



Faculty of Education and Social Work

**Primary school achievement gaps and school  
decisions to support the academic  
achievement of disadvantaged students with  
data: A cross-country comparative study**

Nicole Bien

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

Faculty of Education and Social Work

2016



Faculty of Education and Social Work

*Office of Doctoral Studies*

---

### **AUTHOR'S DECLARATION**

This is to certify that:

- I. this thesis comprises only my original work towards the Doctor of Philosophy Degree
- II. due acknowledgement has been made in the text to all other material used
- III. the thesis does not exceed the word length for this degree.
- IV. no part of this work has been used for the award of another degree.
- V. this thesis meets the *University of Sydney's Human Research Ethics Committee (HREC) requirements for the conduct of research.*

Signature:

Name: Nicole Ma Bien

Date: 29 March 2016

## Acknowledgements

The journey to undertake and complete this thesis has been incredibly rewarding. I am deeply grateful to the many people who accompanied me on my journey and supported me at various junctures so that I could reach my destination. First and foremost, I wish to express my heartfelt gratitude to my Doctoral Supervisor, Dr. Ilektra Spandagou for her generosity, patience and encouragement through the entire course of my candidature. I have benefited immensely from her tireless guidance and academic insights. I also wish to thank my Associate Supervisor, Associate Professor David Evans for his review and constructive feedback.

I owe my greatest thanks to my family for their love and unwavering support. Bill, my partner and soulmate, encouraged me to undertake this doctoral study and was my biggest cheerleader through the entire journey. More importantly, Bill's steadfast support for my career change has been indispensable. My sisters, Dr. Ma and Dr. Ma provided valuable thesis-writing advice and my brothers, Ken and Mitch offered ongoing moral support. My mother, Jenny, taught me perseverance through her own life story. Finally, my children, Bandon and Kaela, who understood the importance of this journey by checking my progress on a regular basis.

My thanks also go to Dr. Alun Pope for his review and feedback of the quantitative work and Sonia Bartoluzzi for proof-reading the final copy of this thesis. I am particularly thankful to the school administrators and teachers who gave up their precious time to participate in the interviews. The University also made the thesis possible by awarding generous funding support through the Thomas and Ethel Mary Ewing Scholarship, the Alexander Mackie Research Scholarship, and the Postgraduate Research Support Scheme. Finally, I am grateful to Dr. Armistead and Carol Mille, leaders of the School Board where I worked for approving a sabbatical for me to complete this thesis.

## Abstract

Reducing the academic disadvantage of all students is a significant educational goal for many countries. Increasingly, education reforms around the world, including those in Australia and the United States have sought to reduce achievement gaps by adopting a strategy of embedding accountability anchored by standardised assessment. Whether to meet federal and state educational requirements, to provide transparency to the general public, or to inform curriculum and instruction at individual schools, policy makers rely on assessment data and data-driven practice to make a difference. Although external forces such as policy expectations are generally the first step in creating social change, the internal beliefs of change agents can impact their course of action. Being agents of change, school educators can choose to adopt data-driven practice for compliance, or also to engage with data for continuous improvement. Applying the efficacy theory and the theory of planned behaviour from the social cognitive tradition, this thesis examined educators' belief mechanism regarding embracing data-practice, and considered the direct and indirect benefits of data-engagement, as well as the costs that ensued for teaching and learning.

Using standardised assessment results from 2008–2013 in Australia and two counties in California, and six case studies across New South Wales, California, and Hawaii, the present mixed methods research found evidence of progress in raising the proficiency of disadvantaged students, but not in narrowing achievement gaps between advantaged and disadvantaged students. The case studies suggest a positive relationship between academic proficiency progress and data engagement. This can be explained by the structural design and operational procedures of the data-driven process enhancing educators' attitudes, intention, perceived efficacy beliefs, and perceived behavioural control relating to the challenging task of raising the educational outcomes of disadvantaged students. As a result, participants could see beyond mere compliance with data-driven practice to its potential for professional improvement.

## Contents

<b>AUTHOR’S DECLARATION</b> .....	<b>ii</b>
<b>Acknowledgements</b> .....	<b>iii</b>
<b>Abstract</b> .....	<b>iv</b>
<b>List of Tables</b> .....	<b>ix</b>
<b>List of Figures</b> .....	<b>xi</b>
<b>Definition of Key Terms</b> .....	<b>xii</b>
<b>General Terminology</b> .....	<b>xii</b>
<b>American Terminology</b> .....	<b>xiii</b>
<b>Australian Terminology</b> .....	<b>xiv</b>
<b>Chapter 1 Introduction</b> .....	<b>1</b>
<b>Context</b> .....	<b>1</b>
<b>Rationale for the Study</b> .....	<b>5</b>
<b>Theoretical Framework</b> .....	<b>11</b>
Self-efficacy and collective-efficacy theory. ....	12
Theory of planned behaviour (TPB).....	14
A combined construct.....	15
<b>Purpose of the Study</b> .....	<b>16</b>
<b>Research Questions</b> .....	<b>18</b>
<b>Significance of the Study</b> .....	<b>18</b>
<b>Overview of the Methods</b> .....	<b>20</b>
<b>Overview of the Chapters</b> .....	<b>21</b>
<b>Chapter 2 Contextual Framework: the Evolution of Education Transparency and Accountability</b> .....	<b>23</b>
<b>Policy Context</b> .....	<b>24</b>
<b>The Race to Produce 21<sup>st</sup> Century Knowledge Workers</b> .....	<b>26</b>
<b>Education Inequity and the Persistent Achievement Gap</b> .....	<b>28</b>
Education inequity in the US .....	30
Education inequity in Australia.....	33
<b>Federal Solutions: Education Accountability</b> .....	<b>37</b>
The rise and fall of No Child Left Behind. ....	38
The Melbourne Declaration and its implementation. ....	40
<b>The Quest for Transparency and the Outsized Role of Testing</b> .....	<b>45</b>
<b>Conclusion</b> .....	<b>48</b>
<b>Chapter 3 The Impact of Test-Based Accountability Reforms</b> .....	<b>50</b>
<b>Needs and Challenges of Disadvantaged Students</b> .....	<b>51</b>
<b>Equal Access to External Assessment – Goal or Reality?</b> .....	<b>53</b>
<b>Effects on Student Performance and the Achievement Gap</b> .....	<b>60</b>
Academic outcomes of disadvantaged students in the US .....	62
Academic outcomes of disadvantaged students in Australia.....	67
<b>Unintended Consequences of Test-Based Accountability</b> .....	<b>71</b>

Cheating to inflate school performance.....	72
Teaching to the test and narrowing the curriculum.....	72
‘Education triage’.....	73
<b>Parallel Evidence in Australia.....</b>	<b>75</b>
<b>Contradictory Findings .....</b>	<b>77</b>
<b>The Promise of Data-Driven Practice and the Reality of Practice.....</b>	<b>79</b>
Perceived trust and benefits in external data. ....	81
The nature of data use.....	84
Data-engagement criteria – human factors. ....	84
Data-engagement criteria – systemic factors.....	86
<b>Conclusion.. .....</b>	<b>89</b>
<b>Chapter 4 Theoretical Framework .....</b>	<b>92</b>
<b>Identifying a Theoretical Framework .....</b>	<b>92</b>
<b>Efficacy Theory – An Overview .....</b>	<b>94</b>
Mastery experience. ....	96
Vicarious experience.....	97
Social persuasion. ....	97
Physiological and emotional state.....	98
Efficacy in the education context. ....	99
Justifications for the efficacy construct.....	103
<b>The Theory of Planned Behaviour .....</b>	<b>105</b>
TPB in the education context.....	109
Justifications for the Planned Behaviour Theory.....	110
<b>Theoretical Issues .....</b>	<b>111</b>
<b>Justifications for Combining the Two Theories .....</b>	<b>112</b>
<b>Conclusion.. .....</b>	<b>115</b>
<b>Chapter 5 Methodology .....</b>	<b>116</b>
<b>Context of the Study .....</b>	<b>116</b>
<b>Philosophical Assumption and Implication for Research.....</b>	<b>118</b>
<b>Mixed Methods Research Paradigm .....</b>	<b>119</b>
Rationale for mixed methods research. ....	120
<b>Research Questions .....</b>	<b>123</b>
<b>Research Design.....</b>	<b>124</b>
<b>The Quantitative Study .....</b>	<b>126</b>
Data sources. ....	128
Sample. ....	130
Definition of progress. ....	131
Definition of participation. ....	132
Data analysis and procedure. ....	133
Limitations. ....	142
<b>The Qualitative Study.....</b>	<b>144</b>
The role of the researcher. ....	146
Ethical considerations.....	147
Sampling procedure.....	148
Participating schools and participants.....	151
Data collection and sources of data. ....	159
Data analysis.....	166
Validity and reliability. ....	169

Limitations.....	169
<b>Conclusion..</b> .....	<b>170</b>
<b>Chapter 6 Quantitative Findings: Academic Progress in California.....</b>	<b>172</b>
<b>Data and Statistical Procedures .....</b>	<b>173</b>
Statistical procedures.....	173
Data summary.....	175
<b>Reading Achievement Trends .....</b>	<b>177</b>
<b>Mathematics Achievement Trends .....</b>	<b>183</b>
<b>Progress on Achievement Gap.....</b>	<b>189</b>
Reading.....	189
Mathematics.....	192
<b>Inclusion in Assessment .....</b>	<b>195</b>
<b>Discussion.....</b>	<b>195</b>
<b>Summary.....</b>	<b>198</b>
<b>Chapter 7 Quantitative Findings: Academic Progress in Australia .....</b>	<b>200</b>
<b>Data and Statistical Procedures .....</b>	<b>200</b>
Data summary.....	201
Statistical analysis.....	202
<b>Reading and Numeracy Achievement Trends .....</b>	<b>205</b>
Subgroup differences in reading and numeracy.....	208
Progress on NAPLAN.....	213
<b>Progress on Achievement Gap.....</b>	<b>217</b>
<b>Trends on NAPLAN Absence and Withdrawal.....</b>	<b>219</b>
<b>Discussion.....</b>	<b>222</b>
<b>Summary.....</b>	<b>225</b>
<b>Chapter 8 Becoming Data-Driven Schools.....</b>	<b>227</b>
<b>Case Context, Data Definition, and Student Support Structure .....</b>	<b>228</b>
Case context.....	228
Data definition.....	243
Support structure for low-performing students.....	245
<b>The Evolution of Data Engagement.....</b>	<b>249</b>
Perception of data utility.....	252
Nature of data use.....	256
<b>Similarities and Differences.....</b>	<b>265</b>
Instructional leadership.....	265
Goal-oriented, accountable and transparent.....	266
Goal attainment – the currency for programming and instructional freedom.....	266
Not 100% on board.....	268
Strategic versus tactical orientation.....	269
Perceived versus actual accountability pressure.....	270
Portfolio view versus strict assessment view.....	271
<b>Conclusion.....</b>	<b>272</b>
<b>Chapter 9 Interpreting Behavioural Change .....</b>	<b>274</b>
<b>The Theoretical Construct .....</b>	<b>275</b>
<b>The Formation and Influence of Efficacy Belief, Intention and Behaviour Control.....</b>	<b>277</b>
Higher degree of perceived behavioural control and mastery experience.....	278

Positive vicarious experience.....	283
An improved physiological and emotional state.....	285
Affirming social persuasion.....	287
Influences of attitude and subjective norms.....	290
<b>Relevance and Implications of Strong Efficacy Belief and Planned Behaviour .....</b>	<b>291</b>
Students are not the problem regardless of background. ....	293
Student-centric practice. ....	294
Collaboration over isolation. ....	296
Subsequent direct and indirect benefits. ....	299
<b>Costs of Data Engagement.....</b>	<b>306</b>
Constant testing.....	306
Seeing the forest for the trees.....	307
Compromising creativity.....	308
Limited visibility of students not at risk of failing.....	309
<b>Conclusion.. ..</b>	<b>310</b>
<b>Chapter 10 Discussion and Conclusions .....</b>	<b>313</b>
<b>Context.....</b>	<b>313</b>
<b>Significance of the Study .....</b>	<b>313</b>
<b>Key Findings.....</b>	<b>314</b>
Quantitative findings. ....	314
Qualitative findings.....	318
<b>Implications of the Study .....</b>	<b>322</b>
Prospects for closing the achievement gaps. ....	322
The ‘real’ value of data-driven practice.....	325
Lessons from policy-borrowing and policy-lending.....	327
<b>Limitations and Further Research Direction .....</b>	<b>330</b>
<b>Conclusion.....</b>	<b>331</b>
<b>References.....</b>	<b>334</b>
<b>Appendices.....</b>	<b>359</b>
Appendix A Recruitment Letter .....	359
Appendix B Participant Information Statement.....	360
Appendix C Participant Consent Form .....	362
Appendix D Interview Guide .....	364
Appendix E Colour-code Worksheet of Student Achievement at Almond .....	366
Appendix F Pacing Guide at Kukui.....	367
Appendix G Data Team Process at Hibiscus.....	368
Appendix H Learning Community Meeting Protocol at Hibiscus.....	369
Appendix I Instructional Leadership Team Meeting Agenda at Kukui.....	370
Appendix J Evaluation of Student Performance and Action Plan Part I at Almond .....	371
Appendix K Evaluation of Student Performance and Action Plan Part II at Almond .....	372



