence in young children's peer relations: Patterns of development and change. *Child psychiatry and human development*, 22(1), 29-44.

<https://doi.org/10.1007/BF00706057>

Denham, S. A., Zoller, D., & Couchoud, E. A. (1994). Socialization of preschoolers' emotion understanding. *Developmental psychology*, 30(6), 928-936.

<https://doi.org/10.1037/0012-1649.30.6.928>

Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y. W., & Cordier, R. (2017).

Psychometric properties of language assessments for children aged 4-12 years:

A systematic review. *Frontiers in Psychology*, 8, 1515.

<https://doi.org/10.3389/fpsyg.2017.01515>

Department of Statistics. (1993). *New Zealand Standard Classification of Ethnicity*.

DeRosier, M. E., Kupersmidt, J. B., & Patterson, C. J. (1994). Children's academic and behavioral adjustment as a function of the chronicity and proximity of peer

rejection. *Child Development*, 65(6), 1799-1813. <https://doi.org/10.1111/j.1467->

[8624.1994.tb00850.x](https://doi.org/10.1111/j.1467-8624.1994.tb00850.x)

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In *Principles of frontal lobe function*. (pp. 466-503). Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780195134971.003.0029>

Diamond, A. (2006). The early development of executive functions. In *Lifespan cognition: Mechanisms of change*. (pp. 70-95). Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780195169539.003.0006>

Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64(1), 135-168.

<https://doi.org/10.1146/annurev-psych-113011-143750>

Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In *Executive function in preschool-age children: Integrating*

measurement, neurodevelopment, and translational research. (pp. 11-43).

American Psychological Association. <https://doi.org/10.1037/14797-002>

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control:

Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental psychobiology*, 29(4), 315-334.

[https://doi.org/10.1002/\(SICI\)1098-2302\(199605\)29:4<315::AID-](https://doi.org/10.1002/(SICI)1098-2302(199605)29:4<315::AID-DEV2>3.0.CO;2-T)

[DEV2>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-2302(199605)29:4<315::AID-DEV2>3.0.CO;2-T)

Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research & Practice*, 16, 186-202. <https://doi.org/10.1111/0938-8982.00019>

Dickinson, D. K., McCabe, A., & Sprague, K. (2003). Teacher Rating of Oral Language and Literacy (TROLL): Individualizing Early Literacy Instruction with a Standards-Based Rating Tool. *The Reading Teacher*, 56(6), 554-564.

<http://www.jstor.org/stable/20205246>

Didham, R. (2005). *Understanding and working with ethnicity data*. Statistics New Zealand.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121-149. <https://doi.org/10.1177/0748175610373459>

Dionne, G., Tremblay, R., Boivin, M., Laplante, D., & Pérusse, D. (2003). Physical aggression and expressive vocabulary in 19-month-old twins. *Developmental psychology*, 39(2), 261-273. <https://doi.org/10.1037/0012-1649.39.2.261>

Dixon, L. Q. (2011). The role of home and school factors in predicting English vocabulary among bilingual kindergarten children in Singapore. *Applied*

Psycholinguistics, 32(1), 141-168. <https://doi.org/10.1017/S0142716410000329>

Dixon, L. Q., Zhao, J., Quiroz, B. G., & Shin, J.-Y. (2012). Home and community factors influencing bilingual children's ethnic language vocabulary development.

International Journal of Bilingualism, 16(4), 541-565.

<https://doi.org/10.1177/1367006911429527>

Dodge, K. A. (1985). Facets of social interaction and the assessment of social competence in children. In *Children's peer relations: Issues in assessment and intervention* (pp. 3-22). Springer.

Dodge, K. A., Pettit, G. S., McClaskey, C. L., Brown, M. M., & Gottman, J. M. (1986). Social Competence in Children. *Monographs of the Society for Research in Child Development*, 51(2), i-85. <https://doi.org/10.2307/1165906>

Dollaghan, C. A., & Campbell, T. F. (2009). How well do poor language scores at ages 3 and 4 predict poor language scores at age 6? *International Journal of Speech-Language Pathology*, 11(5), 358-365.

<https://doi.org/10.1080/17549500903030824>

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: a randomized trial of the preschool "PATHS" curriculum. *Journal of Prevention*, 28(2), 67-91. [https://doi.org/10.1007/s10935-](https://doi.org/10.1007/s10935-007-0081-0)

[007-0081-0](https://doi.org/10.1007/s10935-007-0081-0)

Dong Hwa, C., & Juhu, K. (2003). Practicing social skills training for young children with low peer acceptance: A cognitive-social learning model. *Early Childhood*

Education Journal, 31(1), 41. <https://doi.org/10.1023/A:1025184718597>

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In *Differential item functioning*. (pp. 35-66). Lawrence Erlbaum Associates, Inc. <https://doi.org/10.1075/z.62.13kok>

Downs, A., Strand, P., Heinrichs, N., & Cerna, S. (2012). Use of the teacher version of the Strengths and Difficulties Questionnaire with German and American preschoolers. *Early Education & Development*, 23(4), 493-516.
<https://doi.org/10.1080/10409289.2010.532082>

Dragovic, M. (2004). Towards an improved measure of the Edinburgh Handedness Inventory: A one-factor congeneric measurement model using confirmatory factor analysis. *Laterality*, 9(4), 411-419.
<https://doi.org/10.1080/13576500342000248>

Duncombe, M., Havighurst, S. S., Holland, K. A., & Frankling, E. J. (2013). Relations of emotional competence and effortful control to child disruptive behavior problems. *Early Education and Development*, 24(5), 599-615-615.
<https://doi.org/10.1080/10409289.2012.701536>

Dunn, D. M. (2021). Peabody Picture Vocabulary Test, Fifth Edition. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The twenty-first mental measurements yearbook*.

Dunn, L. M. (1959). *Peabody picture vocabulary test*. American Guidance Service.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test--Fourth Edition (PPVT-4)*. APA PsycTests. <https://doi.org/10.1037/t15144-000>

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. American Guidance Service, Inc.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test--Third Edition (PPVT-III)* APA PsycTests. <https://doi.org/10.1037/t15145-000>

Durie, M. (1994). *Whaiora: Māori Health Development*. Oxford University Press.

Durie, M. (2005). Race and ethnicity in public policy: Does it work. *Social policy journal of New Zealand*, 24(24), 1-11.

Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Weissberg, R. P., & Schellinger, K. B. (2011).

The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405-432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>

Ebert, K. D. (2017). Convergence between parent report and direct assessment of language and attention in culturally and linguistically diverse children. *PLOS ONE*, 12(7), e0180598. <https://doi.org/10.1371/journal.pone.0180598>

Education Counts. (2021, April). *Ethnic Group Codes*.

https://www.educationcounts.govt.nz/data-services/code-sets-and-classifications/ethnic_group_codes

Education Review Office. (2024). *Let's keep talking: Oral language development in the early years*. <https://evidence.ero.govt.nz/documents/lets-keep-talking-oral-language-development-in-the-early-years#Acknowledgements>

Edwards, C., Gandini, L., & Forman, G. (2011). *The Hundred Languages of Children: The*

Reggio Emilia Experience in Transformation (3rd ed.).

Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, 47(3-4), 313-337. <https://doi.org/10.1111/j.1469-7610.2006.01618.x>

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, 65(2), 241-261. <https://doi.org/10.1007/BF02294377>

Eisenberg, N., Damon, W., Lerner, R. M., & (Eds). (2006). Social, emotional, and personality development. In *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 6th ed.* (pp. xxiv, 1128-xxiv, 1128). John Wiley & Sons, Inc.

Eisenberg, N., Eggum-Wilkens, N. D., & Spinrad, T. L. (2015). The development of prosocial behavior. In *The Oxford handbook of prosocial behavior.* (pp. 114-136). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195399813.001.0001>

Eisenberg, N., Hofer, C., & Vaughan, J. (2007). Effortful control and its socioemotional consequences. In J. J. Gross (Ed.), *Handbook of emotion regulation.* (pp. 287-306). The Guilford Press.

Eisenberg, N., & Spinrad, T. L. (2004). Emotion-related regulation: Sharpening the definition. *Child Development*, 75(2), 334-339. <https://doi.org/10.1111/j.1467-8624.2004.00674.x>

- Eivers, A. R., Brendgen, M., Vitaro, F., & Borge, A. I. H. (2012). Concurrent and longitudinal links between children's and their friends' antisocial and prosocial behavior in preschool. *Early Childhood Research Quarterly, 27*(1), 137-146.
<https://doi.org/10.1016/j.ecresq.2011.05.001>
- Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation, 19*, 207-283.
- Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. Times Books/Henry Holt and Co.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124-129.
<https://doi.org/10.1037/h0030377>
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*(3875), 86-88.
<https://doi.org/10.1126/science.164.3875.86>
- Ellis, K., Jones, F. W., & Mallett, J. (2014). Differences in the factor structure of the strengths and difficulties questionnaire in Northern Irish children. *Peace and Conflict: Journal of Peace Psychology, 20*(3), 330-333.
<https://doi.org/10.1037/pac0000041>
- Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLOS ONE, 7*(10), e48433.
<https://doi.org/10.1371/journal.pone.0048433>

- England-Mason, G., Andrews, K., Atkinson, L., & Gonzalez, A. (2023). Emotion socialization parenting interventions targeting emotional competence in young children: A systematic review and meta-analysis of randomized controlled trials. *Clinical Psychology Review, 100*. <https://doi.org/10.1016/j.cpr.2023.102252>
- Espy, K. A. (1997). The shape school: Assessing executive function in preschool children. *Developmental Neuropsychology, 13*(4), 495-499. <https://doi.org/10.1080/87565649709540690>
- Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*(1), 465-486. https://doi.org/10.1207/s15326942dn2601_6
- Espy, K. A., Stalets, M. M., McDiarmid, M. M., Senn, T. E., Cwik, M. F., & Hamby, A. (2002). Executive functions in preschool children born preterm: Application of cognitive neuroscience paradigms. *Child Neuropsychology, 8*(2), 83-92. <https://doi.org/10.1076/chin.8.2.83.8723>
- Eynde, P. O. t., & Turner, J. E. (2006). Focusing on the complexity of emotion issues in academic learning: A dynamical component systems approach. *Educational Psychology Review, 18*(4), 361-376. <https://doi.org/10.1007/s10648-006-9031-2>
- Fabes, R. A., Gaertner, B. M., & Popp, T. K. (2006). Getting along with others: social competence in early childhood. In K. McCartney. & D. Phillips (Eds.), *Blackwell handbook of early childhood development*. (pp. 297-316). Blackwell Publishing.

<https://doi.org/10.1002/9780470757703.ch15>

Fabes, R. A., Leonard, S. A., Kupanoff, K., & Martin, C. L. (2001). Parental coping with children's negative emotions: relations with children's emotional and social responding. *Child Development, 72*(3), 907-920. <https://doi.org/10.1111/1467-8624.00323>

Fabes, R. A., Martin, C. L., Hanish, L. D., Anders, M. C., & Madden-Derdich, D. A. (2003). Early school competence: The roles of sex-segregated play and effortful control. *Developmental psychology, 39*(5), 848-858. <https://doi.org/10.1037/0012-1649.39.5.848>

Fan, L., & Wang, Y. (2023). The relationship between executive functioning and attention deficit hyperactivity disorder in young children: A cross-lagged study. *Current Psychology, 42*(25), 21375-21383-21383. <https://doi.org/10.1007/s12144-022-03233-5>

Fantuzzo, J., Bulotsky, R., McDermott, P., Mosca, S., & Lutz, M. N. (2003). A multivariate analysis of emotional and behavioral adjustment and preschool educational outcomes. *School Psychology Review, 32*(2), 185-203. <https://doi.org/10.1080/02796015.2003.12086193>

Feeney, R., Desha, L., Khan, A., Ziviani, J., & Nicholson, J. (2015). Speech and language difficulties along with other child and family factors associated with health-related quality of life of Australian children. *Applied Research in Quality of Life, 11*. <https://doi.org/10.1007/s11482-015-9443-6>

- Feldman, H. M. (2019). How young children learn language and speech. *Pediatrics in review*, 40(8), 398-411. <https://doi.org/10.1542/pir.2017-0325>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), 32-60. <https://doi.org/10.1111/j.1540-5834.1994.tb00173.x>
- Ferguson, S. L., G. Moore, E. W., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development*, 44(5), 458-468. <https://doi.org/10.1177/0165025419881721>
- Fey, C. F., Hu, T., & Delios, A. (2023). The measurement and communication of effect sizes in management research. *Management and Organization Review*, 19(1), 176-197. <https://doi.org/10.1017/mor.2022.2>
- Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement*, 29(4), 278-295. <https://doi.org/10.1177/0146621605275728>
- Finch, H. (2015). A comparison of statistics for assessing model invariance in latent class analysis. *Open Journal of Statistics*, 5, 191-210. <https://doi.org/10.4236/ojs.2015.53022>
- Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(1), 132-151.

<https://doi.org/10.1080/10705511.2011.532732>

Fischer, K. W., & Bidell, T. R. (1998). Dynamic development of psychological structures in action and thought. In *Handbook of child psychology: Theoretical models of human development, Volume 1, 5th ed.* (pp. 467-561). John Wiley & Sons Inc.

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action and thought. In R. M. Lerner (Ed.), *Handbook of child psychology, Vol. 1: Theoretical models of human development* (6th ed., pp. 313-399). Wiley.

Fischer, K. W., & Daley, S. (2006). Connecting cognitive science and neuroscience to education: Potentials and pitfalls in inferring executive processes. In L. Meltzer (Ed.), *Executive function in education: From theory to practice*. Guilford.

Fletcher, J. M. (1996). Executive functions in children: Introduction to the special series. *Developmental Neuropsychology, 12*(1), 1-3.