## List of Tables

Table 2.1	<i>Government Recurring School Funding</i> .....	32
Table 2.2	<i>Enrolment in Australian Schools (2008)</i> .....	35
Table 3.1	<i>Score Changes on the NAEP (1990 and 2013)</i> .....	63
Table 3.2	<i>NAPLAN Score and Minimum Proficiency Percentage Change between 2008 and 2014</i> .....	68
Table 5.1	<i>Advantaged and Disadvantaged Subgroups for California Data Comparison</i> .....	135
Table 5.2	<i>Standardised Beta Coefficients of Australian School Background Variables .....</i>	139
Table 5.3	<i>Disadvantaged and Advantaged Proxies for NAPLAN Categorical Variables .....</i>	140
Table 5.4	<i>Categorisation of NAPLAN Numeric Variables</i> .....	140
Table 5.5	<i>Summary of Case Study Participant</i> .....	152
Table 5.6	<i>Student Profiles at California Participating Schools and Districts (2011– 2012)</i> .....	154
Table 5.7	<i>Staff Information at Almond and Walnut (2011–2012)</i> .....	155
Table 5.8	<i>Student Profiles at Hawaiian Participating Schools and Complexes (2011– 2012)</i> .....	156
Table 5.9	<i>Staff Information at Kukui and Hibiscus (2011–2012)</i> .....	156
Table 5.10	<i>Australian Participating School Profiles (2012)</i> .....	158
Table 5.11	<i>NAPLAN Mean Scale Scores (2012)</i> .....	159
Table 6.1	<i>Percentage of Schools Represented by Subgroup in Grade 3</i> .....	176
Table 6.2	<i>Percentage of Schools Represented by Subgroup in Grade 5</i> .....	176
Table 6.3	<i>Grade 3 Reading Means and Proficiency Rates at Almond County</i> .....	178
Table 6.4	<i>Grade 5 Reading Means and Proficiency Rates at Almond County</i> .....	178
Table 6.5	<i>Grade 3 Reading Means and Proficiency Rates at Walnut County</i> .....	181
Table 6.6	<i>Grade 5 Reading Means and Proficiency Rates at Walnut County</i> .....	181
Table 6.7	<i>Grade 3 Mathematics Means and Proficiency Rates at Almond County</i> .....	183
Table 6.8	<i>Grade 5 Mathematics Means and Proficiency Rates at Almond County</i> .....	184
Table 6.9	<i>Grade 3 Mathematics Means and Proficiency Rates at Walnut County</i> ..	185
Table 6.10	<i>Grade 5 Mathematics Means and Proficiency Rates at Walnut County</i> ..	185
Table 6.11	<i>Grade 3 Reading Achievement Gaps between Groups at Almond County .....</i>	190
Table 6.12	<i>Grade 5 Reading Achievement Gaps between Groups at Almond County .....</i>	190
Table 6.13	<i>Grade 3 Reading Achievement Gaps between Groups at Walnut County .....</i>	191
Table 6.14	<i>Grade 5 Reading Achievement Gaps between Groups at Walnut County .....</i>	191

Table 6.15 <i>Grade 3 Mathematics Achievement Gaps between Groups at Almond County</i> .....	192
Table 6.16 <i>Grade 5 Mathematics Achievement Gaps between Groups at Almond County</i> .....	193
Table 6.17 <i>Grade 3 Mathematics Achievement Gaps between Groups at Walnut County</i> .....	194
Table 6.18 <i>Grade 5 Mathematics Achievement Gaps between Groups at Walnut County</i> .....	194
Table 7.1 <i>Disadvantaged and Advantaged Proxies for NAPLAN Variables</i> .....	202
Table 7.2 <i>Average NAPLAN Scale Score Means and Standard Deviations (2008–2013)</i> .....	209
Table 7.3 <i>Average NAPLAN Proficiency Rates and Standard Deviations (2008–2013)</i> .....	212
Table 7.4 <i>Difference in NAPLAN Means (2013 less 2008)</i> .....	216
Table 7.5 <i>Average NAPLAN Score Gaps and Growth Trends (2008–2013)</i> .....	217
Table 7.6 <i>Average NAPLAN Proficiency Gaps and Growth Trends (2008–2013)</i> .....	218
Table 7.7 <i>NAPLAN Absentee/Withdrawal Percentage Change Over Time (2013 less 2008)</i> .....	221
Table 8.1 <i>The Meaning of 'Data' in the Interviews</i> .....	244
Table 8.2 <i>Nature of Data Use</i> .....	256

## List of Figures

<i>Figure 1.1.</i> The relationship between accountability and standardised assessment ...	4
<i>Figure 2.1.</i> The Completed Australian Curriculum .....	42
<i>Figure 3.1.</i> Year 3 Literacy and Numeracy Withdrawal Rate .....	57
<i>Figure 3.2.</i> Impact of NCLB on NAEP Outcomes in Grade 4 .....	65
<i>Figure 4.1.</i> Combined Theoretical Framework .....	113
<i>Figure 5.1.</i> Analysis Input and Output Variables .....	127
<i>Figure 6.1.</i> Reading Proficiency between Grade Levels at Almond County .....	179
<i>Figure 6.2.</i> Reading Proficiency between Grade Levels at Walnut County .....	182
<i>Figure 6.3.</i> Mathematics Proficiency between Grade Levels at Almond County .....	186
<i>Figure 6.4.</i> Mathematics Proficiency between Grade Levels at Walnut County .....	187
<i>Figure 7.1.</i> Year 3 Reading and Numeracy Average Score amongst School Groups	207
<i>Figure 7.2.</i> Year 5 Reading and Numeracy Average Score amongst School Groups	207
<i>Figure 7.3.</i> Year 3 NAPLAN Mean Score by LBOTE Quartile .....	211
<i>Figure 7.4.</i> Year 5 NAPLAN Mean Score by LBOTE Quartile .....	211
<i>Figure 7.5.</i> Year 3 Absentee/Withdrawal Percentage by Jurisdiction .....	220
<i>Figure 7.6.</i> Year 3 Absentee/Withdrawal Percentage by School Type .....	220
<i>Figure 8.1.</i> Data Engagement Models .....	241
<i>Figure 8.2.</i> Models Supporting Low-Performing Students .....	246
<i>Figure 8.3.</i> The Relationship between Data Source and Depth of Data-Driven Practice .....	255

## Definition of Key Terms

This study focuses on two countries: the United States of America (US) and Australia. The former uses American English and the latter Australian English. To ensure consistency, Australian English will be used throughout this thesis except for formal names, such as the *Center for American Studies*. Some terms are used more commonly in one country than the other; and some terms apply solely to one country's accountability and transparency reform. For clarity, these terms are identified and defined in the tables below.

### General Terminology

Progress in International Reading Literacy Study (PIRLS)	Administered by the International Association for the Evaluation of Educational Achievement (IEA) every five years, PIRLS is an international literacy comparative study of fourth-grade equivalent students. Australia and the US are both among the 53 participants.
Programme for International Student Assessment (PISA)	Administered by the Organisation for Economic Cooperation and Development (OECD), PISA is a triennial international assessment and survey evaluating the literacy, mathematics and science knowledge of 15-year old students. Australia and the US are both among the 65 participants.
Proficiency standard	The state (in the US) or national (in Australia) reading and mathematics minimum standards that students are expected to meet to be considered proficient.
Trends in International Mathematics and Science Study (TIMSS)	Administered by the International Association for the Evaluation of Educational Achievement (IEA), TIMSS is an international comparative study of mathematics and science among fourth-, eighth- and twelfth-grade equivalent students every four years across 60 countries. Australia and the US are participants.

## American Terminology

Academic Performance Index (API)	Enacted in 1999 in California, API is an annual measure of school academic performance using standardised exams. Scores range from 200 to 1,000, with a state-wide target of 800.
Adequate Yearly Progress (AYP)	Required under No Child Left Behind, AYP is an indicator that measures the extent to which students in a school, taken as a whole, and certain student subgroups within the school, demonstrate proficiency in reading and mathematics.
English Language Learners (ELL)	Students for whom English is not their first language.
Every Student Succeeds Act (ESSA)	Approved by US Congress, President Obama signed ESSA into law in December 2015. Replacing NCLB, ESSA reauthorised the Elementary and Secondary Education Act of 1965.
Individuals with Disabilities Education Act (IDEA)	Enacted in 1975 by US Congress, IDEA requires services to be provided to children with disabilities.
National Assessment of Educational Progress (NAEP)	Authorized by the US Department of Education, NAEP is a continuing and nationally representative measure of achievement in various subjects over time. The assessment is administered to fourth-, eighth- and twelfth -graders across the US through random sample every two years.
National Center for Education Statistics (NCES)	The primary US federal entity for collecting and analyzing data related to education including the NAEP.
No Child Left Behind (NCLB)	Approved by US Congress, President Bush signed NCLB law in 2002. The law reauthorised the Elementary and Secondary Education Act of 1965.
Title I Schools	Title I is a US federal designation designed to provide funding support to schools serving 40% or more students from a disadvantaged background. As a condition for this funding, NCLB held districts receiving Title I funding accountable for making adequate yearly progress (AYP) for two consecutive years.

## Australian Terminology

Australian Curriculum, Assessment and Reporting Authority (ACARA)	An independent statutory authority with responsibility for developing and implementing curriculum and assessments. It is also responsible for collecting and publishing information on <i>My School</i> about the performance and resources of more than 9,500 schools around the country.
The Index of Community Socio-Educational Advantage (ICSEA)	This index is a socio-educational advantage scale of individual schools computed based on the socio-educational advantage (SEA) and other student level background factors.
Indigenous	A NAPLAN classification used to identify students of Aboriginal and/or Torres Strait Islander origin.
Language Background other than English (LBOTE)	A NAPLAN classification used to identify students or parents/guardians who speak a language other than English at home.
<i>My School</i>	Developed by ACARA, <i>My School</i> is an online tool that captures the performance and resources of schools in Australia to provide publically accessible information.
National Assessment Program Literary and Numeracy (NAPLAN)	A series of common literacy and numeracy tests conducted annually across Australia for all students in Years 3, 5, 7 and 9.
Socioeconomical Advantage (SEA)	Socioeconomical advantage indicator of students.

## Chapter 1 Introduction

This comparative study evaluates disadvantaged students' progress and achievement gaps as identified by external assessment. In addition, it explores school-level decisions to engage with data-driven practice as a strategy to raise student outcomes, particularly the outcomes of disadvantaged students. The study employed a mixed methods approach and was conducted in six metropolitan primary schools in New South Wales (NSW), Australia, and in the US in both California and Hawaii. Assessment results from 2008–2013 from Australia and from two Californian counties served as the basis for a quantitative analysis of progress on achievement gaps. Semi-structured interviews with school administrators and teachers provided empirical data to explore data-engagement decisions, data practices and their impact on student outcomes and teacher professionalism. The combined theoretical frameworks of the efficacy theory and the theory of planned behaviour were employed to explain administrators' and teachers' decisions and motivations to engage with data practice for continuous improvement.

### Context

Globally, advocacy for educational rights for disadvantaged children has come a long way: from education exclusion, to segregation, to inclusion. The implementation of educational and disability policies around the world, intended to guarantee equal educational access and participation for traditionally marginalised children, has resulted in a 96% enrolment rate across the developed nations (The United Nations, 2009). However, after decades of education inequality, a gap

remains of an average 39-point difference in achievement between disadvantaged and advantaged students as measured by the Programme for International Student Assessment (PISA) (OECD, 2012b), and one in five disadvantaged students drop out from high school (OECD, 2012a). This achievement gap is equivalent to one year of formal education (OECD, 2012b) and, if persists, disadvantaged students are not expected to gain the basic proficiency to function in society (OECD, 2012a).

Wide achievement gaps have prompted governments around the world to implement transparency- and accountability-related education reforms to address education inequities, such as *Excellence in Schools 1997* in the UK, and *No Child Left Behind* (The U.S. Office of Under Secretary, 2002) in the US. A number of European and Latin American countries have had their own accountability policies for over a decade (Figlio & Loeb, 2011). Borrowing the accountability concept from the US, the Australian government followed suit in 2008 (Lingard, 2010) with its version of education transparency and accountability through *the Melbourne Declaration* (Australian Ministerial Council on Education, 2008). While the US and Australian accountability policies differ in their implementation strategies, each contains three central pillars, which scholars (Ball, 2012; Lingard, 2010) commonly refer to as: managerialism, choice, and performativity. Put it simply, under the current reforms, schools are subject to three forms of accountability: external, market and internal (Goldring & Berends, 2009).

*External accountability or managerialism*, refers to a duty to satisfy the organisational hierarchy, where one level of an organisation is held accountable to the next: teachers to principals, principals to district superintendents or regional directors, and further up the chain to the to the Federal Department of Education



(Goldring & Berends, 2009; Sachs, 2001). In the US and in Australia, it is the higher governing bodies – federal and state – that define benchmarks and requirements for raising student achievement; and other levels of the school system work to meet these regulations.

*Market accountability* refers to school choice (Ball, 2012). Under this accountability reform, parents are no longer constrained by catchment areas for school enrolment. Instead, they have the flexibility to transfer their children to the school of their choice, and the student's local catchment school must comply and relinquish funding for that particular child to the chosen school. Market accountability is intended to exert pressure on schools to deliver academic outcomes or else face the prospect of losing community support (Ball, 2012; Lingard, 2010; Ravitch, 2010).

*Internal accountability*, the final pillar, has been the subject of much less research and discourse than the other two pillars. It is the focus of the current research study. Internal accountability is concerned with “school workplace norms; local decision making; and school goals, assessments, and consequences” (Goldring & Berends, 2009, p. 17). As described by the same authors, it differs from external accountability in that the emphasis is on the local group of educators within a school. Internal accountability is about a school's unique response through programming, staffing and pedagogy to meet external accountability requirements.

The common thread linking all three pillars of accountability is mandated standardised assessment, as demonstrated in Figure 1.1. For external accountability, assessment data serve as the metric used by higher education authorities to evaluate state, district or regional, and school performance. For market

accountability, data forms a common benchmark used by students, parents and the greater community to evaluate their neighbourhood schools. Finally, under internal accountability, assessment data are expected to inform schools about where there are gaps and challenges, and to enable them to monitor their progress against goals. What policy makers rely on when making a case for testing is its diagnostic value (The Honorable Julia Gillard Australian MP, 2010, May 5; US Secretary of Education Ann Duncan, 2010). As the Melbourne Declaration (Australian Ministerial Council on Education, 2008) in Australia articulates:

It [data] supports effective diagnosis of student progress and the design of high-quality learning programs. It also informs schools' approaches to provision of programs, school policies, pursuit and allocation of resources, relationships with parents and partnerships with community and business. (p. 16)



*Figure 1.1.* The relationship between accountability and standardised assessment

As might be expected, the implementation of test-driven accountability reform on this scale has attracted an unprecedented level of discourse and research in a variety of areas including: student performance, teacher and student morale and

well-being, implementation challenges, test reliability, and teacher pedagogy (Hess & Finn Jr, 2007; Lingard, 2013; Polesel, Dulfer, & Turnbull, 2012; Ravitch, 2010).

While a significant amount of research has focused on and found evidence of unintended consequences of accountability testing, a report (Wöbmann, Lüdemann, Schütz, & West, 2007) published by Organisation for Economic Cooperation and Development (OECD) has demonstrated a strong positive correlation between student achievement and different elements of accountability, autonomy and choice. Given these mixed views, the present research set out to evaluate changes in achievement gaps since the implementation of the reforms, and to investigate school and teacher responses to accountability reform through data-driven practice designed to close achievement gaps.

### **Rationale for the Study**

Governments in both the US and Australia have set aggressive goals in support of their respective education reforms. In enacting *No Child Left Behind* (NCLB), US lawmakers aimed to achieve universal literacy and mathematics proficiency by the year 2014 (Ravitch, 2010). Australian policy makers committed to raising reading, mathematics and science ratings as measured by international assessments, to rank within the top five countries globally by 2025 (Ferrari, 2012, September 3). In both countries, the magnitude of these goals stirred passionate debates in all sectors of the community: the media, academics, practitioners, policy analysts, parents and students. Even critics could agree with the value of the policy goals (Rose & Gallup, 2006; Sirotnick, 2004), and acknowledged (Cuban, 2014) the merit of, and the consistent effort required for eradicating inequality in education

“for those children who have been historically least well served: the poor, minorities, newcomers” (p. 29). One critic, Beadie (2004), further observed that, “the general mood in current education reform is that this linkage [student performance to school performance] is good” as it “finally provides the political and financial leverage necessary to make schools and teachers concentrate effectively on meeting the needs of failing students” (p. 36). In practice, however, “*accountability* has become a dirty word” (Reeves, 2004, p. 5) because the high-stakes nature of this approach can have a serious impact on the very survival and reputation of schools and teachers, and on the success of students (Australian Education Union, 2013; Lingard, 2010; Ravitch, 2010). Representing the views of his colleagues, Sirotnik (2004) summed up the sentiment of external accountability as follows:

None of us believe that there are no good uses for test-based assessment or evaluation strategies. All of us believe that appraisal is important, that the public has a right to know how well schools are educating their children, and that the very nature of education itself should model good inquiry and reflective practice. All of us, however, are deeply concerned about what happens when heavy-handed accountability schemes get superimposed on the complexities of schooling practice. (p. 9)

Internal accountability is greeted with more receptivity in the US, although not in Australia. Even critics of the NCLB policy such as Hess and Finn Jr. (2007), contended that, despite its flaws,

We welcome the flood of additional information, and we suspect that, in the long run, NCLB’s greatest accomplishment may be the school-performance data it furnishes to parents, educators, and state and local officials, data on the basis of which they can make desirable changes in their own schools and their choices among schools. (p. 316)

Proponents of data-driven practice could not agree more (Goldring & Berends, 2009; Mandinach & Jackson, 2012; Reeves, 2004). First, because data-driven practice relies

on “objective evidence, rather than on anecdotes” (Mandinach & Jackson, 2012, p. 17) for student, teacher, program and school evaluations. Secondly, using data to analyse gaps between learning goals and actual performance “defines actions of effective schools” (Goldring & Berends, 2009, p. 5). Goldring and Berends (2009) further suggested that “decision making, setting and prioritizing goals, and monitoring progress” (p. 5) is important for school improvement. As opponents and proponents continue their debates, Young and Kim (2010) of the American Education Policy Analysis Archives summed up the current status of data this way, “the importance of using ‘data’ is now taken-for-granted as an essential strategy for educational improvement” (p. 3).

Despite the potential positive impact of data engagement on student achievement, there is less attention given in empirical research to productive or constructive data-driven practice than there is to counter-productive data use. As the literature review in Chapter 3 demonstrates, standardised assessment as a measure of accountability, engendered a significant number of unintended behaviours in the early years of the reform (Dulfer, Polesel, & Rice, 2012; Heilig & Darling-Hammond, 2008; J. Lee, 2010). These unintended behaviours have fuelled a large part of the policy evaluations and discussions in both countries (Nichols & Berliner, 2005; Polesel et al., 2012; Ravitch, 2010). While there is some evidence of constructive data practices (Datnow, Park, & Wohlstetter, 2007; Wayman, Cho, & Johnston, 2007), uptake has been slow. Furthermore, aside from unintended consequences, internal accountability driven by assessment data is not without challenges. Research (Pierce & Chick, 2011; Wayman et al., 2007; Young & Kim, 2010) reveals that schools lack the technological capability to house and disseminate

data, teachers lack the necessary analytic skills to use the large quantity of data, and data culture at schools is not a given. Less explored in the literature are the motivations (aside from benefits articulated by policy makers) that move educators to change their practices and behaviours to adopt data-driven practice.

This study concentrates on internal accountability systems, where the school is the unit of analysis. Unlike external factors such as socioeconomic status (SES), or school funding, the internal accountability measure is where school principals and teachers can directly seek to make real and incremental progress. Specifically, this study explores how schools and teachers respond to policy makers' expectations of the data-driven practice that is inherent in test-based accountability reform designed to raise student achievement. As Goldring and Berends (2009) observed, leading, instructing and improving schools with data is not easy for educators; if it were, "evidence-based decision making [would be] ubiquitous" (p. 2). It is the intent of the present study to build on constructive data-engagement research to support disadvantaged students and to explore educators' decisions to embrace data-practice as the core element of internal accountability.

At the heart of this research is the alignment between reforms' intention to promote data-driven practice and educators' individual and collective decisions to implement such practice within their local school contexts. Does enactment of a policy automatically lead to changed behaviour in educators? Or are educators' intentions to implement policy in practice shaped by strong intrinsic and extrinsic forces that have a greater influence than the policy itself? Based on the unintended consequences uncovered in the media and in research, it is fair to assume that changes in instructional behaviour do not follow merely because policy dictates

them. Indeed, important as policy makers consider data-driven practice, Datnow, Park and Wohlstetter (2007) observed in their study that, “it is at the school level where everything comes together” (p. 16). It is behavioural change at the school level that could deliver the promises of reform. For this reason, the main concern of this study is to evaluate those behavioural beliefs, attitudes and intentions that lead to behavioural changes. Exploring both the internal and external factors that influence school leaders’ and teachers’ engagement with data can contribute to improved implementation of policies designed to raise student outcomes.

It is important to point out that data-driven accountability practice represents just one aspect of the policy, and it is one factor among a complex set of factors that affect the success of a school and its students. These factors include, but are not limited to the historical, social, academic, and financial contexts and other challenges that have resulted in an academic gap between advantaged and disadvantaged students in the first place (Darling-Hammond, 2004, 2007; Ma Rhea, 2012; Noguera, 2004; Oakes, Blasi, & Rogers, 2004; Sirotnick, 2004). However, the critical role of data and the success of data-driven practice in sectors outside of education suggest that it could have a positive impact within education also, and is worth studying.

Furthermore, this study acknowledges that any reform at this level of complexity, scale and goal is political, and carries implementation risks. Therefore, it does not intend to evaluate whether accountability reform is set up with the proper structure, remedies, or incentives to guarantee success. Rather, this study begins with the following premises:

- (1) Any reform, big or small, at federal, state or school level, takes time to realise its goals due to the complex nature of schools, the school system and the individual needs of students.
- (2) Incremental successes may build momentum towards the larger goal of raising the status and performance of disadvantaged students.

As Figlio and Loeb (2011) noted in their evaluation of US accountability policy, “It is clear that there is no one ideal accountability system. The optimal system for one context and one set of policy goals is unlikely to be the optimal system for another context and another set of policy goals” (p. 416). To this end, any reform of this scale would face colossal implementation challenges, as many critics have revealed (Dulfer et al., 2012; Hess & Finn Jr, 2007; Klenowski & Wyatt-Smith, 2012; Sirotnick, 2004). It would therefore be unreasonable to expect sweeping changes within half a decade or even a decade. Incremental changes are more realistic and could have a higher chance of reaching the end goals. Commenting on leadership for school reforms, Cuban (2014), an education historian at Stanford University, makes the following observation on his blog, “school boards would do well to downsize expectations, display more patience, seek leaders who believe in incremental changes toward fundamental ends...”.

It is with the perspective of incremental change that this study utilised a mixed methods approach involving large-scale assessment data analysis and selected case studies in NSW, Hawaii, and California:

- (1) To evaluate the academic progress of disadvantaged students under the current accountability environment.



- (2) To explore the belief mechanisms behind data engagement and the relationship between data use and disadvantaged student achievement.
- (3) To compare the impact of the transparency accountability policy on disadvantaged student learning and teaching in both Australia and the US.

The six participating schools have demonstrated effective data practice or have made visible progress in student performance. For the purpose of this study, disadvantaged students included those who are economically, physically, cognitively, academically, or racially disadvantaged. Some students in this population may experience only one type of challenge, while others might struggle against multiple challenges at the same time. To the extent that data are available, this study has examined each disadvantaged group separately to gauge progress, as well as the support provided to each group in this new accountability era.

## **Theoretical Framework**

Whether school administrators and teachers choose to use assessment data destructively or constructively, they are important agents in any education reform process. Reform, according to the dictionary (Merriam-Webster, n.d.) is improvement by change in behaviour or habits. The social cognitive perspective is most appropriate to explain the ways in which educators exercise their personal agency to embrace data practice at the local level. Conceptualised by Bandura (Bandura, 1986b), this perspective suggests that behaviour change can be influenced by elements in the environment as well as by one's own cognitive process. The present study integrates two complementary theories within the social behavioural

perspective to explain the qualitative findings of this research. The first relates to Bandura's self and collective efficacy theory (1977, 1997), and the second, the theory of planned behaviour (Ajzen, 1991). Both models have been applied extensively to predict behavioural intentions. These theories provide a framework to investigate teachers' personal and collective agency, and the antecedents of behaviour change to embrace data engagement as a central part of instructional practice. These antecedents include intentions, motivations, efficacy beliefs, and outcome expectations. The theory of planned behaviour and collective efficacy, working in aggregate, help to explain participating educators' contemplations of data engagement as part of their daily practice. Understanding their behavioural beliefs can shed light on the motivations behind schools' responses to policy makers' calls for data-driven practice. Illuminating the necessary determinants to behaviour change could increase the number of schools and teachers adopting the desired change for internal accountability to influence student achievements.

#### **Self-efficacy and collective-efficacy theory.**

Bandura theorised self-efficacy (1997) as an individual's beliefs in his or her "capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). The efficacy theory postulates that an individual's action is determined by two sets of beliefs. The first is intrinsic and reflects that individual's belief about his or her capability to take on a particular action. The second is extrinsic and it concerns the expected effects or outcomes of his or her action (Bandura, 1977, 1997). According to Bandura, efficacy beliefs that are internally focused have more lasting power in the face of challenge compared to extrinsic expectancy beliefs. Efficacy is an important construct for the understanding of

behaviour because efficacy belief constitutes “the key factor of human agency” (Bandura, 1997, p. 3). The stronger a person’s self-perceived ability to perform an action, the more effort he or she will exert; in contrast, the weaker that perception, the less likely it is that he or she will take the action (Bandura, 1977, 1993). These efficacy beliefs do not happen in a vacuum; they are shaped by four sources of information: (1) mastery experiences; (2) affective state; (3) vicarious experience; and (4) social persuasion (Bandura, 1977). The concept of self-efficacy has been applied frequently in education for three decades, and in particular to the study of teacher self-efficacy belief and its impact on student outcomes (Shaughnessy, 2004). Many of these studies (Y. L. Goddard, Goddard, & Tschannen-Moran, 2007; Tschannen-Moran & Hoy, 2007) have found that teacher self-efficacy has significant predictive power over student achievement.

Collective efficacy is a newer construct that shares comparable properties to self-efficacy (Bandura, 1995, 1997; R. D. Goddard, 2001). This construct refers to the collective beliefs held by members of an organisation about their collective ability to realise a common goal (Bandura, 1995). In the education context, perceived collective teacher beliefs refer to teachers’ shared beliefs that the entire staff can create and implement an action plan needed for students’ success (R. D. Goddard, Hoy, & Hoy, 2000, 2004). Based on findings in their previous study (R. D. Goddard & Goddard, 2001), Goddard et al. (2004) conclude that, “collective efficacy beliefs are far more strongly related to teachers’ perceptions of self-capacity than many more common measures of school context” (p. 9). Research (Bandura, 1993; R. D. Goddard et al., 2000) has also validated the mediating role of teacher efficacy on student achievement. It is evident from these studies (Bandura, 1997; R. D. Goddard &

Goddard, 2001) that perceptions of efficacy are paramount for the individual and organisational behaviour change at schools that ultimately affects student performance.

**Theory of planned behaviour (TPB).**

Ajzen's theory of planned behaviour (1991, 2002) postulates that a person's intention or readiness to undertake a particular action is a function of three determinants: attitude, subjective norm and perceived behavioural control. All three are functions of behavioural beliefs. The TPB builds on the theory of reasoned action (Ajzen & Fishbein, 1980), which includes attitude and subjective norm. Together, attitude and subjective norm lead to the formation of intention to predict an action. However, volition alone addresses only part of the equation as in some situations, people do not have control over the action of interest (Madden, Ellen, & Ajzen, 1992) (for example, lack of financial resources to start a project, or lack of control over a prospective employer's hiring process to secure a much-desired job). Addressing this gap, Ajzen (1991, 2002) added the perceived behavioural control construct to the extended TPB. TPB states that attitude is the individual's initial inclination to favour or not favour a change. Subjective norm refers to the social pressure influencing the decision to change. Finally, perceived behavioural control pertains to whether the person feels control over the external elements involved in the action being contemplated. The perceived behavioural control construct within the TPB shares similar properties to the efficacy theory (Ajzen, 2002; Manstead & Van Eekelen, 1998). Crediting efficacy theory for the origin of the perceived behavioural control construct, Ajzen (1991) noted that, "both are concerned with perceived ability to perform a behaviour" (Ajzen, 2002, p. 668).

In their aggregate, intention and perceived behavioural control account for both internal attitudes and responses to external influences. Ajzen (1991) further postulates that the likelihood of the intended actions being carried out increases when all three TPB determinants are met. The predictive power of the TPB on behaviour has been empirically proven in multiple disciplines, including healthcare (Godin & Kok, 1996), consumer behaviour (Kalafatis, Pollard, East, & Tsogas, 1999; S. Taylor & Todd, 1995), and unethical behaviours (Beck & Ajzen, 1991; Chang, 1998) to name a few. Application of this theory within the school context has been less prolific (Crawley, 1990; Haney, Czerniak, & Lumpe, 1996; J. Lee, Cerreto, & Lee, 2010). However, findings from the majority of existing studies on teacher intention towards targeted actions corroborate findings in other disciplines; namely, that the three determinants within the TPB significantly predict targeted behaviours, though their weights vary according to context, population, and disciplines.