<https://doi.org/10.1080/87565649609540636>

Fletcher, J. M., & Taylor, H. G. (1984). Neuropsychological approaches to children: towards a developmental neuropsychology. *Journal of Clinical Neuropsychology, 6*(1), 39-56. <https://doi.org/10.1080/01688638408401195>

Flouri, E., Ruddy, A., & Midouhas, E. (2017). Maternal depression and trajectories of child internalising and externalising problems: The roles of child decision making and working memory. *Psychological Medicine, 47*(6), 1138-1148.

<https://doi.org/10.1017/S0033291716003226>

Foster-Cohen, S. H. (2013). *The communicative competence of young children : a*

modular approach. Routledge.

Foster, W. A., & Miller, M. (2007). Development of the literacy achievement gap: A longitudinal study of kindergarten through third grade. *Language, speech, and hearing services in schools*, *38*(3), 173-181. [https://doi.org/10.1044/0161-1461\(2007/018\)](https://doi.org/10.1044/0161-1461(2007/018))

Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A., & Bolhuis, J. J. (2017). Language, mind and brain. *Nature Human Behaviour*, *1*(10), 713-722. <https://doi.org/10.1038/s41562-017-0184-4>

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186-204. <https://doi.org/10.1016/j.cortex.2016.04.023>

Friedrich, M., & Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and Language*, *114*(2), 66-71. <https://doi.org/10.1016/j.bandl.2009.07.004>

Fujiki, M., Brinton, B., Isaacson, T., & Summers, C. (2001). Social behaviors of children with language impairment on the playground: a pilot study. *Language, Speech & Hearing Services in Schools*, *32*(2), 101-113. [https://doi.org/10.1044/0161-1461\(2001/008\)](https://doi.org/10.1044/0161-1461(2001/008))

Fullerton, J. (2021). *Bridging the gaps in education data*.

Fung, W. K., Chung, K. K. H., Lam, I. C. B., & Li, N. X. (2020). Bidirectionality in kindergarten children's school readiness and emotional regulation. *Social*

development, 29(3), 801-817. <https://doi.org/10.1111/sode.12434>

Gagnon, C., Craig, W. M., Tremblay, R. E., Zhou, R. M., & Vitaro, F. (1995). Kindergarten predictors of boys' stable behavior problems at the end of elementary school.

Journal of Abnormal Child Psychology, 23, 751-766.

<https://doi.org/10.1007/BF01447475>

Gagnon, M., Gosselin, P., Hudon-ven der Buhs, I., Larocque, K., & Milliard, K. (2010).

Children's recognition and discrimination of fear and disgust facial expressions.

Journal of Nonverbal Behavior, 34(1), 27-42. <https://doi.org/10.1007/s10919->

[009-0076-z](https://doi.org/10.1007/s10919-009-0076-z)

Gallagher, E. N., & Vella-Brodrick, D. A. (2008). Social support and emotional

intelligence as predictors of subjective well-being. *Personality and Individual*

Differences, 44(7), 1551-1561. <https://doi.org/10.1016/j.paid.2008.01.011>

Galloway, R., Reynolds, B., & Williamson, J. (2016). Strengths-based teaching and

learning approaches for children: A New Zealand case study. In S. Fan & J.

Fielding (Eds.), *What is Next in Educational Research?* (pp. 213-222).

SensePublishers. https://doi.org/10.1007/978-94-6300-524-1_19

Galvin, A., Davis, G., Neumann, D., Underwood, L., Peterson, E. R., Morton, S., & Waldie,

K. E. (2020). Risk factors associated with language delay in preschool children.

International Journal for Research in Learning Disabilities, 4(2), 35-52.

<https://doi.org/10.28987/ijrld.4.2.35>

Garner, P. W., Robertson, S., & Smith, G. (1997). Preschool children's emotional

expressions with peers: The roles of gender and emotion socialisation. *Sex Roles: A Journal of Research*, 36(11-12), 675-691.

<https://doi.org/10.1023/a:1025601104859>

Garner, P. W., & Waajid, B. (2008). The associations of emotion knowledge and teacher-child relationships to preschool children's school-related developmental competence. *Journal of Applied Developmental Psychology*, 29(2), 89-100.

<https://doi.org/10.1016/j.appdev.2007.12.001>

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, 134(1), 31-60.

<https://doi.org/10.1037/0033-2909.134.1.31>

Garraza, L. G., Azur, M., Stephens, R. L., & Walrath, C. M. (2011). Gender differences in patterns of child risk across programmatic phases of the CMHI: A multiple group latent class analysis (LCA). *The journal of behavioral health services & research*, 38(2), 265-277. <https://doi.org/10.1007/s11414-010-9219-6>

Gerde, H. K., & Powell, D. R. (2009). Teacher education, book-reading practices, and children's language growth across one year of Head Start. *Early Education and Development*, 20(2), 211-237. <https://doi.org/10.1080/10409280802595417>

Gergely, G., & Watson, J. S. (1999). Early socio-emotional development: Contingency perception and the social-biofeedback model. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life*. (pp. 101-136).

Lawrence Erlbaum Associates Publishers.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2-7 years old on a Stroop-like day-night test. *Cognition*, *53*(2), 129-153. [https://doi.org/10.1016/0010-0277\(94\)90068-X](https://doi.org/10.1016/0010-0277(94)90068-X)

Gibbs, C. J. (2023). The whole child and the whole teacher within the dynamics of the place – here and beyond. *He Kupu*, *7*(3).

Giddens, A. (1997). *Sociology* (3rd ed.). Polity Press.

Girouard, N., Stack, D., & O'Neill-Gilbert, M. (2011). Ethnic differences during social interactions of preschoolers in same-ethnic and cross-ethnic dyads. *European Journal of Developmental Psychology*, *8*(2), 185-202. <https://doi.org/10.1080/17405620903506988>

Glasser, H. M., & Smith, J. P. (2008). On the vague meaning of “gender” in education research: The problem, its sources, and recommendations for practice. *Educational Researcher*, *37*(6), 343-350. <https://doi.org/10.3102/0013189x08323718>

Golden, C. J., Sweet, J., Hammeke, T., Purisch, A., Graber, B., & Osmon, D. (1980). Factor analysis of the Luria-Nebraska Neuropsychological Battery: I. Motor, rhythm, and tactile scales. *International Journal of Neuroscience*, *11*(2), 91-99. <https://doi.org/10.3109/00207458009150331>

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of executive functioning as a theoretical and clinical construct. In S. Goldstein &

- J. A. Naglieri (Eds.), *Handbook of Executive Functioning* (pp. 3-12). Springer New York. https://doi.org/10.1007/978-1-4614-8106-5_1
- Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., & Hulme, C. (2016). The development of executive function and language skills in the early school years. *Journal of Child Psychology and Psychiatry*, *57*(2), 180-187. <https://doi.org/10.1111/jcpp.12458>
- Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(4), 400-403. <https://doi.org/10.1097/CHI.0b013e3181985068>
- Goodman, A., Joshi, H., Nasim, B., & Tyler, C. (2015). *Social and emotional skills in childhood and their long-term effects on adult life*. University College London & Institute of Education. <http://www.eif.org.uk/wp-content/uploads/2015/03/EIF-Strand-1-Report-FINAL1.pdf>
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of child psychology and psychiatry, and allied disciplines*, *38*(5), 581-586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European child & adolescent psychiatry*, *7*(3), 125. <https://doi.org/10.1007/s007870050057>
- Goodman, R., Meltzer, H., & Bailey, V. (2003). The Strengths and Difficulties

Questionnaire: A pilot study on the validity of the self-report version.

International Review of Psychiatry, 15(1/2), 173.

<https://doi.org/10.1080/0954026021000046137>

Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European child & adolescent psychiatry*, 9(2), 129.

<https://doi.org/10.1007/s007870050008>

Goyer, R. S. (1970). Communication, communicative Process meaning: Toward a unified theory. *Journal of Communication*, 20(1), 4-16.

<https://doi.org/10.1111/j.1460-2466.1970.tb00860.x>

Grammer, J. K., & Ahmed, S. F. (2023). Informing the development of school-based strategies to promote children's executive function skills: Considerations, challenges, and future directions. *Mind, Brain, and Education*, 17(4), 383-390.

<https://doi.org/10.1111/mbe.12368>

Grantham-McGregor, S. P., Cheung, Y. B. P., Cueto, S. P., Glewwe, P. P., Richter, L. P., & Strupp, B. P. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet (British edition)*, 369(9555), 60-70.

[https://doi.org/10.1016/S0140-6736\(07\)60032-4](https://doi.org/10.1016/S0140-6736(07)60032-4)

Green, V. A., & Rechis, R. (2006). Children's cooperative and competitive interactions in limited resource situations: A literature review. *Journal of Applied Developmental Psychology*, 27(1), 42-59. <https://doi.org/10.1016/j.appdev.2005.12.002>

Grissom, N. M., & Reyes, T. M. (2019). Let's call the whole thing off: Evaluating gender and sex differences in executive function. *Neuropsychopharmacology*, *44*, 86-96.

<https://doi.org/10.1038/s41386-018-0179-5>

Guerra, N. G., & Bradshaw, C. P. (2008). Linking the prevention of problem behaviors and positive youth development: Core competencies for positive youth development and risk prevention. *New Directions for Child & Adolescent Development*, *2008*(122), 1-17.

<https://doi.org/10.1002/cd.225>

Guralnick, M. J. (1999). Family and child influences on the peer-related social competence of young children with developmental delays. *Mental Retardation and Developmental Disabilities Research Review*, *5*(1), 21-29.

[https://doi.org/10.1002/\(SICI\)1098-2779\(1999\)5:1<21::AID-MRDD3>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1098-2779(1999)5:1<21::AID-MRDD3>3.0.CO;2-O)

Haggerty, M., & Alcock, S. (2016). The changing roles of early childhood care and education in Aotearoa New Zealand: A shifting policy landscape. *Global Studies of Childhood*, *6*(1), 136-146.

<https://doi.org/10.1177/2043610615625164>

Haitana, T., Pitama, S., & Rucklidge, J. J. (2010). Cultural biases in the Peabody Picture Vocabulary Test-III: Testing tamariki in a New Zealand sample. *New Zealand Journal of Psychology*, *39*(3), 24-34.

Hall, R. A. (1977). Some critiques of Chomskyan theory *Neuphilologische Mitteilungen*,

78(1), 86-95. <http://www.jstor.org/stable/43343116>

Halliday, M. A. K. (2007). *Language and Education* (Vol. 9). Continuum.

Halpin, G., Simpson, R. G., & Martin, S. L. (1990). An investigation of racial bias in the

- Peabody Picture Vocabulary Test-Revised. *Educational and Psychological Measurement*, 50(1), 183-189-189. <https://doi.org/10.1177/0013164490501022>
- Hammersley, M. L., Buchanan, L., Xu, H., & Wen, L. M. (2022). Early childhood dietary intake and subsequent socioemotional and cognitive school readiness among Australian children. *Health Education & Behavior*, 49(5), 861-870. <https://doi.org/10.1177/10901981221096100>
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72(2), 625-638. <https://doi.org/10.1111/1467-8624.00301>
- Han, H. S., & Thomas, M. S. (2010). No child misunderstood: enhancing early childhood teachers' multicultural responsiveness to the social competence of diverse children. *Early Childhood Education Journal*, 37(6), 469-476. <https://doi.org/10.1007/s10643-009-0369-1>
- Hanish, L. D., Martin, C. L., Fabes, R. A., Leonard, S., & Herzog, M. (2005). Exposure to externalizing peers in early childhood: Homophily and peer contagion processes. *Journal of Abnormal Child Psychology*, 33(3), 267-281-281. <https://doi.org/10.1007/s10802-005-3564-6>
- Hannesdottir, D. K., & Ollendick, T. H. (2007). The role of emotion regulation in the treatment of child anxiety disorders. *Clinical Child and Family Psychology Review*, 10(3), 275-293. <https://doi.org/10.1007/s10567-007-0024-6>
- Hardy, I., & Woodcock, S. (2015). Inclusive education policies: Discourses of difference,

diversity and deficit. *International Journal of Inclusive Education*, 19(2), 141-164.

<https://doi.org/10.1080/13603116.2014.908965>

Harmon-Jones, E., & Allen, J. J. B. (1998). Anger and frontal brain activity: EEG asymmetry consistent with approach motivation despite negative affective valence. *Journal of Personality and Social Psychology*, 74, 1310-1316.

<https://doi.org/10.1037/0022-3514.74.5.1310>

Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. *Educational Psychologist*, 51(3/4), 354-367.

<https://doi.org/10.1080/00461520.2016.1207176>

Harris, P. L. (1989). *Children and emotion: The development of psychological understanding*. Basil Blackwell.

Hart, B., & Risley, T. R. (1995). The early catastrophe: The 30 million word gap by age 3. *American Educator*, 27(1), 4-9.

Hartup, W. W. (1996). The company they keep: Friendships and their developmental significance. *Child Development*, 67(1), 1-13. <https://doi.org/10.2307/1131681>

Hartup, W. W., & Moore, S. G. (1990). Early peer relations: Developmental significance and prognostic implications. *Early Childhood Research Quarterly*, 5(1), 1-17.

[https://doi.org/10.1016/0885-2006\(90\)90002-1](https://doi.org/10.1016/0885-2006(90)90002-1)

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569-1579.