#### **A combined construct.**

There are two reasons for combining the efficacy and the TPB constructs. First, many scholars (Maddux & Stanley, 1986), including Ajzen (2002), believe that the two theories are compatible and complementary in relation to the perceived ability to perform an action of interest. Secondly, together, these theories address both internal and external perceived constraints over a contemplated behaviour (Manstead & Van Eekelen, 1998), therefore providing a more comprehensive framework to guide the present study. As the efficacy construct is intended to be more narrow and task specific (Ajzen, 2002; Kirsch, 1986; Pajares, 1996), the additional antecedents of attitude and subjective norm from the TPB allow for a more global interpretation of the factors at play in the high-pressure accountability

school environment under which data-driven practice takes place. At the same time, Bandura's efficacy theory and Ajzen's perceived behavioural control construct can guide evaluation of the internal beliefs that affect the decisions to engage with data of novice and experienced teachers, high- and low-performing schools, or low-SES against average schools. Working together, these two theories provide a framework that is broad enough to address the internal complexity and external factors impacting a school or a group of teachers as they decide whether to embrace data-driven practice.

Unlike testing and public reporting, it is not easy to mandate or enforce data-driven practice as a means to inform curriculum and instruction. As agents in the school reform process, schools and teachers can exercise influence over what they do in their classrooms, at their grade/year levels, and in their schools. To encourage behavioural change, it is necessary to understand educators' personal and collective beliefs and the antecedents influencing those beliefs. Understanding these motivators can contribute to the development of strategies to encourage more data use at the local level.

### **Purpose of the Study**

Current education reforms in the US and in Australia share the two ambitious goals of raising student achievement and closing achievement gaps between advantaged and disadvantaged students. Unfortunately, the high-stakes nature of these accountability reforms continues to dominate debates, while empirical evidence evaluating student academic progress and data use is scant in Australia and comparative research is lacking despite many nations also have an accountability

policy. At the same time, the requirement for public accountability in education is unlikely to fade. In his global review, Fullan (2009) predicts that, “we will see a great expansion and deepening of large-scale reform strategies in the immediate future, not only in the U.S. but across the world” (p. 101). The US Congress had reauthorised NCLB at the end of 2015, and the new Act, the Every Student Succeeds Act (US Department of Education, 2015) maintains two key aspects of NCLB: accountability and annual external assessments. In fact, both of these elements are so much embedded in the US education system, that current US policy discourse has extended beyond school accountability to teacher accountability (Figlio & Loeb, 2011), also based on the same measure – student performance assessment data.

Borrowing Reeves (2004):

As educators, we have two choices. We can rail against the system, hoping that standards and testing are a passing fad, or we can lead the way in a fundamental reformulation of educational accountability. (p. 6)

The current research began with the notion that there are school leaders and teachers who see opportunities to turn what critics perceive as a destructive, unedifying and messy reform into a constructive and transformative improvement process. The first step towards this is to identify areas of improvement because schools “cannot improve what they do not measure” (Barber & Mourshed, 2007, p. 52). This was a sentiment that top-performing schools articulated in Barber and Mourshed’s (2007) global study exploring “*How the World’s Best-Performing School Systems Come Out on Top*”. Data, and the engagement with data, are among the key ingredients in this critical step towards improving student achievement (Halverson, Grigg, Prichett, & Thomas, 2007; Mandinach & Jackson, 2012). Therefore, it is important to evaluate and to compare how academic achievement for less

advantaged students has been maintained or advanced by national policies and local instructional change with the advent of data across both countries.

## Research Questions

The current study began with the following questions:

- (1) What have the test-based accountability policies in Australia and California accomplished in the area of assessment inclusion and achievement of disadvantaged students?
- (2) How and why have school administrators and teachers chosen to invest time and effort in data-driven practice to support student learning?
- (3) How have the different policies in the two countries affected the learning and teaching experiences of disadvantaged students and their teachers?

## Significance of the Study

The current education reforms in the US and in Australia speak to contemporary trends in globalisation and to policy lending and borrowing. These trends lead Darling-Hammond, a noted scholar (2010) to entitle her book *The Flat World and Education* to describe the future of education and of schools as a place to develop global knowledge workers, a thought also articulated by Ball (2012) in *Global Education Inc.* Given these trends, comparative or cross-cultural research is ever more necessary to avoid the pitfall of generalising policy or intervention impacts across countries without paying attention to local contexts and meanings (Booth & Ainscow, 1998; Broadfoot, 2000; Lingard, 2010; Steiner-Khamsi, 2004a; Zymek & Zymek, 2004). Careful analyses of different cross-cultural perspectives,



contexts, and practices relating to similar problems can provide powerful and fresh insights into the home system (Booth, 2003; Grant, 2000; Welch, 2011). Using a different lens can bring to light “features that are normally ignored”, “possibilities that have been overlooked” (Booth & Ainscow, 1998, p. 5), or simply “what we take for granted” (Welch, 2011, p. 190) in a single-context analysis. A comparative analysis of how test-based accountability reforms have played out in each country can help to identify to what extent the concept of accountability has worked under different cultural and educational contexts. These insights are crucial for assessing the educational impact and social implications of an assessment and accountability society (Crossley & Watson, 2003) and for contributing to future policy modification discourse (Zymek & Zymek, 2004). Findings from this comparative analysis could also inform and benefit future policy refinements not merely in the countries studied, but in others contemplating similar reforms.

In regards to data-driven practice, after a decade of reform enactment in the US, data-driven practice has finally garnered momentum in schools. Mandinach and Jackson (2012) believe that “data use is no longer a passing fad, one to which educators can close their doors and assume it will go away until the next innovative idea appears” (p. 11). These authors contend that data-driven decisions contribute to continuous improvement through the process of organisational learning. Even leading critics agree that data-driven education practice could be one reform outcome that brings about desirable changes (Hess & Finn Jr, 2007). For these reasons, further understanding how educators internalise their decisions to embrace data for continuous improvement to meet educational goals could encourage more schools to consider constructive, rather than destructive, use of data. Since much of

the research and commentary on data use in schools focuses narrowly on the nature of data use and the challenges of data use, there is a need for research that examines the motivators behind its adoption.

This study has a particular interest in and focus on the progress of disadvantaged students and on how schools and teachers support this student population. Thus this study speaks not only directly to the goals of the accountability reforms, but also to the need to find solutions towards eliminating inequality in education. As the OECD (2012a) concludes, “school failure penalises a child for life. The student who leaves school without completing upper secondary education or without the relevant skills has fewer life prospects” (p. 3).

## **Overview of the Methods**

A mixed methods approach was selected for the present study, which took place in three geographical locations: Oahu, Hawaii; Northern California; and the Sydney metropolitan area of NSW. The methodological choice of mixed methods was guided by “a practical and applied research philosophy” of the pragmatism worldview (Tashakkori and Teddlie, 2003 cited in Creswell & Plano Clark, 2007, p. 27) where the research aims to “draw from ‘what works’, [by] using diverse approaches and valuing both objective and subjective knowledge” (Creswell & Plano Clark, 2007, p. 26). Together, quantitative and qualitative techniques increase the potential for explanation-building of the observed phenomenon (Howe, 2012) Furthermore, because the study evaluated predictor variables whose “manifestations have already occurred”, as opposed to something that the research

actually manipulated (B. Johnson & Christensen, 2008, p. 357), mixed methods was deemed the most appropriate design.

The mixed methods approach provided an effective way to investigate the multiple research goals previously outlined. The first research question regarding student progress and the achievement gap was answered by quantitative methods. It involved statistical analyses using *T* tests, analysis of variance (ANOVA), and multiple regressions to explore student progress on the Australian National Assessment Program Literary and Numeracy (NAPLAN) and the California Standardised Test (CST) results between 2008 and 2013. Qualitative case studies through the use of semi-structured interviews provided data to explore the second question regarding the motivations behind data engagement. Finally, results from both phases were pulled together to frame the analysis of the third question pertaining to the accountability policies' impact on student performance and teacher professionalism.

## **Overview of the Chapters**

This thesis comprises ten chapters. This introduction has painted a broad picture of the research project. Chapter 2 begins the literature review. It outlines the evolution of recent education reform in Australia and in the US with special attention paid to their respective historical, educational and cultural similarities and differences. The reforms' moral aspirations and philosophical beliefs are also discussed. Chapter 3 continues the literature review by detailing the documented effects of these policies to date, their unintended consequences, and current data engagement practice. Chapter 4 introduces the theoretical frameworks selected to

interpret the qualitative empirical findings of the present study. Chapter 5 describes the chosen research methodology and procedures to conduct the research and to analyse the empirical data. This chapter also addresses ethics issues and limitations of the research. Chapters 6 and 7 report the Australian and the US results respectively, from the quantitative analyses, which explored progress on their respective reform goals. Findings from six case studies exploring the evolution and nature of assessment data engagement follow in Chapter 8, which also explores the impact of context as well as similarities and differences between the two countries. Chapter 9 seeks to interpret participants' behavioural change to embrace data-driven practice through the combined theoretical constructs of efficacy and the theory of planned behaviour, and how the change impacted learning, teaching and school management. Finally, Chapter 10 reflects on the key findings and significance of the study, and draws applicable lessons, including cross-cultural similarities and differences. It closes with a discussion of the study limitations and possible directions for future research.

## Chapter 2 Contextual Framework: the Evolution of Education

### Transparency and Accountability

This chapter outlines the evolution of test-based accountability reform in the US and in Australia through an examination of literature. It begins with a discussion of the global trends in education that led to the current reforms. Next, it traces, compares and contrasts each country's historical, educational and cultural contexts and discusses how the current reform fits into each of the educational contexts. It also reviews the definition of, and the current status of, disadvantaged students, as well as how the accountability reforms intend to support their learning. The chapter ends with an examination of each policy's implementation strategy and the political implications of external assessment, leading the way into Chapter 3 where the literature on the implications of assessments for student achievement and teaching practices will be presented.

The literature review for both chapters draws from a wide range of sources. They include but are not limited to the following: peer-reviewed journals, reports from international organisations such as the World Bank and OECD, legislations, government publications, government official speeches and press releases, education system websites such as those of the departments of education in Australia and in the US and their respective testing entities such as the Australian Curriculum, Assessment and Reporting Authority (ACARA) and the US National Center for Education Statistics (NCES), newspaper articles, non-governmental and non-university affiliated research institutions, and published books.

## Policy Context

In 2001, the US Congress unanimously passed NCLB, a major federal education reform aimed to support K-12 education (Ravitch, 2010). This reform was a reiteration of the initial Elementary and Secondary School Act introduced in 1965. Between 1965 and 2002, US Congress approved seven reiterations (New America EdCentral, n.d.). NCLB introduced the concepts of school accountability, transparency and parental choice through student testing and data reporting. Based on the volume of literature on this topic, NCLB was arguably the most drastic of these reforms and stirred passionate debate among politicians, academics, practitioners and parents. In December 2015, US Congress passed the most recent reiteration called the *Every Student Succeeds Act* (ESSA). Accountability, transparency and choice accomplished through student testing remain the key tenets of the reauthorised Act.

In 2004, the Australian Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) passed the *Schools Assistance (Learning Together – Achievement through Choice and Opportunity) Act 2004* (Attorney-General's Department, n.d.). This Act committed funding of 33 billion Australian dollars to schools for the period 2005 to 2008. In 2008, the Act was reauthorised. Consistent with the spirit of the US NCLB, the focus of the Act was accountability. One element of the Act mandated the implementation of national literacy and numeracy testing for all students in Years 3, 5, 7 and 9, and public reporting of school performance. In 2008, the Australian Education Ministers of all eight Australian jurisdictions, along with the federal government, signed the *Melbourne Declaration on Educational Goals for Young Australians* (Australian Ministerial

Council on Education, 2008) and subsequent education *National Partnerships* to support teaching and learning (Council of Australian Governments, n.d.). The Melbourne Declaration aimed to promote equity and excellence among young Australians and the intention was to achieve these goals through more accountability and partnership between schools, parents and the government.

While the US and Australian education reform Acts were enacted at different times, they are intricately linked in a number of ways and were driven by similar national and global trends. First, the then Australian Labour government drew the accountability and testing concepts from New York State (Lingard, 2010). Secondly, in regards to student performance, both countries had fallen behind on global rankings as assessed by PISA since 2000. Although Australia is still one of the top performers on the PISA, growth has stagnated and its ranking slipped substantially in reading, mathematics and science (Gonski, 2011, December). This trend has not reversed since the enactment of these reforms. For example in mathematics, fewer Australian students in 2012 scored at or above proficiency than did so in 2003, while more students fell below proficiency level in 2012 than did so in 2003 (Schleicher, 2014). This pattern held true for US students (Schleicher, 2014), whose PISA scores stagnated between 2000 and 2009 (OECD, 2010c). More discouragingly, US students performed below average in mathematics and science and around average in reading (OECD, 2010c). Thirdly, two dominant global trends, globalisation and education inequity, converged at the same time and provided further impetus for the implementation of test-based accountability in education.

## **The Race to Produce 21<sup>st</sup> Century Knowledge Workers**

Within the academic and research communities, globalisation is generally accepted as being the catalyst that sparked contemporary education reforms in the Anglo-American countries (Cheng, 2010; Darling-Hammond, 2010; Reimers, 2010; Rizvi & Lingard, 2000; Spring, 2009; Welch, 2001). Globalisation “refers to both the intensity and the extent of international interactions” (Rizvi & Lingard, 2000, p. 421). Specific to the education context, Spring (2009) describes globalisation as, “the worldwide discussions, processes, and institutions affecting local education practices and policies” (p. 1). This line of discourse emerges from the growing global interconnectedness in economic relations, science and technology, communication media and cultural exchange (Darling-Hammond, 2010; Lingard, 2010; Schriewer & Marinez, 2004; Steiner-Khamsi, 2004a).

The interconnectedness of economies and industries has had an impact on education in two significant ways. First, it created urgency among policy makers, such as the then, Australian Education Minister, to articulate the need to develop “one of the most highly educated and skilled nations on Earth” to feed the “global economy hungry for knowledge and skills” (Gillard, 2008, April 4). Justified by the need to ensure that students have the appropriate knowledge and skills to compete in and meet the needs of the global economy, many nations, particularly the Anglo-American countries, intensified their involvement in school education in the form of new public management (Hartley, 2003). Scholars contend that this new form of public management of schools has had far-reaching effects on education, most visibly on the role of the federal government in education (Crossley & Watson, 2003; Welch, 2001). The US NCLB Act is one such example. In fact, in reauthorising the



education Act in 2015, US President Obama (2015, December 10) acknowledged the failure of the “one-size-fits-all approach” imposed by the federal government by reverting school management back to state authorities.

Secondly, globalisation and the need for global knowledge workers have prompted unprecedented financial support, and influence on policy, from well-endowed charitable foundations funded by businesses, particularly in the US (Ravitch, 2010). Some scholars believe that this civic and business coalition could benefit the entire community, because it brings together disparate parties in a quest to find solutions to improve schools (Cuban, 2004). However, like Ravitch (2010) and Ball (2012), Cuban also fears that the private-civic partnership has the potential to narrow the broader goals of education. As Welch (2013) explains, policies are not value free, and, “interventions by key interest groups” are entirely possible (p. 188).

Thirdly, the emergence of the global knowledge economy has also lent prominence to international educational assessment programs such as TIMSS, PIRLS and PISA, whose mission is to provide international benchmarks for the purpose of “assist[ing] policy-makers in identifying the comparative strengths and weaknesses of their educational systems” (International Association for the Evaluation of Educational Achievement, n.d.). Broadfoot (2000), a leading comparative education scholar remarks that in this climate of intense global economic competition, “governments have become increasingly obsessed with the international rankings of measured educational outcomes” (p. 360). Other researchers (Baker and LeTendre 2005, cited in Spring, 2009) also observe that, “around the world, countries are using the results of international tests as a kind of Academic Olympiad, serving as a referendum on their school system’s performance” (p. 62). Indeed, these league

tables' powerful influence on policy formation (Crossley, 2014) is evidenced in the Australian government's education target: to be ranked in the top five by 2025 on these measures (Ferrari, 2012, September 3). Along with comparing themselves on the global assessment benchmarks, "governments around the world [are] anxious to learn about educational practices in other countries" (Broadfoot, 2003, p. 411). In so doing, Welch (2001) argues that education becomes homogenised and social diversity lost. Chapter 3 will further explore the waves of discussion this prompts, among all constituents, regarding the many unintended consequences for students and educators.

### **Education Inequity and the Persistent Achievement Gap**

Globalisation, however, has also created greater inequality between societies and between individuals regarding the skills and means to benefit from the opportunities, and the capacity to enjoy the goods and services created by the global economy (Rizvi & Lingard, 2000). Education inequality permeates most societies, including developed nations such as the US and Australia, with the latter being comparatively more equitable according to the OECD (2010b) in part due to their differing educational funding models, to be discussed later. Among the factors that lie behind education inequity are a long history of structural inequities, exclusion and segregation (Hardman & Dawson, 2008; Lingard, Creagh, & Vass, 2012), and low expectations (Cimera, 2007; Hehir, 2005; Pullin, 2005) towards African and Native Americans in the US and towards the Indigenous and Torres Strait Islander Peoples in Australia. Against the backdrop of education inequality and the desire to develop human capital to participate in and to enjoy the benefits of the global economy,

recent education reforms throughout the Western world have flavoured what policy researchers (Ball, 2005a; Welch, 2013) describe as markets, managerialism and performativity (Ball, 2005a). The then Australian Rudd/Gillard government's *Education Revolution (ER)* articulated these motivations:

As the world changes, the consequences of being left behind are increasingly harsh. So our agenda for national prosperity involves investment in all and high expectations that every Australian who can participate, will participate in our changing economy. (Gillard, 2010, June 10)

The then US Secretary of Education, Duncan, reminded the American public the rationale behind the reform ideals:

Eleven years ago, Congress, with strong bipartisan support in the Senate and the House, rightly said that our schools needed to focus on all students; that for America to continue to succeed, all of our children had to succeed. That is why NCLB sought to hold every State, district, and school accountable for 100 percent of students being proficient in reading and math by the end of the 2013-2014 school year. (US Department of Education, 2013)

Equal education opportunity through high expectations for all resonates widely, for it is a global issue. According to the OECD (2005, 2007, 2010b), students who generally perform at below-grade level include those students: with a physical, intellectual, cognitive or sensory impairment; with general learning difficulties; from disadvantaged or minority backgrounds; and/or who are recent immigrants. Most low performers also share a similarly disadvantaged socioeconomic background (OECD, 2011). This population is disproportionately represented in diverse socio-cultural and ethnic metropolitan and urban neighbourhood schools (Darling-Hammond, 2004; Harris, 2007; Smith, 2005) and in very remote communities (Wigglesworth, Simpson, & Loakes, 2011). Through decades of exclusion, segregation and political marginalisation (Hardman & Dawson, 2008; Lingard et al., 2012),

socioeconomic-related factors (Darling-Hammond, 2010), low expectations (Cimera, 2007; Hehir, 2005; Pullin, 2005), and deficit beliefs among educators (Love & Kruger, 2005; Nelson & Guerra, 2014), disadvantaged students' academic performance and education attainment persistently lag behind their advantaged peers.

### **Education inequity in the US**

In tracing the history of inequity, Darling-Hammond (2010) offers that “institutionally sanctioned discrimination in access to education is older than the American nation itself” (p. 28) and it is “the Achilles’ heel of American education” (p. 8). She further suggests that globalisation exacerbates the scale of education disparity for US students. This is because it has changed the skills required for today’s students to succeed in the 21<sup>st</sup> century, for which the US education system has not prepared them (Darling-Hammond, 2010). The struggle for education equality, according to Darling-Hammond (2010), “has played out in each historical era for racial/ethnic minority groups, new immigrants, and the poor, surfacing in decisions about whom to educate, with what resources, where and how, and toward what ends” (p. 28). The latest involves the changing demographics in the US. In 1990, Hispanic students represented only 7% of eighth graders; by 2011, this figure jumped to 23% (National Center for Education Statistics, 2013b). That puts the US in the sixth place in terms of countries with the largest immigrant population among OECD members (OECD, 2012c). In California, as many as 32% of fourth-grade students are identified as English language learners (ELL) (National Center for Education Statistics, 2013b). With this demographic shift, education inequity between non-immigrant children and immigrant children whose native language is not English has intensified.

Today, among Grade 4 and 8 students, 22% to 27% fewer Black students than White students achieve reading proficiency (National Center for Education Statistics, 2013b). Students with disabilities lag their non-disabled peers by 30% (Eckes & Swando, 2009). Nation-wide, only 16–18% Hispanic, Black and economically disadvantaged fourth graders meet proficiency, as compared to 42% White students and 48% economically advantaged students (National Center for Education Statistics, 2013b). A measure by the OECD (2012a) found similar trends: US students from a low SES background are 2.5 times more likely to score below level 2, the OECD proficiency threshold, and recent immigrants are 1.3 times more likely than their non-immigrant peers to fall below this threshold. On the OECD equity index in education, the US scores below the OECD average (OECD, 2012a).

By many accounts, the US public school funding model, notorious for its inequity in comparison with other Western countries, is a key culprit for today's education inequity (Biddle & Berliner, 2002; Darling-Hammond, 2010). As Table 2.1 demonstrates, US government schools receive 8–9% funding from the federal government, with state income tax and county property tax funding the remaining balance. This funding model results in immense discrepancies between advantaged and disadvantaged neighbourhood schools (Zhou, 2010). Biddle and Berliner (2002) found that the disparity between a wealthy community and an impoverished one can be as large as US\$15,000 versus US\$4,000 per child, respectively. Across the 50 states, the latest census (United States Census Bureau, 2015, June 2) found a discrepancy of US\$13,000 per child between the lowest and highest funding states. While the US federal government, through NCLB and legislation prior to NCLB, provides additional funding to low-income schools through the Title I program, it

allocates funds partly based on state-level funding per student, hence perpetuating the funding gap (Darling-Hammond, 2010). Many critics, therefore argue that social inequities, unrelated to standards and teacher quality, are the locus of the education challenge that leaves the less advantaged students behind (Darling-Hammond, 2004; Noguera, 2004; Oakes et al., 2004).

Table 2.1

*Government Recurring School Funding*

	<i>Australia</i> <sup>1</sup>		<i>The US</i> <sup>2</sup>
	<i>Government Schools</i>	<i>Non-government Schools</i>	<i>Government schools</i> <sup>3</sup>
Funding model	Adjusted by school sector		Differed by local taxes
Range per student expenditure	A\$10,800–15,600		US\$6,600–19,800
Average per student expenditure	A\$12,700		US\$10,000–11,000
Federal funding	15%	33–57%	9%
State/territory funding	80%	12–20%	46%
County/regional funding	0%	0%	45%
Private Sources	5%	23–55%	Not known

Notes: <sup>1</sup>Figures represent 2008 data.

<sup>2</sup>Figures represent 2013 data.

<sup>3</sup>In the US, non-government schools receive no funding from local, state or the federal government. (Dowling, 2007; Gillard, 2010, April 15; Gonski, 2011, December; Rorris et al., 2011; US Department of Education, n.d.; Welch, 2010; Zhou, 2010)

Over the years, multiple reform efforts have been launched to promote educational parity. In the 1950s, *Brown vs. Board of Education* ruled that the segregation of Black and White students is a violation of the American Constitution (Ravitch, 2010). A decade later, the landmark Elementary and Secondary Education Act (ESEA) of 1965, a product of the Civil Rights Movement, followed (Cuban, 2004; Lessinger, 1970). The ESEA extended US federal school funding to those living in poverty and kept up the pressure for desegregation. Another significant piece of legislation, which aimed to support children with disability in education, the

*Education for All Handicapped Children Act*, was passed by the US Congress in 1970. Since then, multiple iterations of these aforementioned education Acts have been implemented, with the most recent iterations being NCLB and ESSA. The strategies employed to bring about change have included funding buildings, texts and teachers; programs to address challenges resulted from student background characteristics (disability, migrants, etc.); and providing parents with school choice (Cuban, 2004; Goldring & Berends, 2009). Legislative efforts over the last few decades may have paid off somewhat, as the OECD (2012c) has noted a decrease in socioeconomic impact on student performance over the years. Nonetheless, socioeconomic factors still explain 15% of the variance in the PISA performance of US disadvantaged students, a proportion that is significantly higher than the top-performing nations such as Hong Kong, Norway, Japan and Finland (OECD, 2012c).

#### **Education inequity in Australia.**

The OECD has given Australia, one of the highest performers on PISA, a positive review regarding equitable and inclusive education (OECD, 2013) that extends to most student subgroups. Along with other top-performing countries such as Singapore, Finland and Canada, Australia ranks above average in education equity (Schleicher, 2014). For example, even though Australia and the US are among the ten OECD countries with the highest immigrant populations, the educational achievement of immigrants in Australia does not differ significantly from non-immigrant cohorts (Schnepf, 2007). Such outcome parity does not hold true in the US (OECD, 2012c). The OECD (2013) attributed the equity outcome in Australia to “fair and inclusive practices” and “positive learning environments, strong pedagogical leadership and well-prepared teachers” (p. 4). Like many European

countries, Australia operates on a significantly more equitable funding model than the US. What sets Australia apart from many other OECD nations is its policy to fund both government and non-government schools (Gonski, 2011, December; Sriprakash & Proctor, 2013). It has been suggested that this funding model creates real school options for parents (Gonski, 2011, December; OECD, 2013).

However, there are reasons to question the equity of this funding model. First, as demonstrated in Table 2.2, a disproportionately large percentage of disadvantaged students are enrolled in government schools (Rorris et al., 2011). Funding provided to non-government schools is therefore distributed at the expense of disadvantaged students at government schools who might require more support. Secondly, the current funding model has been found to be muddled, complicated and incoherent (Gonski, 2011, December; OECD, 2013). The landmark government funding review, the *Gonski Report* (2011) concluded that this funding model leads to duplicate and inefficient funding between Australian federal and state governments that does not differentiate between students with needs and those without needs. Specific to students with disabilities, the same report found that those attending non-government schools are unintentionally penalised, as this inefficiency results in substantially less funding for them than for their counterparts in government schools. For these reasons, one of the key recommendations from Gonski's Report, is to provide separate and significant loadings to support each disadvantaged student group based on specific or multiple needs.



Table 2.2

*Enrolment in Australian Schools (2008)*

	<i>Government schools</i>	<i>Non-government Schools</i>
Students with disability	80%	20%
Indigenous students	86%	14%
English as second language <sup>1</sup>	90%	10%
Students from low-SES background	77%	28%
Students from remote locations	80%	20%
Students from very remote locations	89%	11%

*Note:* <sup>1</sup>The authors (Rorris et al., 2011) noted that these figures are underestimated due to missing data.

Generally positive Australian outcomes overall on the world stage also mask the entrenched inequity experienced by the Aboriginal and Torres Strait Islander peoples. The performance of Indigenous students on the OECD benchmark is over-represented in the lowest-performing level and under-represented in the highest level for all key domains (Sriprakash & Proctor, 2013; Thomson & De Bortoli, 2008). Another study, using one-year data from Queensland, found that the educational outcomes of Indigenous students were significantly behind non-Indigenous peers by age ten, and the gap widened as students aged (Bradley, Draca, Green, & Leeves, 2007). Compared to other first peoples, Australian Aboriginal and Torres Strait Islander students have the worst educational results among Western societies with a settler past (Bradley et al., 2007). The Council of Australian Governments considers this subgroup of students to be “the most educationally disadvantaged” (The Senate Standing Committee on Education and Employment, 2010, p. 19).

Education inequity with respect to Indigenous children is the legacy of institutional racism and discrimination through which Aboriginal and Torres Strait Islander children were denied education, among other government-sanctioned

discriminatory actions (Ford, 2013). For example, according to Pearson, a leading Aboriginal lawyer, academic and land-rights activist, decades of historical “dispossession, trauma, discrimination, and the undermining of Indigenous leadership and authority” (Pearson 2005, p. 7 cited in Ma Rhea, 2012) have denied generations of Aboriginal people the right to take responsibility for themselves and for their families. The remoteness of many Indigenous communities further compounds these complicated challenges (Bradley et al., 2007). Yunupingu (2008, December), another eminent Indigenous Australian, believed that the gap between Indigenous and non-Indigenous Australians had grown wider, despite all the government’s efforts. The legacy of discrimination leaves little faith amongst the Indigenous population that “anyone except the few of us who have lived our lives in the Aboriginal world understand this task that is called ‘closing the gap’” (Yunupingu, 2008, December, p. para. 30).

Multiple governmental programs have been established since the 1970s’ to tackle education inequality in Australia (Sriprakash & Proctor, 2013). Chief among them are special-purpose programs and partnerships with various communities created in the current millennium aimed to narrow achievement gaps (Gonski, 2011, December; Ma Rhea, 2012). The most significant were those created after the signing of the *2008 Melbourne Declaration* by all jurisdictional governments and the Australian federal government (Gonski, 2011, December). Of the nine partnerships under this agreement, four are directly related to supporting disadvantaged students: Smarter Schools, Literacy and Numeracy; Smarter Schools – Low SES School Communities; Closing the Gap in the Northern Territory; and More Support

for Students with Disabilities. Together, these partnerships received A\$11.5 billion of additional funding between 2008 and 2015 (Gonski, 2011, December).