<https://doi.org/10.1017/CBO9780511817755.002>

Hay, D. F., Payne, A., & Chadwick, A. (2004). Peer relations in childhood. *Journal of Child Psychology and Psychiatry*, 45(1), 84-108. <https://doi.org/10.1046/j.0021-9630.2003.00308.x>

Healey, D., & Healey, M. (2019). Randomized Controlled Trial comparing the effectiveness of structured play (ENGAGE) and behaviour management (TRIPLE P) in reducing problem behaviours in preschoolers. *Scientific Reports*, 9(1), 3497. <https://doi.org/10.1038/s41598-019-40234-0>

Healey, D., Milne, B., & Healey, M. (2022). Adaption and Implementation of the Engage Programme: Teaching self-regulation through play, within the early childhood curriculum. *Research Square* <https://doi.org/10.21203/rs.3.rs-1359890/v1>

Healey, D. M., & Halperin, J. M. (2015). Enhancing Neurobehavioral Gains with the Aid of Games and Exercise (ENGAGE): Initial open trial of a novel early intervention fostering the development of preschoolers' self-regulation. *Child Neuropsychology*, 21(4), 465-480. <https://doi.org/10.1080/09297049.2014.906567>

Hendriks, P. (2013). *Asymmetries between language production and comprehension*. Springer.

Henricsson, L., & Rydell, A.-M. (2006). Children with behaviour problems: the influence of social competence and social relations on problem stability, school achievement and peer acceptance across the first six years of school. *Infant and Child Development*, 15(4), 347-366. <https://doi.org/10.1002/icd.448>

- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 202-226. <https://doi.org/10.1080/10705510709336744>
- Hill, A. L., Degnan, K. A., Calkins, S. D., & Keane, S. P. (2006). Profiles of externalizing behavior problems for boys and girls across preschool: The roles of emotion regulation and inattention. *Developmental psychology*, *42*(5), 913-928. <https://doi.org/10.1037/0012-1649.42.5.913>
- Horowitz-Kraus, T., Holland, S. K., & Freund, L. S. (2016). Imaging executive functions in typically and atypically developed children. In *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research*. (pp. 181-213). American Psychological Association. <https://doi.org/10.1037/14797-009>
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. Yale University Press.
- Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods*, *21*(4), 846-876. <https://doi.org/10.1177/1094428117744021>
- Howe, G. W., Dagne, G., Valido, A., Espelage, D. L., Abram, K. M., Brown, C. H., & Gallo, C. (2024). The impact of sparse datasets when harmonizing data from studies with different measures of the same construct. *Prevention Science*, *25*(6), 989-

1002. <https://doi.org/10.1007/s11121-024-01704-8>

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus. *Structural Equation Modeling*, 6(1), 1.

<https://doi.org/10.1080/10705519909540118>

Hudson, M., Milne, M., Reynolds, P., Russell, K., & Smith, B. (2019). *Te Ara Tika - Guidelines for Māori research ethics: A framework for researchers and ethics committee members*. Health Research Council of New Zealand,.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 16(2), 233-253.

<https://doi.org/10.1111/j.2044-835X.1998.tb00921.x>

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviours and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychology*, 108, 663-676. <https://doi.org/10.1016/j.jecp.2010.06.005>

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: problems and solutions? *Child and Adolescent Mental Health*, 7(3), 131-142.

<https://doi.org/10.1111/1475-3588.00024>

Humphrey, N., Kalambouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M. (2011). Measures of social and emotional skills for children and young people: A systematic review. *Educational and Psychological Measurement*, 71(4), 617-637. <https://doi.org/10.1177/0013164410382896>

- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581-592. <https://doi.org/10.1037/0003-066X.60.6.581>
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, *74*(2), 171-193. <https://doi.org/10.1037/amp0000307>
- Hymes, D. (1972). On Communicative Competence. In J. B. Pride & A. Holmes (Eds.), *Sociolinguistics: Selected Readings*. Penguin.
- Ibbotson, P. (2022). *Language Acquisition: The Basics* (1st ed.). Routledge. <https://doi.org/10.4324/9781003156536>
- Institute of Medicine (US) Committee. (1999). The national academies collection: Reports funded by National Institutes of Health. In M. A. Haynes & B. D. Smedley (Eds.), *The Unequal Burden of Cancer: An Assessment of NIH Research and Programs for Ethnic Minorities and the Medically Underserved*. National Academies Press (US). <https://doi.org/10.17226/6377>
- Ioannis, T., & Fathima, M. J. (2023). Identifying bias in social and health research: Measurement invariance and latent mean differences using the Alignment Approach. *Mathematics*, *11*(18), 4007-4007. <https://doi.org/10.3390/math11184007>
- Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, *12*(1), 18-23.

<https://doi.org/10.1017/s0954579401001043>

Izard, C. E. (1984). Emotion-cognition relationships and human development. *Emotions, cognition, and behavior*, 17-37.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, 99, 561-565. <https://doi.org/10.1037/0033-295X.99.3.561>

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115, 288-299.
<https://doi.org/10.1037/0033-2909.115.2.288>

Jan, L. W., Sarah, B., Polly Atatoa, C., Elizabeth, R. P., Karen, E. W., Emma, M., Stephanie, D. S., & Susan, M. B. M. (2021). Patterns of risk exposure in first 1,000 days of life and health, behaviour, and education-related problems at age 4.5: Evidence from Growing Up in New Zealand, a longitudinal cohort study. *BMC Pediatrics*, 21(1), 1-14. <https://doi.org/10.1186/s12887-021-02652-w>

Janus, M., Brinkman, S. A., & Duku, E. K. (2011). Validity and psychometric properties of the Early Development Instrument in Canada, Australia, United States, and Jamaica. *Social Indicators Research*, 103(2), 283-297.
<https://doi.org/10.1007/s11205-011-9846-1>

Janus, M., & Offord, D. R. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du*

comportement, 39(1), 1-22. <https://doi.org/10.1037/cjbs2007001>

Jeffreys, M., Smiler, K., Ellison Loschmann, L., Pledger, M., Kennedy, J., & Cumming, J.

(2022). Consequences of barriers to primary health care for children in Aotearoa

New Zealand. *SSM - Population Health*, 17, 101044.

<https://doi.org/10.1016/j.ssmph.2022.101044>

Johnson, A. (2023). Holistic learning theory: More than a philosophy. *Journal of*

Contemplative and Holistic Education, 1(2).

<https://doi.org/10.25035/jche.01.02.03>

Johnson, C., Ironsmith, M., Snow, C. W., & Poteat, G. M. (2000). Peer acceptance and

social adjustment in preschool and kindergarten. *Early Childhood Education*

Journal, 27(4), 207. <https://doi.org/10.1023/B:ECEJ.0000003356.30481.7a>

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning

and public health: The relationship between kindergarten social competence and

future wellness. *American Journal of Public Health*, 105(11), 2283-2290.

<https://doi.org/10.2105/AJPH.2015.302630>

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and

growth mixture modeling. *Social and Personality Psychology Compass*, 2(1),

302-317. <https://doi.org/10.1111/j.1751-9004.2007.00054.x>

Kafetsios, K. (2004). Attachment and emotional intelligence abilities across the life

course. *Personality and Individual Differences*, 37(1), 129-145.

<https://doi.org/10.1016/j.paid.2003.08.006>

- Kalkan, Ö. K., & Kelecioğlu, H. (2016). The effect of sample size on parametric and nonparametric factor analytical methods. *Educational Sciences: Theory and Practice*, 16(1), 153-171. <https://doi.org/10.12738/estp.2016.1.0220>
- Kam, C., Morin, A. J. S., Meyer, J. P., & Topolnytsky, L. (2016). Are commitment profiles stable and predictable? A latent transition analysis. *Journal of Management*, 42(6), 1462-1490. <https://doi.org/10.1177/0149206313503010>
- Karlsson, P., Larm, P., Svensson, J., & Raninen, J. (2022). The factor structure of the Strength and Difficulties Questionnaire in a national sample of Swedish adolescents: Comparing 3 and 5-factor models. *PLOS ONE*, 17(3), e0265481. <https://doi.org/10.1371/journal.pone.0265481>
- Keenan, K., & Shaw, D. (1997). Developmental and social influences on young girls' early problem behaviour. *Psychological Bulletin*, 121(1), 95-113. <https://doi.org/10.1037/0033-2909.121.1.95>
- Keller, H. (2018). Universality claim of attachment theory: Children's socioemotional development across cultures. *Proceedings of the National Academy of Sciences of the United States of America*, 115(45), 11414-11419. <https://doi.org/10.1073/pnas.1720325115>
- Kellermann, K. (1992). Communication: Inherently strategic and primarily automatic. *Communications Monographs*, 59(3), 288-300. <https://doi.org/10.1080/03637759209376270>
- Kersten, P., Drew, A., McPherson, K., Elder, H., Nayar, S., & Dudley, M. (2014). A

validation and norming study of the Strengths and Difficulties Questionnaire in the New Zealand context: final reports: executive summary.

<https://www.health.govt.nz/system/files/documents/publications/validation-norming-study-strengths-difficulties-questionnaire-nz-executive-summary-aug15.pdf>

Kersten, P., Vandal, A. C., Elder, H., & McPherson, K. M. (2018). Strengths and Difficulties Questionnaire: Internal validity and reliability for New Zealand preschoolers. *BMJ Open*, 8(4), e021551. <https://doi.org/10.1136/bmjopen-2018-021551>

Khademi, A., Wells, C. S., Oliveri, M. E., & Villalonga-Olives, E. (2023). Examining appropriacy of CFI and TLI cutoff value in multiple-group CFA test of measurement invariance to enhance accuracy of test score interpretation. *Sage Open*, 13(4), 21582440231205354. <https://doi.org/10.1177/21582440231205354>

Khawaja, M., Boddington, B., & Didham, R. (2007). *Growing ethnic diversity in New Zealand and its implications for measuring differentials in fertility and mortality.*

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 524-544.

<https://doi.org/10.1080/10705511.2017.1304822>

Kim, E. S., & Wang, Y. (2019). Investigating sources of heterogeneity with three-step multilevel factor mixture modeling: Beyond testing measurement invariance in cross-national studies. *Structural Equation Modeling: A Multidisciplinary Journal*,

26(2), 165-181. <https://doi.org/10.1080/10705511.2018.1521284>

Kim, M., Vermunt, J., Bakk, Z., Jaki, T., & Van Horn, M. L. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling, 23*(4), 601-614. <https://doi.org/10.1080/10705511.2016.1158655>

Kim, S., Nordling, J. K., Yoon, J. E., Boldt, L. J., & Kochanska, G. (2013). Effortful control in “hot” and “cool” tasks differentially predicts children’s behaviour problems and academic performance. *Journal of Abnormal Child Psychology, 41*(1), 43-56. <https://doi.org/10.1007/s10802-012-9661-4>

King, T. L., Scovelle, A. J., Meehl, A., Milner, A. J., & Priest, N. (2021). Gender stereotypes and biases in early childhood: A systematic review. *Australasian Journal of Early Childhood, 46*(2), 112-125. <https://doi.org/10.1177/1836939121999849>

Kline, R. B. (2013). Assessing statistical aspects of test fairness with structural equation modelling. *Educational Research and Evaluation, 19*(2-3), 204-222. <https://doi.org/10.1080/13803611.2013.767624>

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford publications.

Kochenderfer, B. J., & Ladd, G. W. (1996). Peer victimization: Cause or consequence of school maladjustment? *Child Development, 67*, 1305-1317. <https://doi.org/10.2307/1131701>

Koglen, U., Barquero, B., Mayer, H., Scheithauer, H., & Petermann, F. (2007). German version of Strengths and Difficulties Questionnaire (T4-16-SDQ): Psychometric

quality of the teacher version for preschoolers. *Diagnostica*, 53, 175-183.

<https://doi.org/10.1026/0012-1924.53.4.175>

Konold, T. R., & Pianta, R. C. (2005). Empirically-derived, person-oriented patterns of school readiness in typically-developing children: Description and prediction to first-grade achievement. *Applied Developmental Science*, 9(4), 174-187.

https://doi.org/10.1207/s1532480xads0904_1

Krieger, N. (2003). Genders, sexes, and health: What are the connections and why does it matter? *International journal of epidemiology*, 32(4), 652-657.

<https://doi.org/10.1093/ije/dyg156>

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831-843. <https://doi.org/10.1038/nrn1533>

Kukutai, T., & Callister, P. (2009). A "main" ethnic group? Ethnic self-prioritisation among New Zealand youth. *Social Policy Journal of New Zealand*(36), 16-31.

Ladd, G. W. (2007). Social learning in the peer context. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on socialization and social development in early childhood education*. (pp. 133-164). IAP Information Age Publishing.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*, 70(6), 1373-1400. <https://doi.org/10.1111/1467-8624.00101>

Lambie, I., & Krynen, A. (2017). The utility of the Strengths and Difficulties Questionnaire as a screening measure among children and adolescents who light fires. *Journal*

of Forensic Psychiatry & Psychology, 28(3), 313-330.

<https://doi.org/10.1080/14789949.2016.1275747>

Landa, R. J. (2005). Assessment of social communication skills in preschoolers. *Mental Retardation and Developmental Disabilities Research Reviews*, 11(3), 247-252.

<https://doi.org/10.1002/mrdd.20079>

Lane, R. D., & Nadel, L., (Eds.). (2000). *Cognitive neuroscience of emotion*. Oxford University Press.

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(1), 1-26.

<https://doi.org/10.1080/10705511.2013.742377>

Laursen, B. P., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly* 52(3), 377-389.

<https://doi.org/10.1353/mpq.2006.0029>

Law, Boyle, Harris, Harkness, & Nye. (1998). Screening for speech and language delay: A systematic review of the literature. *Health Technology Assessment*, 2(9).

<https://doi.org/10.3310/hta2090>

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the literature. *Developmental Medicine and Child Neurology* 42(3),

190-200. <https://doi.org/10.1017/s0012162200000335>

Lawrence, K., Campbell, R., & Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Frontiers in Psychology, 6*, 761.

<https://doi.org/10.3389/fpsyg.2015.00761>

Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford University Press.

Leather, F. (2009). Prioritising ethnic data for schools: Who are we missing? A research note. *Social Policy Journal of New Zealand., 36*, 16-31.

Levin, H. S., Fletcher, J. M., Kufera, J. A., Harward, H., Lilly, M. A., Mendelsohn, D., Bruce, D., & Eisenberg, H. M. (1996). Dimensions of cognition measured by the tower of London and other cognitive tasks in head-injured children and adolescents. *Developmental Neuropsychology, 12*(1), 17-34.

<https://doi.org/10.1080/87565649609540638>

Lewis, C., Koyasu, M., Oh, S., Ogawa, A., Short, B., & Huang, Z. (2009). Culture, executive function, and social understanding. *New Directions for Child and Adolescent Development, 2009*(123), 69-85. <https://doi.org/10.1002/cd.236>

Lewis, M. D. (2000). The promise of dynamic systems approaches for an integrated account of human development. *Child Development, 71*(1), 36-43.

<https://doi.org/10.1111/1467-8624.00116>

Lewis, M. D. (2007). Emotional habits in brain and behavior: A window on personality development. In A. Fogel, B. J. King, & S. G. Shanker (Eds.), *Human Development in the Twenty-First Century: Visionary Ideas from Systems Scientists* (pp. 72-80).

Cambridge University Press. <https://doi.org/10.1017/CBO9780511489693.009>

- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment, 5th ed.* Oxford University Press.
- Lim, A., Rumball-Smith, J., Jones, R., & Kawachi, I. (2017). The rise and fall of hospitalizations for skin infections in New Zealand, 2004–2014: Trends by ethnicity and socioeconomic deprivation. *Epidemiology & Infection*, *145*(4), 678-684. <https://doi.org/10.1017/S0950268816002685>
- Livesey, D. J., & Morgan, G. A. (1991). The development of response inhibition in 4- and 5-year-old children. *Australian Journal of Psychology*, *43*(3), 133-137. <https://doi.org/10.1080/00049539108260137>
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767-778. <https://doi.org/10.1093/biomet/88.3.767>
- Londono, M. C., Dionne, C., & Lacharité, C. (2024). Authentic assessment of executive functions in early childhood: A scoping review. *Journal of Early Intervention*. <https://doi.org/10.1177/10538151241271134>
- Longobardi, C., Settanni, M., Lin, S., & Fabris, M. A. (2021). Student–teacher relationship quality and prosocial behaviour: The mediating role of academic achievement and a positive attitude towards school. *British Journal of Educational Psychology*, *91*(2), 547-562. <https://doi.org/10.1111/bjep.12378>
- Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological methods*, *10*(1), 21-39.

<https://doi.org/10.1037/1082-989X.10.1.21>

Lukac, M., Doerflinger, N., & Pulignano, V. (2019). Developing a cross-national comparative framework for studying labour market segmentation: Measurement equivalence with latent class analysis. *Social Indicators Research*, *145*(1), 233-255. <https://doi.org/10.1007/s11205-019-02101-3>

Luria, A. R. (1963). *Restoration of function after brain injury*. Pergamon Press.

Luria, A. R. (1966). *Human brain and psychological processes*. Harper and Row Publishers.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology*. Basic Books.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). Basic Books.

Ly, K., Lai, H., Smith, A., Walker, C., Morton, S. M. B., & Neumann, D. (2020). *Growing Up in New Zealand Technical Report: Children's performance on the adapted Peabody Picture Vocabulary Test (3rd Edition)*.

Machado, P. M., M. Vepíssimo, & Denham, S. (2012). The Affect Knowledge Test for preschool children. *Revista Iberoamericana de Diagnostico y Evaluacion Psicologica*, *1*(34), 201-222.

Madsen Sjö, N., Kiil, A., & Jensen, P. (2021). Teachers' perspectives on strength-based and deficit-based instruments for assessing socioemotional development in early childhood. *Infants & Young Children: An Interdisciplinary Journal of Early Childhood Intervention*, *34*(1), 33-45.

<https://doi.org/10.1097/IYC.0000000000000180>

Magai, C., Distel, N., & Liker, R. (1995). Emotion socialisation, attachment, and patterns of adult emotional traits. *Cognition and Emotion*, *9*, 461-481.

<https://doi.org/10.1080/02699939508408976>

Magnusson, D. (1995). Individual development: A holistic, integrated model. In *Examining lives in context: Perspectives on the ecology of human development*.

(pp. 19-60). American Psychological Association. <https://doi.org/10.1037/10176-001>

Magnusson, D. (2003). The person approach: Concepts, measurement models, and research strategy. *New Directions for Child and Adolescent Development*,

2003(101), 3-23. <https://doi.org/10.1002/cd.79>

Magnusson, D., & Allen, V. L. (1983). *Human development: an interactional perspective*. Academic Press.

Magnusson, D., & Törestad, B. (1993). A holistic view of personality: A model revisited. *Annual review of psychology*, *44*, 427-452.

<https://doi.org/10.1146/annurev.ps.44.020193.002235>

Magnusson, K. R., & Brim, B. L. (2014). The aging brain. In *Reference Module in Biomedical Sciences*. Elsevier. [https://doi.org/10.1016/B978-0-12-801238-](https://doi.org/10.1016/B978-0-12-801238-3.00158-6)

[3.00158-6](https://doi.org/10.1016/B978-0-12-801238-3.00158-6)

Main, M., & Solomon, J. (1986). Discovery of an insecure-disorganized/disoriented attachment pattern. In *Affective development in infancy*. (pp. 95-124). Ablex Publishing.

- Malti, T., Gummerum, M., Keller, M., Chaparro, M. P., & Buchmann, M. (2012). Early sympathy and social acceptance predict the development of sharing in children. *PLOS ONE*, 7(12), e52017. <https://doi.org/10.1371/journal.pone.0052017>
- Mandell, D. J., & Sackett, G. P. (2009). Comparability of developmental cognitive assessments between standard and computer testing methods. *Developmental psychobiology*, 51(1), 1-13. <https://doi.org/10.1002/dev.20329>
- Marjolein, W., & Maddie, H. (2023). Teaching approaches exploring multiple pathways to holistic development. *He Kupu*, 7(3).
- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(2), 191-225. <https://doi.org/10.1080/10705510902751010>
- Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13(2), 67-70. <https://doi.org/10.1111/j.0963-7214.2004.00276.x>
- Martinez, A., & Du, S. (2012). A Model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *Journal of machine learning research: JMLR*, 13, 1589-1608.
- Martinsone, B., Supe, I., Stokenberga, I., Damberga, I., Cefai, C., Camilleri, L., Bartolo, P., O'Riordan, M. R., & Grazzani, I. (2022). Social emotional competence, learning

outcomes, emotional and behavioral difficulties of preschool children: Parent and teacher evaluations. *Frontiers in Psychology*, *12*.

<https://doi.org/10.3389/fpsyg.2021.760782>

Masten, A. S., Herbers, J. E., Desjardins, C. D., Cutuli, J. J., McCormick, C. M., Sapienza, J. K., Long, J. D., & Zelazo, P. D. (2012). Executive function skills and school success in young children experiencing homelessness. *Educational Researcher*, *41*(9), 375-384. <https://doi.org/10.3102/0013189x12459883>

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis* (pp. 551-611). Oxford University Press.

<https://doi.org/10.1093/oxfordhb/9780199934898.013.0025>

Masyn, K. E. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 180-197.

<https://doi.org/10.1080/10705511.2016.1254049>

Matias, C., O'Connor, T. G., Futh, A., & Scott, S. (2014). Observational attachment theory-based parenting measures predict children's attachment narratives independently from social learning theory-based measures. *Attachment & human development*, *16*(1), 77-92.

<https://doi.org/10.1080/14616734.2013.851333>

May, H. (2014). New Zealand: A narrative of shifting policy directions for early

childhood education and care. In L. Gambaro, K. Stewart, & J. Waldfogel (Eds.), *An Equal Start?: Providing Quality Early Education and Care for Disadvantaged Children* (pp. 147-170). Bristol University Press.

<https://doi.org/10.46692/9781447310532.007>

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. J. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications*. (pp. 3-34). Basic Books.

Mays, V. M., Ponce, N. A., Washington, D. L., & Cochran, S. D. (2003). Classification of race and ethnicity: Implications for public health. *Annual Review of Public Health*, *24*(1), 83-110. <https://doi.org/10.1146/annurev.publhealth.24.100901.140927>

McAloney-Kocaman, K., & McPherson, K. (2017). Factor structure and reliability of the parent-informant Strengths and Difficulties Questionnaire in a Scottish preschool sample. *Early Education & Development*, *28*(3), 368-376.

<https://doi.org/10.1080/10409289.2017.1228367>

McArthur, B. A., Tough, S., & Madigan, S. (2022). Screen time and developmental and behavioral outcomes for preschool children. *Pediatric research*, *91*(6), 1616-1621. <https://doi.org/10.1038/s41390-021-01572-w>

McCabe, P. C., & Meller, P. J. (2004). The relationship between language and social competence: How language impairment affects social growth. *Psychology in the Schools*, *41*(3), 313-321. <https://doi.org/10.1002/pits.10161>

McClelland, M. M., Cameron, C. E., Wanless, S. B., Murray, A., Saracho, O., & Spodek, B.

(2007). Executive function, behavioural self-regulation, and social-emotional competence. *Contemporary perspectives on social learning in early childhood education, 1*, 113-137.

McCormack, J., Harrison, L. J., McLeod, S., & McAllister, L. (2011). A nationally representative study of the association between communication impairment at 4-5 years and children's life activities at 7-9 years. *Journal of Speech, Language & Hearing Research, 54*(5), 1328-1348. [https://doi.org/10.1092-4388\(2011/10-0155\)](https://doi.org/10.1092-4388(2011/10-0155))

McCutcheon, A. (1987). *Latent Class Analysis*. SAGE Publications, Inc.
<https://doi.org/10.4135/9781412984713>

McCutcheon, A. L. (2002). Basic concepts and procedures in single- and multiple-group latent class analysis. In J. A. Hagenars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 56-86). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511499531.003>

McGovern, J., & Lowe, P. (2018). Measurement invariance. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (Vol. 4, pp. 1035-1037). SAGE Publications, Inc. <https://doi.org/10.4135/9781506326139>

McKinlay, A. (2011). Peabody Picture Vocabulary Test –Third Edition (PPVT-III). In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of Child Behavior and Development* (pp. 1072-1072). Springer US. https://doi.org/10.1007/978-0-387-79061-9_2093

McKown, C. (2019). Reliability, factor structure, and measurement invariance of a web-based assessment of children's social-emotional comprehension. *Journal of Psychoeducational Assessment, 37*(4), 435-449.

<https://doi.org/10.1177/0734282917749682>

McLachlan, C. (2018). *Te Whāriki* revisited: How approaches to assessment can make valued learning visible. *He Kupu, 5*(3), 45-56.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. Wiley.

<https://doi.org/10.1002/0471721182>

McLaughlin, T., Allen, J., & Zeng, E. (2024). *Children's learning and development data in early childhood education: An exploration of analysis methods underpinned by the principles of Te Whāriki and standards for educational assessment*. Ministry of Social Development. <https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/children-and-families-research-fund/childrens-learning-and-development-data-in-early-childhood-education.pdf>

McLaughlin, T., Aspden, K., & Clarke, L. (2017). How do teachers support children's social-emotional competence? Strategies for teachers. *Early Childhood Folio, 21*(2), 21-27. <https://doi.org/10.18296/ecf.0041>

McLaughlin, T. W., Snyder, P. A., & Algina, J. (2015). Characterizing early childhood disabilities in a nationally representative sample using functional profiles. *Exceptional Children, 81*(4), 471-488.

<https://doi.org/10.1177/0014402914563696>

McLaughlin, T. W., Snyder, P. A., & Algina, J. (2018). Examining young children's social competence using functional ability profiles. *Disability and rehabilitation*, *40*(25), 2987-2997. <https://doi.org/10.1080/09638288.2017.1363823>

McLeod, S., Harrison, L. J., & Wang, C. (2019). A longitudinal population study of literacy and numeracy outcomes for children identified with speech, language, and communication needs in early childhood. *Early Childhood Research Quarterly*, *47*, 507-517. <https://doi.org/10.1016/j.ecresq.2018.07.004>

McLeod, S., Harrison, L. J., Whiteford, C., & Walker, S. (2016). Multilingualism and speech-language competence in early childhood: Impact on academic and social-emotional outcomes at school. *Early Childhood Research Quarterly*, *34*, 53-66. <https://doi.org/10.1016/j.ecresq.2015.08.005>

McMillen, J. C., Morris, L., & Sherraden, M. (2004). Ending social work's grudge match: Problems versus strengths. *Families in Society*, *85*(3), 317-325. <https://doi.org/10.1177/104438940408500309>

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *The Journal of applied psychology*, *93*(3), 568-592. <https://doi.org/10.1037/0021-9010.93.3.568>

Meissel, K., Reese, E., & Turnbull, S. (2019). *Factors of the early learning environment that promote early learning outcomes in Aotearoa / New Zealand*. Ministry of Social Development. <https://www.msd.govt.nz/documents/about-msd-and->

[our-work/publications-resources/research/children-and-families-research-fund/earlylearningoutcomes-november2019.pdf](#)

Melis, A. P., Grocke, P., Kalbitz, J., & Tomasello, M. (2016). One for you, one for me: humans' unique turn-taking skills. *Psychological Science*, *27*(7), 987-996.
<https://doi.org/10.1177/0956797616644070>

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*(2), 127-143. [https://doi.org/10.1016/0883-0355\(89\)90002-5](https://doi.org/10.1016/0883-0355(89)90002-5)

Meltzer, L., Dunstan-Brewer, J., & Krishnan, K. (2018). Learning differences and executive function: Understandings and misunderstandings. In L. Meltzer (Ed.), *Executive function in education: From theory to practice, 2nd ed.* (pp. xix, 396-xix, 396). The Guilford Press.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by Functional MRI. *Psychological Science*, *22*(9), 1173-1182.
<https://doi.org/10.1177/0956797611418347>

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543. <https://doi.org/10.1007/BF02294825>

Metge, J. (1995). *New growth from old: The whānau in the modern world*. Victoria University Press.