### **Federal Solutions: Education Accountability**

Globalisation, education inequality and decline in student performance have given rise to federal governments' deeper involvement in school education. In the US, as in Australia, states and territories have constitutional responsibility over the delivery of education matters (Cranston, Kimber, Mulford, Reid, & Keating, 2010; Whitty, Power, & Halpin, 1998). The Commonwealth or federal government has always had an indirect influence over state education systems through funding for collaborative programs and special education (Ravitch, 2010; Reid, 2009) or through passing general legislation (Whitty et al., 1998) such as disability-related Acts to protect the rights to education of marginalised children. In the last two decades, however, the political climate has shifted from considering education as a 'public welfare' to viewing it as a private or individual good in the interest of economics (Hartley, 2003; Reid, 2009; Welch, 2010). Along with this shift, the federal government's indirect involvement has gradually become more direct in coordinating national policy priorities for school education and investment (Cranston et al., 2010; Donnelly & Wiltshire, 2014; Jennings & Stark Rentner, 2006).

Few debate the nobility of the goals of reform, particularly the goal to raise the achievement of all students, regardless of background. However, critics on both sides of the ocean believe that the focus of current reforms on output rather than input is failing to address the real educational issues (Darling-Hammond, 2010; Harris, 2007; Klenowski, 2009; Lingard, 2013; Ma Rhea, 2012; Ravitch, 2010). These

issues include: SES conditions (Darling-Hammond, 2004; Hattie, 2003; Kozol, 2005; Ladwig, 2010); non-inclusive practice regarding students with disability (L. Graham & Jahnukainen, 2011); and cultural understanding in reference to Indigenous children in Australia (Ford, 2013; Yunupingu, 2008, December) and immigrant students in both countries (Creagh, 2014; National Council of Teachers of English, 2008).

### **The rise and fall of No Child Left Behind.**

NCLB was passed with bipartisan support in the US Congress with the grand goal of closing achievement gaps (Ravitch, 2010). Specific elements encapsulated in NCLB included: a common national curriculum standard and formal assessment; publication of school and student performance to inform government funding decisions and for public consumption; parental choice in schooling; and performance management and school restructuring schemes (Darling-Hammond, 2004; ED.gov, 2010; Hartley, 2003; Ravitch, 2010). NCLB expected competition and parental pressures to raise standards and strengthen accountabilities (Ravitch, 2010; G. Wood, 2004). To date, the US federal government has appropriated roughly US\$25 to US\$27 billion annually to these initiatives (US Department of Education, n.d.). NCLB mandated all publically funded schools annually assess students in Grades 3 to 8 on basic literacy and mathematics. Every school was expected to meet *adequate yearly progress* (AYP), an index developed to reach the universal proficiency goal for every student in literacy and mathematics by 2014 (Goldring & Berends, 2009; Hess & Finn Jr, 2007).

Under NCLB, a school that failed to meet the annual AYP goal for two consecutive years was flagged for *program improvement* (US Department of Education, 2003). Schools in program improvement were expected to offer students

the choice to attend a different school and to provide opportunity for students to obtain supplemental education services, such as after-school tutoring by outside service providers. Furthermore, a school in program improvement had to embark on various corrective actions, including changing teaching staff or administrative staff, reorganising, or closing entirely. The rewards and punishments for meeting these goals were directly linked with Title I funding and other federal grants meant to serve the disadvantaged student population (Goldring & Berends, 2009; US Department of Education, 2003).

The year 2014 came and went, and the US national proficiency level did not come close to 100% (A. Klein, 2014, January 8), despite signs of improvement (which are described in Chapter 3). In fact, fewer than 50% of students in disadvantaged subgroups had reached proficiency as of 2013 (National Center for Education Statistics, 2013a). Anticipating this outcome, the US Federal Department of Education made legislative amendments in 2012 enabling the states to apply for waivers and to meet the requirements using other indicators (A. Klein, 2014, January 8; US Department of Education, 2013). However, funding remains tied to test performance (Obama, 2012, February 9; The White House, n.d.). At the end of 2015, the US Congress reauthorised the education Act and renamed it the *Every Student Succeeds Act* (US Department of Education, 2015). All elements of accountability are retained in ESSA. The key difference is that the management of accountability is returned to state governments. Furthermore, the accountability target has moved from absolute or status achievement which measures the number of students reaching a benchmark, to achievement growth, which measures the actual growth in

scores at a particular school district or of a particular disadvantaged student subgroup (A. Klein, 2014, January 8).

Just as NCLB before it, ESSA received bipartisan support from the US Congress because the implementation of NCLB was fraught with issues (Tooley, Dec 24, 2015). While AYP was put in place, NCLB neither mandated a national curriculum nor provided guidance on test materials, alternative assessment for students with special needs, or AYP targets with which schools were being evaluated (Smith, 2005). The development of *Common Core State Standards* for literacy and numeracy was not launched until 2009 and it was an effort coordinated by US state governments and territories (Common Core State Standards Initiative, n.d.-a) rather than by the US federal government. States have only begun to adopt the Common Core Standards since 2013. For over a decade, the lack of a common benchmark and platform rendered neither comparable nor generalisable achievement data. The lack of comparability created major validity controversies, because one state's proficiency level could be another's failing level (Heilig & Darling-Hammond, 2008; Pullin, 2005; Thurlow & Johnson, 2000). Whether ESSA will succeed where NCLB did not, only time will tell. For the time being, the general consensus is that scaling back federal intervention and returning schooling responsibility to state governments is a positive move (Tooley, Dec 24, 2015).

### **The Melbourne Declaration and its implementation.**

With economic rationalisation in recent decades tilting the balance of education towards outcome-based reforms (Welch, 2010), the Australian Commonwealth looked abroad and cited the New York City school system as an exemplary model to support its educational objectives. Policy borrowing from the US

puzzled the Australian education community, since US students trailed Australian students in all PISA and TIMSS measures (Caldwell, 2010; J. Graham, 2010; Lingard, 2010; OECD, 2010a). Policy borrowing resulting from cross-national attraction, however, is not uncommon (Ochs & Phillips, 2004) and there are more similarities than differences between the US and Australia with respect to constitutional responsibility for education and the shift in global educational agendas. New York City was a pioneer of accountability-based reforms (J. Klein, 2011) and it “became a national symbol of success in the late 1990s” (Ravitch, 2010, p. 33), due in part to its success in raising achievement scores across many student subgroups (Ravitch, 2010). Thus there was some legitimacy to using the New York City school system as a model.

A range of motives can drive nations to borrow policy (Phillips, 2000; Steiner-Khamsi, 2004b). In this case, Australia’s motivation appears to have been a wish to advance domestic reforms, using policies that are gaining ground elsewhere. This motive can be especially important when the policy to be introduced might be highly contested. Borrowing from New York City appears to satisfy what Ochs and Phillip (2004) describe as the political and social agendas aimed to justify the introduction of a policy. After all, the concepts of transparency and accountability were gaining ground in both the UK and the US and they continue to be endorsed by powerful agents such as the World Bank, the OECD and business philanthropists (for example, the Gates Foundation and Broad Foundation) for their affinity to the economic, or business model of value creation (Ball, 2005a; Lingard, 2010; Ravitch, 2010).

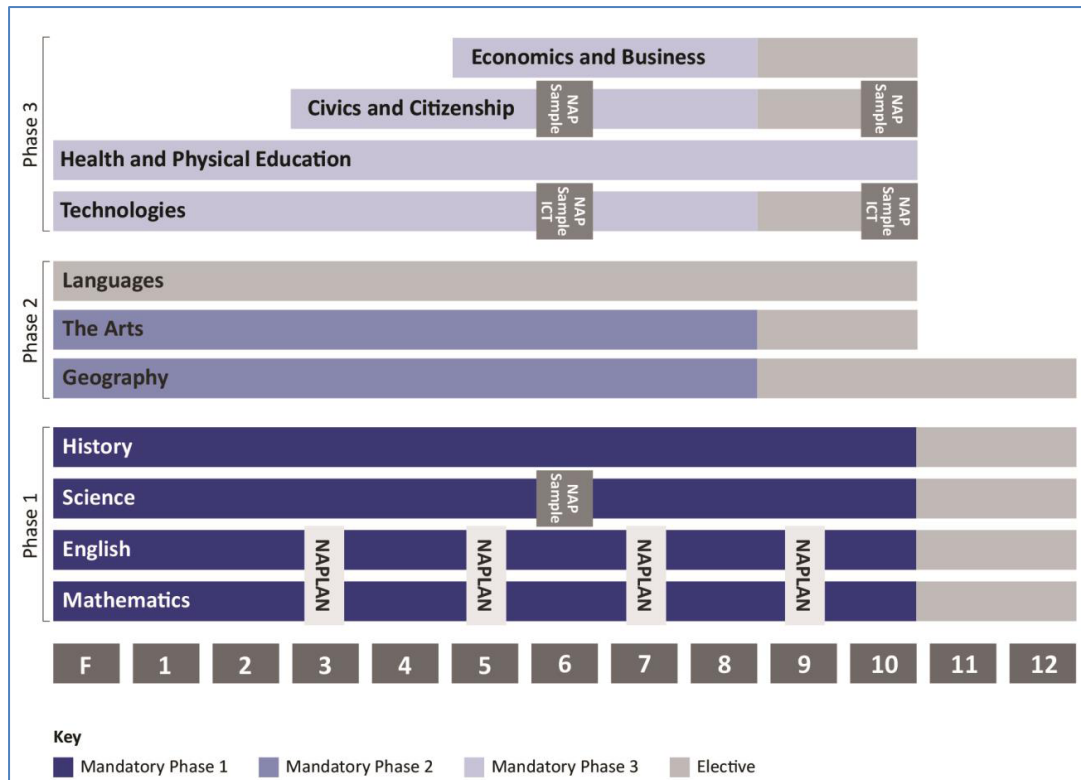


Figure 2.1. The Completed Australian Curriculum

Reproduced from *Review of the “Australian Curriculum: Final report”*, by the Australian Government Department of Education, 2014, Figure 5.

Conceptually, the Australian education reform reflects many key ideas anchoring the American education reform – namely: national standards, assessments, accountability, quality and choice. However, its policy development process is viewed as more collaborative (Cranston et al., 2010) and its implementation of the reform concepts differs considerably from the US approach. To begin with, the Australian government was committed to replacing individual state assessment and curricula with national assessment and a national school curriculum for all key subjects (Figure 2.1). In November 2013, ACARA completed the Foundation to Year 10 curriculum for all 11 domains (Australian Government Department of Education, 2014). This was a not so insignificant feat, as multiple attempts to create a national curriculum had been made over 30 years (Australian



Government Department of Education, 2014; Cranston et al., 2010). As the curriculum was being developed, ACARA launched NAPLAN, the first nationwide assessment to test all qualified students in Years 3, 5, 7, and 9 and the *My School* website to publish school-level data and NAPLAN results. Unlike the US reform, Australian education authority does not use NAPLAN results to label schools as 'failing' or 'passing' (Polesel et al., 2012). Instead, the results inform additional funding support from the government for underperforming schools (Polesel et al., 2012)

Compared to earlier reforms, the funding support of the latest reform has been substantial. To provide some context, between 1999–2000, the federal government invested a total of A\$4.8 billion in school education; between 2009–2010, total school funding quadrupled to A\$20 billion (Gonski, 2011, December). Almost half of the increase (A\$9.4 billion) is associated with the Building the Education Revolution initiatives. Additionally, A\$24.78 billion have been allocated to nine National Partnerships to achieve the reform outcomes from 2009–2017 (Gonski, 2011, December). Among these Partnerships, the initiatives directly related to disadvantaged students include three key 'Smarter Schools National Partnerships', totalling A\$2.6 billion over a five-year period to address economic and cultural disadvantages, to support teacher quality, and to improve literacy and numeracy (The Senate Standing Committee on Education and Employment, 2010). In addition, in 2010, the Australian federal government launched the Aboriginal and Torres Strait Islander Education Action Plan, with the hope of narrowing the gap between Indigenous and mainstream students (Gonski, 2011, December; OECD, 2013).

Despite the Australian reform's lack of contentious punitive measures, the introduction of external testing and the open publication of data have been the subject of controversy. The media use the data published to construct school league tables, and parents presume data comparability "notwithstanding the questionable validity of the use of the data" (Polesel et al., 2012, p. 7). This type of comparative activity has led to claims and findings of negative impacts on student well-being, teaching and learning (Athanasou, 2010; Australian Education Union, 2013; Thompson & Harbaugh, 2013). On the positive side, in interviews conducted by the Australian Council for Education Research (ACER) (Rorris et al., 2011), schools have been very receptive to the partnership funding model, because it is efficient, flexible and targeted to schools' and students' needs. One of the Indigenous communities that the Australian federal government and Queensland state government supported attested to "'statistically significant improvements' in school attendance... from 46.1 per cent in the first term of 2008 to 70.9 per cent in the first term of 2012" (Karvelas, 2013, March 28). Improvement in attendance can have significant impact on student outcome, as the multiple regression result in Chapter 5 will demonstrate. Given its short history, the long-term impact of the recent reform remains to be seen. At present, however, it appears that Australia's reform implementation strategies have avoided the mistakes of NCLB in the US – namely: (1) overreach of federalism; (2) avoidance of punitive measures; and (3) identification of what schools should be without prescriptions for how to get there.

## The Quest for Transparency and the Outsized Role of Testing

Standardised assessment is a familiar concept in most developed countries around the world (Irons & Harris, 2007; The Senate Standing Committee on Education and Employment, 2010). However, prior to the turn of the millennium, its predominant role was to serve as a teaching tool to inform educational practice, professional training development, and student grade promotion (Ravitch, 2010; The Senate Standing Committee on Education and Employment, 2010). Few scholars disagree with the benefits of assessments, especially tests that align with standards and curriculum (Klenowski & Wyatt-Smith, 2012; Ravitch, 2010). Testing can have value for teaching and learning modification based on analyses of students' responses, as practiced in Queensland (Klenowski, 2011) and in a few large school districts in the US (Reeves, 2004). The intentional use of assessments, both formative and summative, can contribute to teachers' understandings of students' strengths and weaknesses, support teachers in developing consistent student evaluation, and enable comparison of student progress and achievement (Goldring & Berends, 2009; Klenowski, 2011). What critics of external assessment do not support is reliance on a single measure to drive multiple critical decisions. Linking student promotion, school funding, school restructuring, and teacher bonuses, among other incentives and sanctions, to a single benchmark has made these external standardised assessments high-stakes. Standardised assessments, are thus given an outsized role, whereby they effectively serve as a monitoring tool for policy administrators and the public (Hardy & Boyle, 2011; Polesel et al., 2012). Sociologists and education historians believe that pressure from high-stakes testing is a method whereby Western societies are turning children into adults before they are developmentally ready

(Bousfield & Ragusa, 2014; Ravitch, 2010). Using evidence from the Australian Senate Inquiry, Bousfield and Ragusa (2014) cite inappropriate pressure, expectations, and precocious knowledge, phenomena that researchers across different disciplines consider as 'adultification'.

The high-stakes nature of the assessment under the current reforms has become the focal point of discourse within the academic and teaching community in both countries (Australian Primary Principals Association, 2009, 2010; Klenowski, 2011; Ravitch, 2010; Supovitz, 2009). Among this community, it is generally believed that social pressure for both teachers and students has serious impacts on the very survival or reputation of schools and teachers, and on the success of students (Australian Primary Principals Association, 2009, 2010; Lingard, 2010; Ravitch, 2010). In addition, some scholars raise concerns about the true diagnostic value of using Standard English as a benchmark for populations who use non-standard varieties of English (Wigglesworth et al., 2011), which is particularly relevant for the Australian Indigenous population, and for English language learners in both countries.

Despite controversies, the OECD considered the Australian NAPLAN 'well-conceived' and 'coherent' (OECD, 2013). In fact, the OECD report noted NAPLAN's potential, but stated that this is impeded by "less clear articulation on how it can generate improvements in classroom practice" (p. 12). In the US, supporters of student inclusion and education accountability are keen to point out that, until the implementation of accountability reforms, schools systematically excluded students with special needs from assessments due to low expectations (Hehir, 2005). The accountability provision in the US addressed the inclusion issue by setting assessment-inclusion guidance. US schools are expected to disaggregate student

performance by every demographic subgroup, including ethnicity, special needs such as disability, ELL, and SES background. In contrast, NAPLAN does not offer the same fine-grained disaggregation. Instead, it distinguishes students by only two characteristics: Indigenous heritage and language background other than English (LBOTE). The former category includes both Torres Strait Islander and Aboriginal students; the latter concerns the population of students whose family's first language is not English, even if the student's first language is English. Scholars have labelled the LBOTE disaggregation design as 'misrecognition' (Lingard et al., 2012) and 'ill-conceived' (Dixon & Angelo, 2014). Under the current disaggregation design, neither the 'Indigenous' nor the 'LBOTE' category helps further educators' understanding of the needs of these student subgroups.

Data, or numbers are central to the concept of accountability, whether it is being used within the context of the market, for managerialism, or for performance measurement (Lingard et al., 2012). These numbers are derived directly from standardised assessment. Data provide a common thread linking all three pillars of accountability. For external accountability, the hierarchy of the education apparatus relies on comparative data to evaluate school performance and to make funding decisions. For market accountability, parents and the greater community use the same data to evaluate neighbourhood schools and to inform enrolment decisions. Finally, the success of internal accountability relies on data to monitor instruction and progress against the standards and goals. Data-driven practice, especially at the school level, is seen as the future of school practice (Mandinach & Jackson, 2012). Opponents of data-driven accountability predict negative responses to this approach, while proponents tout it as the systemic change necessary to raise

achievement. Chapter 3 presents empirical evidence supporting both sides of the discourse.

## **Conclusion**

The Australian and the US contemporary education reforms signal a trend towards federal government involvement in school education (Ball, 2012). This trend emphasises outcomes and accountability through the principles of market competition and transparency (Ball, 2012; Hartley, 2003; Tolofari, 2005). At the heart of the opposition to accountability-based reform is what many scholars and practitioners see as the federal government's intrusion into state matters, and this perception of intrusion into local affairs makes it challenging to gain implementation buy-in (Cranston et al., 2010; Irons & Harris, 2007). Although Australia imported elements of its reform from the US, its implementation differs in three key aspects: partnership with state government versus federal-centric policy making; common versus disparate academic standards; and low versus high stakes. The Australian reform adopts a supportive and gentler approach to realise the desired long-term outcomes of achievement growth for all. In contrast, the US NCLB placed less emphasis on the structure and foundations needed for attaining the goals of closing the achievement gap, and put more emphasis on punitive measures and pressures to effect change. While NCLB's recent reauthorisation has transferred management of school accountability to state governments, school accountability requirements remain. The present research explores whether these implementation differences between the two countries affect schools' and teachers' interpretations of policy goals, and ultimately their readiness to use the key ingredient, data, to inform and

align classroom practices. Has federal policy changed school and teacher behaviours to improve the learning experience and achievement, particularly of those students that these policies intend to help – the disadvantaged? Specifically, has the “tremendous faith in the power of data – especially standardised test data – to effect school performance” (Young, 2006, p. 521) motivated and changed the way schools set internal policies and the way teachers deliver curricula and set expectations for students? Chapter 3 reviews empirical data on the impact of data-driven accountability reform to date on student achievement and teacher data engagement.

## Chapter 3 The Impact of Test-Based Accountability Reforms

The effectiveness of test-driven education transparency and accountability is the focus of a large body of research in both the UK and the US, where there is a longer history of test-based accountability than in Australia. A handful of Australian surveys (Athanasou, 2010; Dulfer et al., 2012; Pierce & Chick, 2011; Thompson, 2013) have begun to emerge in recent years exploring educator and student attitudes and experiences regarding NAPLAN testing. However, to date, there is a paucity of empirical data investigating the academic impact of test-based accountability reform in Australia. This may in part be explained by researchers' initial lack of access to NAPLAN data, which has only slowly become available since the Australian Senate's recommendation in the first Senate Inquiry into NAPLAN testing (The Senate Standing Committee on Education and Employment, 2010). This chapter presents empirical findings on the impact of test-based accountability on four areas: (1) assessment inclusion; (2) student achievement; (3) student learning experience; and (4) teacher data engagement. Through existing empirical evidence, it reviews whether the focus on external and market accountability has delivered the desired progress in overall student achievement, particularly among disadvantaged students. Specific to the element of internal accountability, it examines whether each country's respective policy has motivated and encouraged school leaders and teachers to utilise the massive amount of available assessment data to inform their practice and how data are being used.

Available research on the impact of accountability reform appears to fall into two phases, as evidenced by the US experience. In the initial years following



implementation of reform in the US, a large volume of research focused on the impact of standardised testing and the transparency requirement on school and teacher integrity. Specifically, investigations on all activities related to what critics called the 'unintended consequences' and discourse were dominated by consideration of factors such as test exclusion, teaching to the test, narrowing of the curriculum, and students' and teachers' affective states. In recent years in the US, as the concepts of accountability and transparency became the norm, public and academic commentary and research, have shifted to an examination of outcomes-related impact. Research coverage now includes: the effects of accountability on student achievement, curriculum and instruction alignment, assessment design, teacher training, and data-engagement and data-driven decisions, to name but a few. Research in Australia appears to be following a similar pattern, with more currently available research focusing on education experiences and teacher integrity, while outcomes related research has only just begun to surface.

### **Needs and Challenges of Disadvantaged Students**

Disadvantaged students are identified for this literature review, and for the present research, as comprising those listed in Table 2.2. This is consistent with the identification scheme in OECD (2003) and government (Rorris et al., 2011) reports. The needs and challenges of disadvantaged students vary. A student can be disadvantaged based on his/her economic background, race, physical or cognitive ability, general learning difficulty, linguistics needs, or multiple needs. The diverse education needs of this student population are not only distinct from those of general students but also from each other (Cimera, 2007; R. S. Johnson, 2002;

Westwood, 2008). For example, one study found that Indigenous Australians are not only racially disadvantaged, but also linguistically and socioeconomically disadvantaged (Bradley et al., 2007). This holds true also for immigrant groups such as Hispanic Americans, and other migrant groups across Europe (Schnepf, 2007). Some needs, such as physical disabilities are permanent while other needs, such as English language learning or learning difficulty can change. Depending on the “interaction between the student and the educational context” (OECD, 2007, p. 3), the latter set of needs can either be reduced or exacerbated. Therefore, improving the educational context could alter academic achievement (Cimera, 2007; Farrell, 2004, p. 20; *Opening All Options*, n.d.); for example, setting high expectations, providing effective interventions, or allowing accommodations. On the other hand, socioeconomic disadvantage, considered the primary determinant of student outcomes, brings cascading effects on multiple educational experiences including: low-funded schools; lack of resources; scarcity of infrastructure and qualified teachers; or an absence of role models (Darling-Hammond, 2010; Oakes et al., 2004; Schnepf, 2007). Here, what is known as the ‘achievement gap’ in fact derives from socioeconomic status, school characteristics and language skills (Schnepf, 2007). These negative factors give rise to low expectations, low-ability grouping, and a reduced curriculum – practices that were found to be common within low-performing schools (Cimera, 2007; R. S. Johnson, 2002; National Council of Teachers of English, 2008). It is particularly unsettling that these practices are often based on students’ backgrounds rather than their actual ability (Cimera, 2007; National Council of Teachers of English, 2008). These school practices perpetuate disparities in school and student achievement, higher-education access and attainment, and in

employment and income (Bradley et al., 2007; Darling-Hammond, 2010; R. S. Johnson, 2002; Oakes et al., 2004; Schnepf, 2007). The OECD (2012a) concluded that fewer than one in five disadvantaged students ever achieve the basic skills required to function in today's societies.

### **Equal Access to External Assessment – Goal or Reality?**

As discussed in Chapter 2, a key goal of both the US and Australian reforms is to shine more light on the achievement of disadvantaged students as a means to achieve education parity for all. In the US, NCLB required schools to include all students in external assessment and to disaggregate student performance by ethnicity, economic status, English background and disability background (Eckes & Swando, 2009). These steps were intended to ensure that (1) all students are truly accounted for; and (2) disadvantaged students are held to the same curriculum standards as other students (Cole, 2006). Under NCLB, meeting the AYP required testing 95% of every student subgroup and meeting proficiency goals for each group. The intention was that schools should not be able to “hide’ the low performance of any particular group of students” (Eckes & Swando, 2009, p. 2480). Cole (2006) noted that, “many parents, advocates, and educators have touted NCLB as the most significant piece of legislation that affects the education of students with disabilities since the passage of the first IDEA legislation in 1975” (p. 2), because the disaggregation requirement placed students with disabilities on the same playing field as the others. In a survey among 282 administrators and special education directors in the state of Indiana, over three-quarters of respondents agreed that the

legislation had raised expectations and had resulted in students with disabilities being held to the same standards as other students (Cole, 2006).

However, actual practice in the US paints a different picture. The same survey (Cole, 2006) also revealed that over 50% of all participants disagreed that the inclusion of students with disabilities in the assessment and accountability reporting was a benefit of the NCLB legislation. While this survey measured attitude, and not practice, attitudes influence intentions, which in turn drive behaviour, (a concept discussed in greater detail in Chapter 4). Evidence of inclusion violations has been widely documented in the literature (Booher-Jennings, 2005; Heilig & Darling-Hammond, 2008; Nichols & Berliner, 2005). Heilig and Darling-Hammond's (2008) analysis using seven years of achievement data for over 250,000 students in Texas, suggested there was systematic gaming with enrolment numbers to muddle the actual rate of test inclusion, achievement and graduation. Among the evidence is, "the large number of disappearances of students from the data set with no codes, most withdrawals appear to be dropouts" (p. 106).

Researchers (Cole, 2006; Goertz, 2005) believe the stringent requirements of NCLB, no matter how well intended, were responsible for operational challenges in assessment inclusion. The requirement for schools to meet AYP targets for every student subgroup resulted in many schools being placed on watch as well as feeling the pressure to deliver good results. Cole (2006) reported that, in 2005, 76% of schools in the State of Indiana failed to meet AYP in the special education subgroup. Goertz (2005) noticed similar challenge in other states in the early years of NCLB. Evaluating 2004–2005 school year data from California, Texas and Florida, Eckes and Swando (2009) found that the performance and participation metrics for students

with disability contributed to a large proportion of the schools' failing AYP status – it explained 17% of the schools' failing AYP status in California; 32% in Texas; and 28% in Florida. These rates exceeded those of other low-performing subgroups, such as students from low-SES background or African American students, by at least two-fold. Eckes and Swando further concluded that, in California and Texas, schools that included results from the students with disability were 72% and 80% less likely, respectively, to meet AYP than schools that did not include this subgroup.

It is therefore not surprising that most educators felt this particular NCLB requirement, while laudable, posed serious implementation challenges for schools (Center on Education Policy, 2005). The inability of schools to meet AYP resulted in multiple unintended consequences, discussed in the latter part of this chapter, and raised questions about the suitability of external assessment for all students (Goertz, 2005). As discussed earlier, some students have a permanent disability such as a severe cognitive processing disorder, that makes requiring them to participate in the general external assessment unrealistic (Goertz, 2005). Holding them to the same proficiency standards as for the average student is also inappropriate (Center on Education Policy, 2005; Goertz, 2005). Similar challenges were also raised about ELLs and the validity and reliability of testing them in a language they do not understand (Center on Education Policy, 2005).

Responding to some of these challenges, in 2005, the US Department of Education relaxed requirements for students with disability and for ELLs in accountability reporting (O'Day, Elledge, Le Floch, Taylor, & Anderson, 2009). Specifically, schools could offer alternative and modified assessments to students who needed them. However, only 2% of alternative or modified assessments that

met proficiency standards could be counted towards a school's AYP calculation (Irons & Harris, 2007; O'Day et al., 2009). These changes brought about a steady increase in state assessment inclusion (Irons & Harris, 2007). In an empirical evaluation of AYP data from three big states, Eckes and Swando (2009) observed an upward trend in AYP and proficiency target for every subgroup between 2001 and 2006. By 2006, more than 80% of the states tested 90% of students with disabilities in either the general or modified assessments (O'Day et al., 2009). Inclusion on the national benchmark assessment (NAEP) also improved, albeit at a slower pace as inclusion is encouraged, but not mandated for this assessment. Between 2005 and 2009, inclusion of students with disabilities on the NAEP mathematics assessment increased from 82% to 85% for fourth graders and 77% to 79% for eighth graders (Kitmitto, 2011).

In the same year, the US Department of Education also introduced the *safe harbor* provision intended to forgive a school for below-AYP test scores from one or more subgroups, if those students and the school's overall scores demonstrate yearly progress (Spelling, 2005). This provision allowed schools to avoid being placed in 'program improvement' status as long as progress is demonstrated. By providing a "more sensible and workable path for implementing NCLB" (Spelling, 2005, p. 4), law-makers attempted to address NCLB's inflexibility and responded to the challenge of expecting every student subgroup to meet the same goal despite persistent inequity.