Michela, B., Giovanni, L., & Verdiana, T. (2014). What do facial expressions of emotion

express in young children? The relationship between facial display and EMG measures. *Neuropsychological Trends*, 15, 7-23. <https://doi.org/10.7358/neur-2014-015-balc>

Miller, A., Fine, S., Kiely Gouley, K., Seifer, R., Dickstein, S., & Shields, A. (2006). Showing and telling about emotions: Interrelations between facets of emotional competence and associations with classroom adjustment in Head Start preschoolers. *Cognition & Emotion*, 20(8), 1170-1192. <https://doi.org/10.1080/02699930500405691>

Miller, A. L., Gouley, K. K., Seifer, R., Zakriski, A., Eguia, M., & Vergnani, M. (2005). Emotion knowledge skills in low-income elementary school children: Associations with social status and peer experiences. *Social development*, 14(4), 637-651. <https://doi.org/10.1111/j.1467-9507.2005.00321.x>

Miller, K. (2005). *Communication theories : perspectives, processes, and contexts* (2nd ed ed.). McGraw-Hill.

Mills, D., & Conboy, B. (2005). Do changes in brain organization reflect shifts in symbolic functioning? In L.Namy (Ed.), *Symbol use and symbolic representation: developmental and comparative perspectives* (pp. 123-153). Erlbaum.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, 5(3), 317-334. <https://doi.org/10.1162/jocn.1993.5.3.317>

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language

experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience*, *16*(8), 1452-1464. <https://doi.org/10.1162/0898929042304697>

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge/Taylor & Francis Group.

Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological methods*, *9*(1), 93-115. <https://doi.org/10.1037/1082-989x.9.1.93>

Min, S. H., Topaz, M., Lee, C., & Schnall, R. (2023). Racial differences in older adult's mental health and cognitive symptomatology: Identifying subgroups using multiple-group latent class analysis. *Journal of aging and health*. <https://doi.org/10.1177/08982643231212547>

Ministry of Education. (2002). *Pathways to the Future: A 10-year strategic plan for early childhood education*. Wellington: Ministry of Education

Ministry of Education. (2009). *Te whatu pōkeka: Kaupapa Māori assessment for learning: Early childhood exemplars*.

Ministry of Education. (2017). *Te Whāriki: He Whāriki Mātauranga mō ngā Mokopuna o Aotearoa / Early Childhood Curriculum*. Ministry of Education.

Ministry of Education. (2019a). *He māpuna te tamaiti: Supporting social and emotional competence in early learning*. Cognition Education for the New Zealand Ministry of Education.

Ministry of Education. (2019b). *He taonga te tamaiti: Every child a taonga: Early learning action plan 2019-2029*. New Zealand: Ministry of Education, Retrieved from

https://conversation-space.s3-ap-southeast-2.amazonaws.com/SES_0342_ELS_10YP_Final+Report_Web.pdf

Ministry of Education. (2020). *Te Tūāpapa o He Pikorua*

<https://hepikorua.education.govt.nz/how-we-work/flexible-tailored-model-of-support/>

Ministry of Health. (2017). *HISO 10001:2017 Ethnicity Data Protocols*. Wellington: Ministry of Health.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100. <https://doi.org/10.1006/cogp.1999.0734>

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2693-2698. <https://doi.org/10.1073/pnas.1010076108>

Moffitt, T. E., Poulton, R., & Caspi, A. (2013). Lifelong impact of early self-control: Childhood self-discipline predicts adult quality of life. *American Scientist*, 101(5), 352-359. <https://doi.org/10.1511/2013.104.352>

Montessori, M. (1995). *The absorbent mind* (1st ed.). Henry Holt.

Morgan, J. K., Izard, C. E., & King, K. A. (2010). Construct validity of the Emotion

Matching Task: Preliminary evidence for convergent and criterion validity of a new emotion knowledge measure for young children. *Social development, 19*, 52-70. <https://doi.org/10.1111/j.1467-9507.2008.00529.x>

Morgan, P. L., Farkas, G., Wang, Y., Hillemeier, M. M., Oh, Y., & Maczuga, S. (2019).

Executive function deficits in kindergarten predict repeated academic difficulties across elementary school. *Early Childhood Research Quarterly, 46*, 20-32. <https://doi.org/10.1016/j.ecresq.2018.06.009>

Morgan, P. L., Hillemeier, M. M., Farkas, G., Hammer, C. S., & Maczuga, S. (2015). 24-

month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development, 86*(5), 1351-1370. <https://doi.org/10.1111/cdev.12398>

Moriguchi, Y., Shinohara, I., Todo, N., & Meng, X. (2020). Prosocial behavior is related to

later executive function during early childhood: A longitudinal study. *European Journal of Developmental Psychology, 17*(3), 352-364-364. <https://doi.org/10.1080/17405629.2019.1628737>

Morin, A. J., Meyer, J. P., Creusier, J., & Biétry, F. (2016). Multiple-group analysis of

similarity in latent profile solutions. *Organizational Research Methods, 19*(2), 231-254. <https://doi.org/10.1177/1094428115621148>

Morin, A. J. S., & Marsh, H. W. (2015). Disentangling shape from level effects in person-

centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 39-59. <https://doi.org/10.1080/10705511.2014.919825>

Morkel, V., & McLaughlin, T. (2015). Promoting social and emotional competencies in early childhood: Strategies for teachers. *Kairaranga*, 16(1), 45-51. <https://doi.org/10.54322/kairaranga.v16i1.194>

Morning, A. (2008). Ethnic classification in global perspective: A cross-national survey of the 2000 census round. *Population Research and Policy Review*, 27(2), 239-272. <https://doi.org/10.1007/s11113-007-9062-5>

Morton, S., Peterson, E., Buckley, J., & Murray, S. (2020). *Early self-control development: prevalence, persistence and change in a NZ cohort*.

Morton, S. M., Atatoa Carr, P. E., Grant, C. C., Robinson, E. M., Bandara, D. K., Bird, A., Ivory, V. C., Kingi, T. K. R., Liang, R., & Marks, E. J. (2013). Cohort profile: growing up in New Zealand. *International journal of epidemiology*, 42(1), 65-75.

Morton, S. M. B. (2019). Growing Up in New Zealand: A prebirth cohort study of child wellbeing and development. In F. Sata, H. Fukuoka, & M. Hanson (Eds.), *Pre-emptive Medicine: Public Health Aspects of Developmental Origins of Health and Disease* (pp. 117-129). Springer Singapore. https://doi.org/10.1007/978-981-13-2194-8_7

Morton, S. M. B., Grant, C. C., Berry, S. D., Walker, C. G., Corkin, M., Ly, K., de Castro, T. G., Atatoa Carr, P. E., Bandara, D. K., Mohal, J., Bird, A. L., Underwood, L., &

- Fa'alili-Fidow, J. (2017). *Growing Up in New Zealand: A longitudinal study of New Zealand children and their families. Now We Are Four: Describing the preschool years*. https://cdn.auckland.ac.nz/assets/growingup/research-findings-impact/GUiNZ_Now%20we%20are%20four%20report.pdf
- Morton, S. M. B., Ramke, J., Kinloch, J., Grant, C. C., Carr, P. A., Leeson, H., Lee, A. C. L., & Robinson, E. (2015). Growing Up in New Zealand cohort alignment with all New Zealand births. *Australian and New Zealand Journal of Public Health*, *39*(1), 82-87. <https://doi.org/10.1111/1753-6405.12220>
- Moser, S. (2010). Odds ratio. In *Encyclopedia of Research Design* (Vol. 0, pp. 958-959). SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288>
- Mostow, A. J., Izard, C. E., Fine, S., & Trentacosta, C. J. (2002). Modeling emotional, cognitive, and behavioral predictors of peer acceptance. *Child Development*, *73*(6), 1775-1787. <https://doi.org/10.1111/1467-8624.00505>
- Muehlenhard, C. L., & Peterson, Z. D. (2011). Distinguishing between sex and gender: History, current conceptualisations, and implications. *Sex Roles: A Journal of Research*, *64*(11-12), 791-803. <https://doi.org/10.1007/s11199-011-9932-5>
- Mulkey, S. B., Bearer, C. F., & Molloy, E. J. (2023). Indirect effects of the COVID-19 pandemic on children relate to the child's age and experience. *Pediatric research*, *94*(5), 1586-1587. <https://doi.org/10.1038/s41390-023-02681-4>
- Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: comment on Bauer and Curran (2003). *Psychological methods*, *8*(3), 369-377;

discussion 384-393. <https://doi.org/10.1037/1082-989x.8.3.369>

Muthén, B. (2004). Latent variable analysis. Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345–368). Sage Publications.

Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research, 47*(4), 637-664.

<https://doi.org/10.1177/0049124117701488>

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research, 24*(6), 882-891.

<https://doi.org/10.1111/j.1530-0277.2000.tb02070.x>

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations.

Psychometrika, 54(4), 557-585. <https://doi.org/10.1007/BF02296397>

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.

Naragon-Gainey, K., McMahon, T. P., & Chacko, T. P. (2017). The structure of common emotion regulation strategies: A meta-analytic examination. *Psychological Bulletin, 143*(4), 384-427-427. <https://doi.org/10.1037/bul0000093>

National Academy of Education, Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (2020).

Comparability of large-scale educational assessments: Issues and recommendations.

Nelson, J. M., James, T. D., Chevalier, N., Clark, C. A. C., & Espy, K. A. (2016). Structure, measurement, and development of preschool executive function. In *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research*. (pp. 65-89). American Psychological Association. <https://doi.org/10.1037/14797-004>

Neumann, D., Herbert, S. E., Peterson, E. R., Underwood, L., Morton, S. M. B., & Waldie, K. E. (2019). A longitudinal study of antenatal and perinatal risk factors in early childhood cognition: Evidence from Growing Up in New Zealand. *Early Human Development, 132*, 45-51. <https://doi.org/10.1016/j.earlhumdev.2019.04.001>

Neumann, D., Peterson, E. R., Underwood, L., Morton, S. M. B., & Waldie, K. E. (2021a). The development of cognitive functioning indices in early childhood. *Cognitive Development, 60*. <https://doi.org/10.1016/j.cogdev.2021.101098>

Neumann, D., Peterson, E. R., Underwood, L., Morton, S. M. B., & Waldie, K. E. (2021b). Exploring the factor structure of the nih toolbox cognition battery in a large sample of 8-year-old children in Aotearoa New Zealand. *Journal of the International Neuropsychological Society*. <https://doi.org/10.1017/S1355617720001265>

Neumann, D., Yao, E., Fenaughty, J., Liang, R., Kingi, T. K., Taufa, S., Atatoa Carr, P., & Paine, S.-J. (2023). *Ethnic and gender identity at 12 years old*. Growing Up in

New Zealand. <https://www.growingup.co.nz/growing-up-report/ethnic-and-gender-identity-at-12-years-old>

Niles, A. (2015). Assessment practices within New Zealand early childhood settings. *He Kupu*, 4(2).

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition and Emotion*, 27(7), 1326-1334. <https://doi.org/10.1080/02699931.2013.777660>

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. E. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1-14). Plenum Press.

Nugent, W. R. (2017). Understanding DIF and DTF: Description, methods, and implications for social work research. *Journal of the Society for Social Work and Research*, 8(2), 305-334. <https://doi.org/10.1086/691525>

Nurmi, J.-E. (2012). Students' characteristics and teacher-child relationships in instruction: A meta-analysis. *Educational Research Review*, 7(3), 177-197. <https://doi.org/10.1016/j.edurev.2012.03.001>

Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(6), 782-797. <https://doi.org/10.1080/10705511.2016.1221313>

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo

Simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535-569. <https://doi.org/10.1080/10705510701575396>

O'Sullivan, C. (2012). The epidemiology of serious skin infections in New Zealand children: Comparing the Tairāwhiti region with national trends. *New Zealand Medical Journal.*, 125, 40-54.

O'Toole, S. E., Monks, C. P., & Tsermentseli, S. (2017). Executive function and theory of mind as predictors of aggressive and prosocial behavior and peer acceptance in early childhood. *Social development*, 26(4), 907-920.

<https://doi.org/10.1111/sode.12231>

Olivera-Aguilar, M., & Rikoon, S. H. (2018). Assessing measurement invariance in multiple-group latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 439-452.

<https://doi.org/10.1080/10705511.2017.1408015>

Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2009). Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*, 62(11), 1126-

1137. <https://doi.org/10.1016/j.jclinepi.2009.03.013>

Opoku, A., Ahmed, V., & Akotia, J. (2016). Choosing an appropriate research methodology and method. In V. Ahmed, A. Opoku, & Z. Aziz (Eds.), *Research Methodology in the Built Environment: A Selection of Case Studies* (pp. 32-49).

Ornaghi, V., Brazzelli, E., Grazzani, I., Agliati, A., & Lucarelli, M. (2017). Does training

toddlers in emotion knowledge lead to changes in their prosocial and aggressive behavior toward peers at nursery? *Early Education and Development*, 28(4), 396-414. <https://doi.org/10.1080/10409289.2016.1238674>

Otto, B. (2014). *Language development in early childhood education* (Fourth edition ed.). Pearson.

Owens, R. E., Jr. (2020). *Language development: An introduction* (10th ed.). Pearson Education, Inc.

Padilla, A. M., & Medina, A. (2001). Issues in culturally appropriate assessment. In Lisa A. Suzuki, Joseph G. Ponterotto, & P. J. Melle (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (Vol. 2, pp. 5-28).

Pala, F. C., & Lewis, C. (2021). Do preschoolers grasp the importance of regulating emotional expression? *European Journal of Developmental Psychology*, 18(4), 494-519. <https://doi.org/10.1080/17405629.2020.1814251>

Palmer, S. L. (2004). Homai Te Waiora Ki Ahau: A tool for the measurement of wellbeing among Maori-The evidence of construct validity. *New Zealand Journal of Psychology*, 33, 50.

Parke, R. D. (1994). Progress, paradigms, and unresolved problems: A commentary on recent advances in our understanding of children's emotions. *Merrill-Palmer Quarterly*, 40, 157-169.

Pauls, L. J., & Archibald, L. M. (2016). Executive functions in children with specific language impairment: A meta-analysis. *Journal of speech, language, and*

hearing research: JSLHR, 59(5), 1074-1086. https://doi.org/10.1044/2016_jslhr-l-15-0174

Paulus, M., Licata, M., Kristen, S., Thoermer, C., Woodward, A., & Sodian, B. (2015). Social understanding and self-regulation predict pre-schoolers' sharing with friends and disliked peers: A longitudinal study. *International Journal of Behavioral Development*, 39(1), 53-64.

<https://doi.org/10.1177/0165025414537923>

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews. Technical Report.* (Collaborative for Academic, Social, and Emotional Learning (NJ1), Issue. <https://files.eric.ed.gov/fulltext/ED505370.pdf>

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological methods*, 23(2), 208-225.

<https://doi.org/10.1037/met0000126>

Perez, A. D., & Hirschman, C. (2009). The changing racial and ethnic composition of the US Population: Emerging American identities. *Population and Development Review*, 35(1), 1-51. <https://doi.org/10.1111/j.1728-4457.2009.00260.x>

Perren, S., & Alsaker, F. D. (2006). Social behavior and peer relationships of victims, bully-victims, and bullies in kindergarten. *Journal of Child Psychology and Psychiatry*, 47(1), 45-57. <https://doi.org/10.1111/j.1469-7610.2005.01445.x>

- Perren, S., & Alsaker, F. D. (2009). Depressive symptoms from kindergarten to early school age: Longitudinal associations with social skills deficits and peer victimization. *Child and Adolescent Psychiatry and Mental Health, 3*(1), 28.
<https://doi.org/10.1186/1753-2000-3-28>
- Perren, S., Forrester-Knauss, C., & Alsaker, F. D. (2012). Self- and other-oriented social skills: Differential associations with children's mental health and bullying roles. *Journal for Educational Research Online, 4*(1), 99-123.
- Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 616-639.
<https://doi.org/10.1080/10705511.2013.824780>
- Phillips, D. A., & Shonkoff, J. P. (2000). *From neurons to neighborhoods: The science of early child development*. National Academy Press.
- Phillips, G., McNaughton, S., & MacDonald, S. (2004). Managing the mismatch: Enhancing early literacy progress for children with diverse language and cultural identities in mainstream urban schools in New Zealand. *Journal of educational psychology, 96*(2), 309-323. <https://doi.org/10.1037/0022-0663.96.2.309>
- Phinney, J. S. (1990). Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin, 108*(3), 499-514. <https://doi.org/10.1037/0033-2909.108.3.499>
- Piaget, J. (1948). *The moral judgment of the child*. Free Press.

- Piaget, J. (1983). Piaget's Theory. In P. H. Mussen & W. Kessen (Eds.), *Handbook of Child Psychology* (Vol. I, pp. 41-102). John Wiley.
- Pianta, R. C., & Caldwell, C. B. (1990). Stability of externalising symptoms from kindergarten to first grade and factors related to instability. *Development and Psychopathology*, 2, 247-258. <https://doi.org/10.1017/S0954579400000754>
- Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review*, 33(3), 444-458. <https://doi.org/10.1080/02796015.2004.12086261>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329-347. <https://doi.org/10.1017/S0140525X12001495>
- Plaisted, J. R., Wilkening, G. N., Gustavson, J. L., & Golden, C. J. (1983). The Luria-Nebraska Neuropsychological Battery- children's revision: Theory and current research findings. *Journal of Clinical Child Psychology*, 12(1), 13. https://doi.org/10.1207/s15374424jccp1201_4
- Plutchik, R. (1980). *Emotion, a psychoevolutionary theory*. New York: Harper and Row.
- Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution*. American Psychological Association.
- Pomeroy, D. (2020). Educational equity policy as human taxonomy: Who do we compare and why does it matter? *Critical Studies in Education*, 61(3), 329-344. <https://doi.org/10.1080/17508487.2018.1440615>

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1*(2), 127-152.

<https://doi.org/10.1080/17405620344000022>

Pons, F., Lawson, J., Harris, P. L., & De Rosnay, M. (2003). Individual differences in children's emotion understanding: Effects of age and language. *Scandinavian Journal of Psychology, 44*(4), 347-353. <https://doi.org/10.1111/1467-9450.00354>

Prager, E. O., Ernst, J. R., Mazzocco, M. M. M., & Carlson, S. M. (2023). Executive function and mathematics in preschool children: Training and transfer effects. *Journal of Experimental Child Psychology, 232*, 105663.

<https://doi.org/10.1016/j.jecp.2023.105663>

Pribram, K. H. (1973). The primate frontal cortex - executive of the brain. In K. H. Pribram & A. R. Luria (Eds.), *Psychophysiology of the Frontal Lobes* (pp. 293-314). Academic Press. <https://doi.org/10.1016/B978-0-12-564340-5.50019-6>

Prosen, S., & Smrtnik Vitulić, H. (2018). Children's emotional expression in the preschool context. *Early Child Development and Care, 188*(12), 1673-1681-1681.

<https://doi.org/10.1080/03004430.2016.1278367>

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>

Putnick, D. L., Trinh, M. H., Sundaram, R., Bell, E. M., Ghassabian, A., Robinson, S. L., &

Yeung, E. (2023). Displacement of peer play by screen time: associations with toddler development. *Pediatric research*, *93*(5), 1425-1431.

<https://doi.org/10.1038/s41390-022-02261-y>

Qi, C. H., & Kaiser, A. P. (2004). Problem behaviors of low-income children with language delays: an observation study. *Journal of Speech, Language, and Hearing Research*, *47*(3), 595-609. [https://doi.org/10.1044/1092-4388\(2004/046\)](https://doi.org/10.1044/1092-4388(2004/046))

Quintana, S. M. (2007). Racial and ethnic identity: Developmental perspectives and research. *Journal of Counseling Psychology*, *54*(3), 259-270.

<https://doi.org/10.1037/0022-0167.54.3.259>

Rameka, L., Soutar, B., Paki, V., & Clayton, L. (2021). Taunaki puna reo: Kaiako considerations of mana and kaitiakitanga. *Waikato Journal of Education*, *26*.

<https://doi.org/10.15663/wje.v26i2.867>

Redmond, S. M., & Rice, M. L. (1998). The socioemotional behaviors of children with SLI: Social adaptation or social deviance? *Journal of Speech, Language, and Hearing Research*, *41*, 688-700. <https://doi.org/10.1044/jslhr.4103.688>

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: Results of a 24-year longitudinal study. *Social Psychiatry and Psychiatric Epidemiology*, *46*(12), 1233-1241.

<https://doi.org/10.1007/s00127-010-0297-9>

Reese, E., Peterson, E. R., Waldie, K., Schmidt, J., Bandara, D., Carr, P. A., Grant, C., Pryor,

- J., & Morton, S. M. B. (2016). High Hopes? Educational, Socioeconomic, and Ethnic Differences in Parents' Aspirations for their Unborn Children. *Journal of Child and Family Studies, 25*(12), 3657. <https://doi.org/10.1007/s10826-016-0521-7>
- Reese, E., & Read, S. (2000). Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and sentences. *Journal of Child Language, 27*(2), 255-266. <https://doi.org/10.1017/S0305000900004098>
- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *Am Psychol, 74*(4), 445-458. <https://doi.org/10.1037/amp0000356>
- Rhoades, B. L., Greenberg, M. T., & Domitrovich, C. E. (2009). The contribution of inhibitory control to preschoolers' social-emotional competence. *Journal of Applied Developmental Psychology, 30*(3), 310-320. <https://doi.org/10.1016/j.appdev.2008.12.012>
- Richards, N., Reith, D., Stitely, M., & Smith, A. (2019). The Before School Check (B4SC): reporting outcomes and referral rates for all New Zealand children. *New Zealand Medical Journal, 132*(1496), 9-19.
- Riglin, L., Dennison, C., Martin, J., Tseliou, F., Armitage, J. M., Shakeshaft, A., Heron, J., Tilling, K., Thapar, A., & Collishaw, S. (2024). Emotional problems across development: Examining measurement invariance across childhood,

adolescence and early adulthood. *European child & adolescent psychiatry*.

<https://doi.org/10.1007/s00787-024-02461-3>

Rimm-Kaufman, S. E., & Pianta, R. C. (2000). An ecological perspective on the transition to kindergarten: A theoretical framework to guide empirical research. *Journal of Applied Developmental Psychology, 21*, 491-511.

[https://doi.org/10.1016/S0193-3973\(00\)00051-4](https://doi.org/10.1016/S0193-3973(00)00051-4)

Robitzsch, A., & Lüdtke, O. (2023). Why full, partial, or approximate measurement invariance are not a prerequisite for meaningful and valid group comparisons. *Structural Equation Modeling: A Multidisciplinary Journal, 30*(6), 859-870.

<https://doi.org/10.1080/10705511.2023.2191292>

Rochford, T. (2004). Whare Tapa Whā: A Māori model of a unified theory of health. *Journal of Primary Prevention, 25*(1), 41-57.

<https://doi.org/10.1023/B:JOPP.0000039938.39574.9e>

Rodriguez, V. J., La Barrie, D. L., Zegarac, M. C., & Shaffer, A. (2023). A systematic review of parenting scales measurement invariance/ equivalence of by race and ethnicity: Recommendations for inclusive parenting research. *Assessment, 30*(1), 22-36. <https://doi.org/10.1177/10731911211038630>

Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social development, 6*(1), 111-135. <https://doi.org/10.1111/j.1467-9507.1997.tb00097.x>

Rothman, S. (2005). *Report on adapted PPVT-III and Who Am I? Growing Up in*

Australia: The Longitudinal Study of Australian Children.

<https://hdl.handle.net/10620/17497>

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762-1774. <https://doi.org/10.1111/j.1467-8624.2012.01805.x>

Roy-Charland, A., Perron, M., Beaudry, O., & Eady, K. (2014). Confusion of fear and surprise: A test of the perceptual-attentional limitation hypothesis with eye movement monitoring. *Cognition and Emotion, 28*(7), 1214-1222.

<https://doi.org/10.1080/02699931.2013.878687>

Rubie-Davies, C., Asil, M., & Teo, T. (2016). Assessing measurement invariance of the student personal perception of classroom climate across different ethnic groups. *Journal of Psychoeducational Assessment, 34*(5), 442-460.

<https://doi.org/10.1177/0734282915612689>

Rubin, K. H., Bukowski, W. M., & Parker, J. G. (2006). Peer interactions, relationships, and groups. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 6th ed.* (pp. 571-645). John Wiley & Sons, Inc.

Rubin, K. H., Burgess, K. B., Dwyer, K. M., & Hastings, P. D. (2003). Predicting preschoolers' externalising behaviours from toddler temperament, conflict, and maternal negativity. *Developmental psychology, 39*(1), 164-176.

<https://doi.org/10.1037/0012-1649.39.1.164>

- Rubin, K. H., & Clark, M. L. (1983). Preschool teachers' ratings of behavioral problems: Observational, sociometric, and social-cognitive correlates. *Journal of Abnormal Child Psychology*, *11*(2), 273-285. <https://doi.org/10.1007/BF00912091>
- Rubin, K. H., Coplan, R., Chen, X., Bowker, J., & McDonald, K. L. (2011). Peer relationships in childhood. In *Developmental science: An advanced textbook, 6th ed.* (pp. 519-570). Psychology Press.
- Rubin, K. H., & Rose-Krasnor, L. (1992). Interpersonal problem solving and social competence in children. In *Handbook of social development* (pp. 283-323). Springer.
- Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 6th ed.* (pp. 858-932). John Wiley & Sons, Inc.
- Rudas, T. (1998). *Odds ratios in the analysis of contingency tables*. Sage Publications.
- Russell, J., Grant, C. C., Morton, S., Denny, S., & Paine, S.-J. (2022). Prevalence and predictors of developmental health difficulties within New Zealand preschool-aged children: A latent profile analysis. *Journal of the Royal Society of New Zealand*, 1-28. <https://doi.org/10.1080/03036758.2022.2083188>
- Rusten, N. F., Peterson, E. R., Underwood, L., Verbiest, M. E. A., Waldie, K. E., Berry, S., Carr, P. A., Grant, C., Pryor, J., Nicolson, J., & Morton, S. M. B. (2019). Psychological distress among resident and nonresident fathers: Findings from

- New Zealand's Who Are Today's Dads? Project. *Journal of Family Issues*, 40(3), 293-314. <https://doi.org/10.1177/0192513X18808827>
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31-57. <https://doi.org/10.1177/0013164413498257>
- Rutkowski, R. E., Salemi, J. L., Tanner, J. P., Matas, J. L., & Kirby, R. S. (2017). Assessing the impact of different race-bridging algorithms on the reported rate of birth defects. *Journal of Registry Management*, 44(4), 146-156.
- Rydell, A.-M., Berlin, L., & Bohlin, G. (2003). Emotionality, emotion regulation, and adaptation among 5- to 8-year-old children. *Emotion*, 3(1), 30-47. <https://doi.org/10.1037/1528-3542.3.1.30>
- Rydell, A. M., Hagekull, B., & Bohlin, G. (1997). Measurement of two social competence aspects in middle childhood. *Developmental psychology*, 33(5), 824-833. <https://doi.org/10.1037//0012-1649.33.5.824>
- Saarni, C. (1990). Emotional competence. In Ross Thompson (Ed.), *Nebraska symposium: Socioemotional development* (pp. 115-161). University of Nebraska Press.
- Saarni, C. (1999a). *The development of emotional competence* Guilford Press.
- Saarni, C. (1999b). *A skill-based model of emotional competence: A developmental perspective*.
- Saarni, C. (2007). The development of emotional competence: Pathways for helping

children to become emotionally intelligent. In R. Bar-On, J. G. Maree, & M. J. Elias (Eds.), *Educating people to be emotionally intelligent*. (pp. 15-35). Praeger Publishers/Greenwood Publishing Group.

Saarni, C. (2011). Emotional Development in Childhood. *Encyclopedia on Early Childhood Development*.

Saarni, C., Campos, J. J., Camras, L. A., & Witherington, D. (2006). Emotional development: action, communication, and understanding. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 6th ed.* (pp. 226-299). John Wiley & Sons, Inc.

Saarni, C., Campos, J. J., Camras, L. A., & Witherington, D. (2008). Principles of emotion and emotional competence. In William Damon, Richard M. Lerner (Ed.), *Child and adolescent development: An advanced course* (Vol. 23, pp. 361-405).

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17*(1), 74-81.

<https://doi.org/10.1111/j.1467-9280.2005.01667.x>

Sandilos, L. E., Whittaker, J. V., Vitiello, V. E., & Kinzie, M. B. (2019). Preschoolers' school readiness profiles and the teacher-child relationship: A latent transition approach. *Journal of Applied Developmental Psychology, 62*, 185-198.

<https://doi.org/10.1016/j.appdev.2019.02.010>

- Sanne, B., Heiervang, E., Stormark, K. M., & Torsheim, T. (2009). The Strengths and Difficulties Questionnaire in the Bergen child study: A conceptually and methodically motivated structural analysis. *Psychological assessment, 21*(3), 352-364. <https://doi.org/10.1037/a0016317>
- Sanson, A., Smart, D., Prior, M., & Oberklaid, F. (1993). Precursors of hyperactivity and aggression. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*, 1207-1216. <https://doi.org/10.1097/00004583-199311000-00014>
- Santiago, P. H. R., Manzini Macedo, D., Haag, D., Roberts, R., Smithers, L., Hedges, J., & Jamieson, L. (2021). Exploratory graph analysis of the Strengths and Difficulties Questionnaire in the longitudinal study of Australian children. *Assessment, 29*(8), 1622–1640. <https://doi.org/10.1177/107319112111024338>
- Sarason, B. R. (1981). The dimensions of social competence: Contributions from a variety of research areas. In J. D. Wine & M. D. Smye (Eds.), *Social competence* (pp. 100-122). Guilford Press.
- Sbordone, R. J. (1996). Ecological validity: Some critical issues for the neuropsychologist. In *Ecological validity of neuropsychological testing*. (pp. 15-41). Gr Press/St Lucie Press, Inc.
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210-222. <https://doi.org/10.1016/j.hrmr.2008.03.003>
- Schutte, N. S., Malouff, J. M., Thorsteinsson, E. B., Bhullar, N., & Rooke, S. E. (2007). A

meta-analytic investigation of the relationship between emotional intelligence and health. *Personality and Individual Differences*, 42(6), 921-933.

<https://doi.org/10.1016/j.paid.2006.09.003>

Scott, N. (2014). A Māori cultural reluctance to present for care, or a systems and quality failure? How we pose the issue, informs our solutions. *New Zealand Medical Journal*.

Sebanic, A. M. (2003). The friendship features of preschool children: Links with prosocial behavior and aggression. *Social development*, 12(2), 249-268.

<https://doi.org/10.1111/1467-9507.00232>

Serrat, E., Amadó, A., Rostan, C., Caparrós, B., & Sidera, F. (2020). Identifying emotional expressions: Children's reasoning about pretend emotions of sadness and anger. *Frontiers in Psychology*, 11, 602385. <https://doi.org/10.3389/fpsyg.2020.602385>

Sette, S., Bassett, H. H., Baumgartner, E., & Denham, S. A. (2015). Structure and Validity of Affect Knowledge Test (AKT) in a Sample of Italian Preschoolers. *Journal of Genetic Psychology*, 176(5), 330-347.

<https://doi.org/10.1080/00221325.2015.1075466>

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511526817>

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379-423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>

Sharp, P. (2001). *Nurturing emotional literacy: A practical guide for teachers, parents*

and those in the caring professions. *Health Education*, 101(6), 292-294.

<https://doi.org/10.1108/he.2001.101.6.292.1>

Shields, A., Dickstein, S., Seifer, R., Giusti, L., Magee, K. D., & Spritz, B. (2001). Emotional competence and early school adjustment: A study of preschoolers at risk. *Early Education and Development*, 12(1), 73-96.

https://doi.org/10.1207/s15566935eed1201_5

Shoshani, A., Aharon-Dvir, O., Hain, D., & Yaffe, A. (2021). Situational determinants of young children's gratitude: The effects of perceived intentionality and the value of the benefit on gratitude and prosocial behavior. *Journal of Personality and Social Psychology*, 121(4), 914-932. <https://doi.org/10.1037/pspp0000384>

Shugar, G. W. (1986). The acquisition of communicative competence, R. L. Schiefelbusch & J. Pickar (Eds.). Language Intervention Series Volume VIII, Baltimore, MD.: University Park Press, 1984. Pp. x + 550. *Applied Psycholinguistics*, 7(1), 96-100.

<https://doi.org/10.1017/S0142716400007244>

Silberman, E. K., & Weingartner, H. (1986). Hemispheric lateralization of functions related to emotion. *Brain and Cognition*, 5, 322-353.

[https://doi.org/10.1016/0278-2626\(86\)90035-7](https://doi.org/10.1016/0278-2626(86)90035-7)

Silva, C., Sousa-Gomes, V., Fávero, M., Oliveira-Lopes, S., Merendeiro, C. S., Oliveira, J., & Moreira, D. (2022). Assessment of preschool-age executive functions: A systematic review. *Clinical Psychology Psychotherapy*, 29(4), 1374-1391.

<https://doi.org/10.1002/cpp.2718>

- Simon, H. A. (1975). The functional equivalence of problem-solving skills. *Cognitive Psychology*, 7(2), 268-288. [https://doi.org/10.1016/0010-0285\(75\)90012-2](https://doi.org/10.1016/0010-0285(75)90012-2)
- Simsek, E., Jones, I., Hunter, J., & Xenidou-Dervou, I. (2021). Mathematical equivalence assessment: Measurement invariance across six countries. *Studies in Educational Evaluation*, 70, N.PAG-N.PAG. <https://doi.org/10.1016/j.stueduc.2021.101046>
- Skalická, V., Stenseng, F., & Wichstrøm, L. (2015). Reciprocal relations between student-teacher conflict, children's social skills and externalizing behavior: A three-wave longitudinal study from preschool to third grade. *International Journal of Behavioral Development*, 39(5), 413-425. <https://doi.org/10.1177/0165025415584187>
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *Neuroimage*, 25(1), 76-89. <https://doi.org/10.1016/j.neuroimage.2004.11.006>
- Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, 66(4), 563-575. <https://doi.org/10.1007/BF02296196>
- Slykerman, R. F., Neumann, D., Underwood, L., Hobbs, M., & Waldie, K. E. (2023). Age at first exposure to antibiotics and neurodevelopmental outcomes in childhood. *Psychopharmacology*, 240(5), 1143-1150. <https://doi.org/10.1007/s00213-023-06351-5>
- Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for

- field-based research. *Early Childhood Research Quarterly*, 22(2), 173-187.
<https://doi.org/10.1016/j.ecresq.2007.01.002>
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283(5408), 1657-1661. <https://doi.org/10.1126/science.283.5408.1657>
- Smith, G. (2000). Maori education: Revolution and transformative action. *Canadian Journal of Native Education*, 24(1), 57-72.
- Smith, J., & Abell, C. (2016). Introduction: Emotional expression. In C. Abell & J. Smith (Eds.), *The Expression of Emotion: Philosophical, Psychological and Legal Perspectives* (pp. 1-22). Cambridge University Press.
<https://doi.org/10.1017/CBO9781316275672.001>
- Smith, T. E., Polloway, E. A., Patton, J. R., Dowdy, C. A., & Doughty, T. T. (2014). *Teaching students with special needs in inclusive settings* (6 ed.). Pearson
- Smolak, L. (1982). Cognitive precursors of receptive vs. expressive language. *Journal of Child Language*, 9(1), 13-22-22. <https://doi.org/10.1017/S0305000900003573>
- Snyder, J., Horsch, E., & Childs, J. (1997). Peer relationships of young children: Affiliative choices and the shaping of aggressive behavior. *Journal of Clinical Child Psychology*, 26(2), 145. https://doi.org/10.1207/s15374424jccp2602_3
- Soland, J., McGinty, A., Gray, A., Solari, E. J., Herring, W., & Xu, R. (2022). Early literacy, equity, and test score comparability during the pandemic. *Educational Assessment*, 27(2), 98-114. <https://doi.org/10.1080/10627197.2022.2087622>
- Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis:

A review and “how to” guide of its application within vocational behavior research. *Journal of Vocational Behavior*, 120, 103445.

<https://doi.org/10.1016/j.jvb.2020.103445>

Statistics New Zealand. (2004). *Report of the review of the measurement of ethnicity*.

<https://statsnz.contentdm.oclc.org/digital/collection/p20045coll1/id/1460/>

Statistics New Zealand. (2005). Statistical standard for ethnicity. In Statistics New Zealand (Ed.). Wellington: Statistics New Zealand.

Statistics New Zealand. (2005, updated 2023). Ethnicity New Zealand Standard Classification 2005V2.1.0. In: Statistics New Zealand.

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement enough? *Methodology*, 9(1), 1-12. <https://doi.org/10.1027/1614-2241/a000049>

Sterba, S. K. (2013). Understanding linkages among mixture models. *Multivariate Behavioral Research*, 48(6), 775-815.

<https://doi.org/10.1080/00273171.2013.827564>

Stockman, I. J. (2000). The new Peabody Picture Vocabulary Test-III: An illusion of unbiased assessment? *Language, Speech and Hearing Services in Schools*, 31(4), 340-353. <https://doi.org/10.1044/0161-1461.3104.340>

Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-year-olds: A review. *Clinical Child &*

Family Psychology Review, 13(3), 254-274. <https://doi.org/10.1007/s10567-010-0071-2>

Strayer, F. F., & Santos, A. J. (1996). Affiliative structures in preschool peer groups. *Social development*, 5, 117-130. <https://doi.org/10.1111/j.1467-9507.1996.tb00075.x>

Strayer, J., & Roberts, W. (2004). Empathy and observed anger and aggression in five-year-olds. *Social development*, 13(1), 1-13-13. <https://doi.org/10.1111/j.1467-9507.2004.00254.x>

Stremel, K., et al. (1994). *Communication Interactions, Receptive Communication, Expressive Communication, and Communication Intervention*.

Stuss, D. T., Shallice, T., Alexander, M. P., & Picton, T. W. (1995). A multidisciplinary approach to anterior attentional functions. In *Structure and functions of the human prefrontal cortex*. (pp. 191-211). New York Academy of Sciences.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857), 1285-1293. <https://doi.org/10.1126/science.3287615>

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *The American psychologist*, 47(4), 522-532. <https://doi.org/10.1037//0003-066x.47.4.522>

the Department of the Prime Minister and Cabinet. (2019). *The Child and Youth Wellbeing Strategy*, the Department of the Prime Minister and Cabinet (DPMC),

The Education Review Office. (2023). *Education for all our children: Embracing diverse ethnicities*.

- Thelen, E., & Smith, L. B. (2006). Dynamic Systems Theories. In *Handbook of child psychology: Theoretical models of human development, Vol. 1, 6th ed.* (pp. 258-312). John Wiley & Sons, Inc.
- Thomas, A., Cairney, S., Gunthorpe, W., Paradies, Y., & Sayers, S. (2010). Strong souls: Development and validation of a culturally appropriate tool for assessment of social and emotional well-being in indigenous youth. *Australian & New Zealand Journal of Psychiatry, 44*(1), 40-48. <https://doi.org/10.3109/00048670903393589>
- Thompson, J. M. D., Ayrey, S. L., Slykerman, R. F., Stone, P. R., & Fowler, P. V. (2021). Behavioural outcomes in children with an orofacial cleft in a national study in New Zealand. *The Cleft Palate-Craniofacial Journal, 58*(8), 1032-1039. <https://doi.org/10.1177/1055665620972862>
- Thompson, R. A., & Goodvin, R. (2007). Taming the tempest in the teapot: Emotion regulation in toddlers. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations.* (pp. 320-341). The Guilford Press.
- Thompson, R. A., & Raikes, H. A. (2007). The social and emotional foundations of school readiness. In D. F. Perry, R. K. Kaufmann, & J. Knitzer (Eds.), *Social and emotional health in early childhood: Building bridges between services and systems.* (pp. 13-35). Paul H. Brookes Publishing Co.
- Thomson, K. C., Guhn, M., Richardson, C. G., Ark, T. K., & Shoveller, J. (2017). Profiles of children's social-emotional health at school entry and associated income,

gender and language inequalities: a cross-sectional population-based study in British Columbia, Canada. *BMJ Open*, 7(7), e015353.

<https://doi.org/10.1136/bmjopen-2016-015353>

Thomson, K. C., Richardson, C. G., Gadermann, A. M., Emerson, S. D., Shoveller, J., & Guhn, M. (2019). Association of childhood social-emotional functioning profiles at school entry with early-onset mental health conditions. *JAMA Network Open*, 2(1), e186694-e186694. <https://doi.org/10.1001/jamanetworkopen.2018.6694>

Thorell, L. B., Bohlin, G., & Rydell, A.-M. (2004). Two types of inhibitory control: Predictive relations to social functioning. *International Journal of Behavioral Development*, 28(3), 193-203. <https://doi.org/10.1080/01650250344000389>

Tierney, R. D. (2017). Fairness in educational assessment. In M. A. Peters (Ed.), *Encyclopedia of Educational Philosophy and Theory* (pp. 793-798). Springer Singapore. https://doi.org/10.1007/978-981-287-588-4_400

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, 11, 61-82. <https://doi.org/10.1515/cogl.2001.012>

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology & Sociobiology*, 11, 375-424. [https://doi.org/10.1016/0162-3095\(90\)90017-Z](https://doi.org/10.1016/0162-3095(90)90017-Z)

Torkildsen, J. v. K., Hansen, H. F., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*, 108(2), 73-88.

<https://doi.org/10.1016/j.bandl.2008.09.005>

Trentacosta, C. J., & Fine, S. E. (2010). Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Social development, 19*(1), 1-29. <https://doi.org/10.1111/j.1467-9507.2009.00543.x>

Trentacosta, C. J., Izard, C. E., Mostow, A. J., & Fine, S. E. (2006). Children's emotional competence and attentional competence in early elementary school. *School Psychology Quarterly, 21*(2), 148-170.

<https://doi.org/10.1521/scpq.2006.21.2.148>

Tsaousis, I., Sideridis, G. D., & AlGhamdi, H. M. (2020). Measurement invariance and Differential Item Functioning across gender within a latent class analysis framework: Evidence from a high-stakes test for university admission in Saudi Arabia. *Frontiers in Psychology, 11*. <https://doi.org/10.3389/fpsyg.2020.00622>

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10. <https://doi.org/10.1007/BF02291170>

Turk, D. J., Heatherton, T. F., Kelley, W. M., Funnell, M. G., Gazzaniga, M. S., & Macrae, C. N. (2002). Mike or me? Self-recognition in a split-brain patient. *Nature Neuroscience 5*(9), 841-842. <https://doi.org/10.1038/nn907>

Uddin, L. Q. (2021). Cognitive and behavioural flexibility: Neural mechanisms and clinical considerations. *Nature Reviews Neuroscience, 22*(3), 167-179. <https://doi.org/10.1038/s41583-021-00428-w>

Ulloa, M. L. (2011). *Teaching to care: emotionally intelligent teachers support preschool*

children's emotional competence: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Psychology at Massey University, Wellington, New Zealand, October 2011.

Unger, R. K. (1979). Toward a redefinition of sex and gender. *American Psychologist*, *34*(11), 1085-1094. <https://doi.org/10.1037/0003-066X.34.11.1085>

Unger, R. K., & Crawford, M. (1993). Sex and gender—The troubled relationship between terms and concepts. *Psychological Science*, *4*(2), 122-124. <https://doi.org/10.1111/j.1467-9280.1993.tb00473.x>

Vallotton, C., & Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly*, *26*(2), 169-181. <https://doi.org/10.1016/j.ecresq.2010.09.002>

van Anders, S. M. (2015). Beyond sexual orientation: Integrating gender/sex and diverse sexualities via sexual configurations theory. *Archives of Sexual Behavior*, *44*(5), 1177-1213. <https://doi.org/10.1007/s10508-015-0490-8>

Van de Looij-Jansen, P. M., Goedhart, A. W., de Wilde, E. J., & Treffers, P. D. (2011). Confirmatory factor analysis and factorial invariance analysis of the adolescent self-report Strengths and Difficulties Questionnaire: How important are method effects and minor factors? *The British journal of clinical psychology*, *50*(2), 127-144. <https://doi.org/10.1348/014466510x498174>

Van De Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthen, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the

- novel possibility of approximate measurement invariance. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00770>
- Van Hulle, C. A., Goldsmith, H. H., & Lemery, K. S. (2004). Genetic, environmental, and gender effects on individual differences in toddler expressive language. *Journal of Speech, Language & Hearing Research*, 47(4), 904-912.
[https://doi.org/10.1044/1092-4388\(2004/067\)](https://doi.org/10.1044/1092-4388(2004/067))
- Van Leeuwen, K., Meerschaert, T., Bosmans, G., De Medts, L., & Braet, C. (2006). The strengths and difficulties questionnaire in a community sample of young children in Flanders. *European Journal of Psychological Assessment*, 22(3), 189-197. <https://doi.org/10.1027/1015-5759.22.3.189>
- Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology & Psychiatry*, 49(12), 1304-1312.
<https://doi.org/10.1111/j.1469-7610.2008.01942.x>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
<https://doi.org/10.1177/109442810031002>
- Veldman, S. L. C., Hammersley, M. L., Howard, S. J., Stanley, R. M., Okely, A. D., & Jones, R. A. (2023). Associations of gross motor skills with self-regulation and executive function in preschool-aged children. *Australasian Journal of Early Childhood*,

48(3), 234-246. <https://doi.org/10.1177/18369391231175524>

Vitaro, F., Gagnon, C., & Tremblay, R. E. (1990). Predicting Stable Peer Rejection From Kindergarten to Grade One. *Journal of Clinical Child Psychology, 19*(3), 257-264. https://doi.org/10.1207/s15374424jccp1903_9

Vugteveen, J., de Bildt, A., Serra, M., Timmerman, M. E., & de Wolff, M. S. (2020). Psychometric properties of the Dutch Strengths and Difficulties Questionnaire (SDQ) in adolescent community and clinical populations. *Assessment, 27*(7), 1476-1489-1489. <https://doi.org/10.1177/1073191118804082>

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. Harvard University Press.

Walker, S., & Berthelsen, D. (2017). Gender differences in early literacy and mathematics achievement and self-regulatory behaviours in the first year of school: An Australian study. *Australasian Journal of Early Childhood, 42*(1), 70-78. <https://doi.org/10.23965/ajec.42.1.08>

Wallin, G., Chen, Y., & Moustaki, I. (2024). DIF analysis with unknown groups and anchor items. *Psychometrika, 89*(1), 267-295. <https://doi.org/10.1007/s11336-024-09948-7>

Walrath, C. M., Petras, H., Mandell, D. S., Stephens, R. L., Holden, E. W., & Leaf, P. J. (2004). Gender differences in patterns of risk factors among children receiving mental health services: Latent class analyses. *The journal of behavioral health services & research, 31*(3), 297-311. <https://doi.org/10.1007/bf02287292>

- Walsh, M. C., Joyce, S., Maloney, T., Vaithianathan, R., & Walsh, M. (2019). Adverse childhood experiences and school readiness outcomes: Results from the Growing Up in New Zealand study. *New Zealand Medical Journal*, *132*(1493), 15-24.
- Wang, J., & Wang, X. (2019). *Structural equation modeling: Applications using Mplus* (2 ed.). John Wiley & Sons.
- Wang, Y., Kim, E., & Yi, Z. (2022). Robustness of latent profile analysis to measurement noninvariance between profiles. *Educational and Psychological Measurement*, *82*(1), 5-28. <https://doi.org/10.1177/0013164421997896>
- Wasserman, T., & Wasserman, L. D. (2013). Toward an integrated model of executive functioning in children. *Applied neuropsychology. Child*, *2*(2), 88-96. <https://doi.org/10.1080/21622965.2013.748394>
- Waters, E., & Sroufe, L. A. (1983). Social competence as a developmental construct. *Developmental Review*, *3*(1), 79-97. [https://doi.org/10.1016/0273-2297\(83\)90010-2](https://doi.org/10.1016/0273-2297(83)90010-2)
- Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of educational psychology*, *102*(1), 43-53. <https://doi.org/10.1037/a0016738>
- Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, *4*(3),

199-230. <https://doi.org/10.1080/87565648809540405>

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children.

Developmental Neuropsychology, 7(2), 131-149.

<https://doi.org/10.1080/87565649109540483>

Wen Xia, R., Phang Wai, L., & Mazlina Che, M. (2023). A case study on kindergarten children's social behaviors during peer play. *Journal of Research, Policy &*

Practice of Teachers & Teacher Education, 13(2), 40-58.

<https://doi.org/10.37134/jrpptte.vol13.2.3.2023>

Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. . Harvard University Press.

West, R. L., & Turner, L. H. (2017). *Introducing communication theory: Analysis and application* (6th ed.). McGraw-Hill Education.

Whitters, H. G. (2019). Attainment and executive functioning. In *Attainment and executive functioning in the early years: research for inclusive practice and lifelong learning* (1st ed.). Routledge.

Wichstrøm, L., Berg-Nielsen, T. S., Angold, A., Egger, H. L., Solheim, E., & Sveen, T. H.

(2012). Prevalence of psychiatric disorders in preschoolers. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 53(6), 695-705.

<https://doi.org/10.1111/j.1469-7610.2011.02514.x>

Wickrama, K. A. S., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher-order growth*

curves and mixture modeling with Mplus: A practical guide. Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9781315642741>

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental psychology*, *39*(1), 114-128.
<https://doi.org/10.1037//0012-1649.39.1.114>

Wigelsworth, M., Humphrey, N., Kalambouka, A., & Lendrum, A. (2010). A review of key issues in the measurement of children's social and emotional skills. 173-186.
<https://doi.org/10.1080/02667361003768526>

Wiig, E. H., Secord, W. A., & Semel, E. (2006). *Clinical evaluation of language fundamentals-preschool: Australian standardised edition* (2nd ed.). Harcourt Assessment.

Wilcox, R. (2017). *Modern statistics for the social and behavioral sciences: A practical introduction* (2nd ed.). Chapman and Hall/CRC.
<https://doi.org/10.1201/9781315154480>

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological psychiatry*, *57*(11), 1336-1346.
<https://doi.org/10.1016/j.biopsych.2005.02.006>

Williams, K. T. (2007). *Expressive vocabulary test, second edition (EVT-2)*. Pearson Education.

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood:

- longitudinal measurement invariance and developmental change. *Psychological assessment*, 24(2), 418-431. <https://doi.org/10.1037/a0025779>
- Wilson, C. J., Bowden, S. C., Vannier, L.-C., Byrne, L. K., & Weiss, L. G. (2023). Measurement invariance of the Wechsler Intelligence Scale for children, fifth edition, in Australian and New Zealand and U.S. standardization samples. *Psychological assessment*, 35(6), 510-521. <https://doi.org/10.1037/pas0001221>
- Wood, C., & Schatschneider, C. (2019). Item bias: Predictors of accuracy on Peabody Picture Vocabulary Test–Fourth edition items for Spanish-English–speaking children. *Journal of Speech, Language, and Hearing Research*, 62(5), 1392-1402. https://doi.org/10.1044/2018_jslhr-l-18-0145
- Woody, S. R., & Teachman, B. A. (2000). Intersection of disgust and fear: Normative and pathological views. *Clinical Psychology: Science and Practice*, 7(3), 291-311. <https://doi.org/10.1093/clipsy.7.3.291>
- Wu, A. D., Liu, Y., Stone, J. E., Zou, D., & Zumbo, B. D. (2017). Is difference in measurement outcome between groups differential responding, bias or disparity? A methodology for detecting bias and impact from an attributional stance. *Frontiers in Education*, 2. <https://doi.org/10.3389/feduc.2017.00039>
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014-1045. <https://doi.org/10.1007/s11336-016-9506-0>
- Yao, E. S., Meissel, K., Bullen, P., Atatoa Carr, P., Clark, T., & Morton, S. (2021). Classifying

multiple ethnic identifications: Methodological effects on child, adolescent, and adult ethnic distributions. *Demographic Research*, 44(21), 481-512.

<https://doi.org/10.4054/DemRes.2021.44.21>

Yao, E. S., Meissel, K., Bullen, P., Clark, T. C., Atatoa Carr, P., Tiatia-Seath, J., Peiris-John, R., & Morton, S. M. B. (2022). Demographic discrepancies between administrative-prioritisation and self-prioritisation of multiple ethnic identifications. *Social Science Research*, 103.

<https://doi.org/10.1016/j.ssresearch.2021.102648>

Yoder, J. D. (1999). *Women and gender: Transforming psychology*. Prentice-Hall, Inc.

Zava, F., Sette, S., Vecchio, G. M., Laghi, F., Baumgartner, E., & Coplan, R. J. (2024).

Exploring links between social competence, immigrant background, and children's preschool adjustment. *Early Education and Development*, 1-21.

<https://doi.org/10.1080/10409289.2024.2389367>

Zelazo, P. D., Jacques, S., Burack, J. A., & Frye, D. (2002). The relation between theory of mind and rule use: Evidence from persons with autism spectrum disorders.

Infant and Child Development, 11(2), 171-195. <https://doi.org/10.1002/icd.304>

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K.,

Hongwanishkul, D., Schuster, B. V., & Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in*

Child Development, 68(3), vii-137. <https://doi.org/10.1111/j.0037->

[976x.2003.00260.x](https://doi.org/10.1111/j.0037-976x.2003.00260.x)

Zelazo, P. D., Reznick, J. S., & Piñon, D. E. (1995). Response control and the execution of verbal rules. *Developmental psychology*, *31*(3), 508-517.

<https://doi.org/10.1037/0012-1649.31.3.508>

Zimmer, A. E., Zordan, E., & Montes, G. (2017). *Comparison of social emotional assessments, curricula, and competencies*. Children's Institute, T17-008.

<https://www.childrensinstitute.net/sites/default/files/documents/comparison-social-emotional-assessments-curricula-competencies.pdf>

Zins, J. E., Payton, J. W., Weissberg, R. P., & O'Brien, M. U. (2007). Social and emotional learning for successful school performance. In *The science of emotional intelligence: Knowns and unknowns*. (pp. 376-395). Oxford University Press.