In comparison, Australia's accountability reform has not afforded the same attention to all disadvantaged student groups. The subgroup disaggregation is provided only for Indigenous students and for students whose family language

background is other than English (ACARA, n.d.-b). Unlike the US standardised assessment, participation in the Australian NAPLAN is, by design, not compulsory and parents can withdraw their children if they wish. Since the launch of NAPLAN, inclusion patterns have reversed (Figure 3.1) as more students are being withdrawn from the test. If the rate of *absentee* is combined with the rate of *withdrawal*, the non-participation rate becomes even higher.

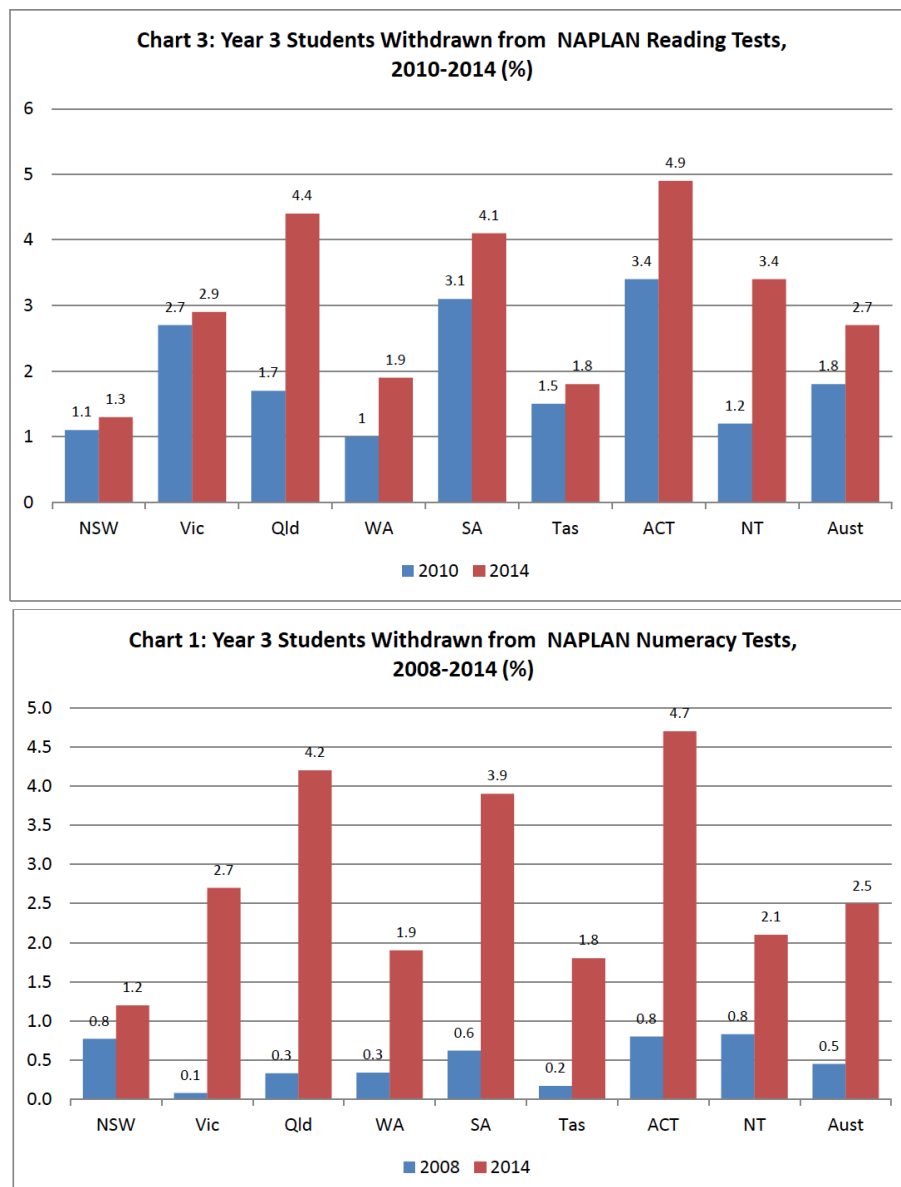


Figure 3.1. Year 3 Literacy and Numeracy Withdrawal Rate

Reproduced from Ray Adams, “Modelling the effects of participation on achievement in NAPLAN testing”, COAG Reform Council 2012 cited in Cobbold, T. (2012).

The 2014 NAPLAN summary (ACARA, 2014) showed that the national non-participation rate (withdrawal and absentee) between 2010 and 2014 jumped from 5.7% to 7.2% in reading and 5.9% to 6.7% in numeracy among Year 3 students. Year 5 students followed similar patterns. Between 2010 and 2014, non-participation in reading moved from 5.5% to 6.4% and in numeracy from 5.9% to 6.8%. More drastic are the non-participation rates of Indigenous students, which doubled the aforementioned national averages in both year levels and calendar years (ACARA, 2014).

The Council of Australian Governments' (COAG) Reform Council has validated the trend in non-participation, particularly for secondary students and students who generally score lower on the assessment (COAG Reform Council, 2013). Allegations of the deliberate exclusion of students with disabilities, in particular, from the test in an attempt to lower the school average were presented to the Australian Senate Inquiry Committee. These allegations led the Committee to recommend that actions be taken to prevent discrimination against students with special needs (The Senate Standing Committee on Education and Employment, 2010).

It is even more disconcerting that existing NAPLAN non-participation data do not provide information about who is being excluded. The current NAPLAN inclusion and outcome publication of the LBOTE and Indigenous categories is only available in the annual NAPLAN summary report, and the information is limited to the jurisdiction, not the school level. The individual view of each school's data on *My School*, the online tool, provides only the school average non-participation rate. Multiple disadvantaged subgroups are also missing both in enrolment numbers and in performance results on *My School*: students with disabilities, English language



learners, and students with low-SES status. It is no accident that a review (Forlin, Chambers, Loreman, Deppeler, & Sharma, 2013) commissioned by the Australian Research Alliance for Children and Youth, concluded that, “students with disability are currently under-represented in national and state testing and accountability measures” (p. 28). Some scholars (Elliott, Davies, & Kettler, 2012) believe that this lack of accurate benchmarking for students with disabilities violates the Australian Disability Standards for Education. Others feel that the lack of representation on NAPLAN “could easily give the impression that these students do not exist in the education system” (Dempsey & Davies, 2013, p. 9) as “they are out of the game” given the discriminatory treatment (Elliott et al., 2012, p. 8). In an effort to unearth the unknown, Dempsey and Davies (2013) used the Longitudinal Study of Australian Children to estimate the NAPLAN participation rate of this student subgroup. The authors concluded that more than a third did not participate in NAPLAN. This proportion is significantly higher than the national average participation rate.

Practitioners and scholars have both expressed concerns in the Senate Inquiry (The Senate Standing Committee on Education and Employment, 2013). They believe that the lack of a narrower disaggregation scheme could negatively impact resource allocations to schools whose overall positive performance score may mask the needs of some students. Even between the two categories that are disaggregated on *My School*, neither LBOTE nor the Indigenous data are deemed reliable, because they do not accurately reflect the needs of these groups (Creagh, 2014; Dixon & Angelo, 2014; Wigglesworth et al., 2011). In a survey (Dixon & Angelo, 2014) of 86 schools in Queensland, only two reported positively that Indigenous student language background was accurately identified in the school systems. Dixon

and Angelo (2014) and other researchers (Creagh, 2014; Lingard et al., 2012; Wigglesworth et al., 2011) all contend that the language needs of Torres Strait Islanders, Aboriginals and ELLs from different countries vary significantly, and the current generic categorisation scheme can be best characterised as “a pervasive blindness about all languages” (Dixon & Angelo, 2014, p. 220). As such, the disaggregated data provide neither explanatory nor diagnostic value (Dixon & Angelo, 2014; Lingard et al., 2012). The Multicultural Development Association (2010) raises the same concern in their submissions to the Australian Senate Inquiry, noting the diverse characteristics and needs of students categorised as LBOTE.

Judging from the current disaggregation effort, or lack thereof, in the official government testing data, it appears that the Australian government’s reform effort is more concerned with achieving a top-five performance ranking for Australia on the PISA than with raising the performance of every Australian student, regardless of their background. As multiple experts and scholars have warned, the school system cannot support what it does not know, and currently, the Australian education system knows very little about many disadvantaged student subgroups (Elliott et al., 2012; Forlin et al., 2013).

### **Effects on Student Performance and the Achievement Gap**

In the US, despite being far from the 100% proficiency goal, a positive, albeit modest, picture of student outcome has emerged over the past decade. The outcomes, however, have not been without issues. The most significant challenge is that the outcomes are not generalisable. The challenge stems from the lack of a national assessment equivalent to the Australian NAPLAN. Every state has its own

version of external assessment. Hence, one state's 'proficiency' definition could be another state's 'fail' rate. However, since 2009, state governments have collaborated to create the *Common Core Standards* for the core subjects (Common Core State Standards Initiative, n.d.-a). To date, 42 States have adopted this standard and common assessment was launched in 2014. Until the Common Core assessment becomes more entrenched, the NAEP serves as the national benchmark assessment. The NAEP is administered every two years to a sample population nation-wide and is limited to only Grades 4, 8, and 12. Researchers have used both state assessment data and NAEP data to evaluate student progress. The literature review of the US context covers literature from both strands.

To date, there is a dearth of independent research in Australia investigating the empirical impact of test-based accountability on academic outcomes. At the present time, the annual NAPLAN National Report published by ACARA is the main source of information about student outcomes. A very small number of large-scale analyses have come from ACER (Ainley & Gebhardt, 2103) and the government-sponsored COAG (2013). As noted earlier, independent research evaluating the impact of external testing on student and educator well-being has emerged since the turn of the decade. However, research on student progress post-NAPLAN is scarce. The following two sections discuss achievement progress for literacy and mathematics, as reported in the literature, since the implementation of NCLB and the Melbourne Declaration. The discussion focuses on studies that used large-scale data as a means to evaluate progress.

**Academic outcomes of disadvantaged students in the US**

A comprehensive review and analysis by the Center on Education Policy in the US combining large-scale annual surveys and case studies over four years (Jennings & Stark Rentner, 2006) concluded that nearly 75% of the states and school districts saw increases in reading and mathematics proficiency as measured by state-level standardised tests. It also reported a narrowing of the achievement gap between advantaged and disadvantaged students, as measured by state external assessments. This finding corroborated trends reported by NCES (National Center for Education Statistics, 2013a). Between 1990 and 2013, reading and mathematics for Grades 4 and 8 Black and Hispanic students achieved statistically greater gains than the national student average. During this period, national average scores in mathematics for Grades 4 and 8 rose by 28 and 22 points, respectively. Increases in reading were more modest at only five and eight points. However, as shown in Table 3.1, Black and Hispanic students (who traditionally score lower than White students) achieved larger gains, albeit lower than gains achieved by Asian students. In comparison, students with disability, those from a low-SES background, and ELLs have made less progress. In mathematics, students with disabilities achieved 50% less than the other groups.

Dee and Jacob's study (2011) using NAEP data supported NCES' findings regarding the mathematics progress of Hispanic and Black students. Other researchers (Carnoy & Loeb, 2002; Hanushek & Raymond, 2005) who used NAEP data to evaluate cross-state progress also observed a positive correlation between high-stakes accountability and student improvement on the NAEP in mathematics. It is important to highlight that overall progress did not come at the expense of the

high-performing students, as critics had predicted (Dee & Jacob, 2011; National Center for Education Statistics, 2013a). Both studies showed comparable score increases at every level of the achievement distribution.

Table 3.1

*Score Changes on the NAEP (1990 and 2013)*

	<i>Mathematics</i>		<i>Reading</i>	
	<i>Grade 4</i>	<i>Grade 8</i>	<i>Grade 4</i>	<i>Grade 8</i>
All students	28	22	5	8
White	30	24	8	9
Black	37	26	14	13
Hispanic	30	26	10	15
Asian	33	31	19	12
With disability	15	18	-- <sup>2</sup>	8
ELL	18	20	13	7
Low-SES <sup>1</sup>	-- <sup>2</sup>	25	-- <sup>2</sup>	8

*Note.* <sup>1</sup>Students whose parents have not graduated high school.

<sup>2</sup>Data are not available.

(National Center for Education Statistics, 2013a)

Two studies also examined the impact of accountability expectations on achievement outcomes. One (Wong, Cook, & Steiner, 2009) categorised the proficiency standard requirement of each state into ‘high’, ‘medium’ and ‘low’, and compared their NAEP results pre- and post-NCLB. Their investigation reported a positive effect on reading associated with NCLB in Grade 4. The largest effects were associated with states that had implemented ‘high’ standards and those with sanctions that were both rewarding and punitive in response to NCLB. For mathematics, the study not only found significant gains but that states in the high proficiency standards category performed best, with the reverse being true for states with low proficiency standards.

Similarly, comparing states that had implemented an accountability policy prior to NCLB with those that had not, Dee and Jacob (2011) found compelling evidence supporting the positive impact of accountability on NAEP outcomes (Figure 3.2). Their data included 39 states for mathematics and 37 for reading at the fourth-grade level and showed that scores for all states grew at a substantial rate after NCLB. While the authors did not provide an explanation, it is interesting to note that those states without prior accountability system grew more after NCLB than those with a prior system. For reading, the growth curve was steeper overall. Both studies demonstrated a strong relationship between expectations (via proficiency standards or accountability measures) and outcomes.

An independent review from the coalition of the US large urban school systems (Casserly, 2007) identified another encouraging trend: a narrowing of achievement gaps on the NAEP among the largest central city school systems, that generally serve a disproportionately large population of disadvantaged students. This observation is supported by another study (Lauen & Gaddis, 2012) using eight years' longitudinal data from the State of North Carolina to evaluate the impact of accountability pressure. Lauen and Gaddis concluded that subgroup-specific accountability pressure has had positive effects on Black and Hispanic students, and on students from a low-SES background. More tellingly, the effects on the Hispanic subgroups were most significant at the lowest-performing schools. Furthermore, the largest effects were evident in schools furthest from the AYP benchmark, lending credence to the impact of accountability pressure. On the NAEP, Grade 4 White–Black score gaps in mathematics and reading narrowed somewhat, from an over 30 point consistent average before 2002 to a below 30 point average since 2003

(National Center for Education Statistics, 2013a). Movement in the Grade 4 White–Hispanic gap followed that of the White–Black gap in reading but was negligible in mathematics.

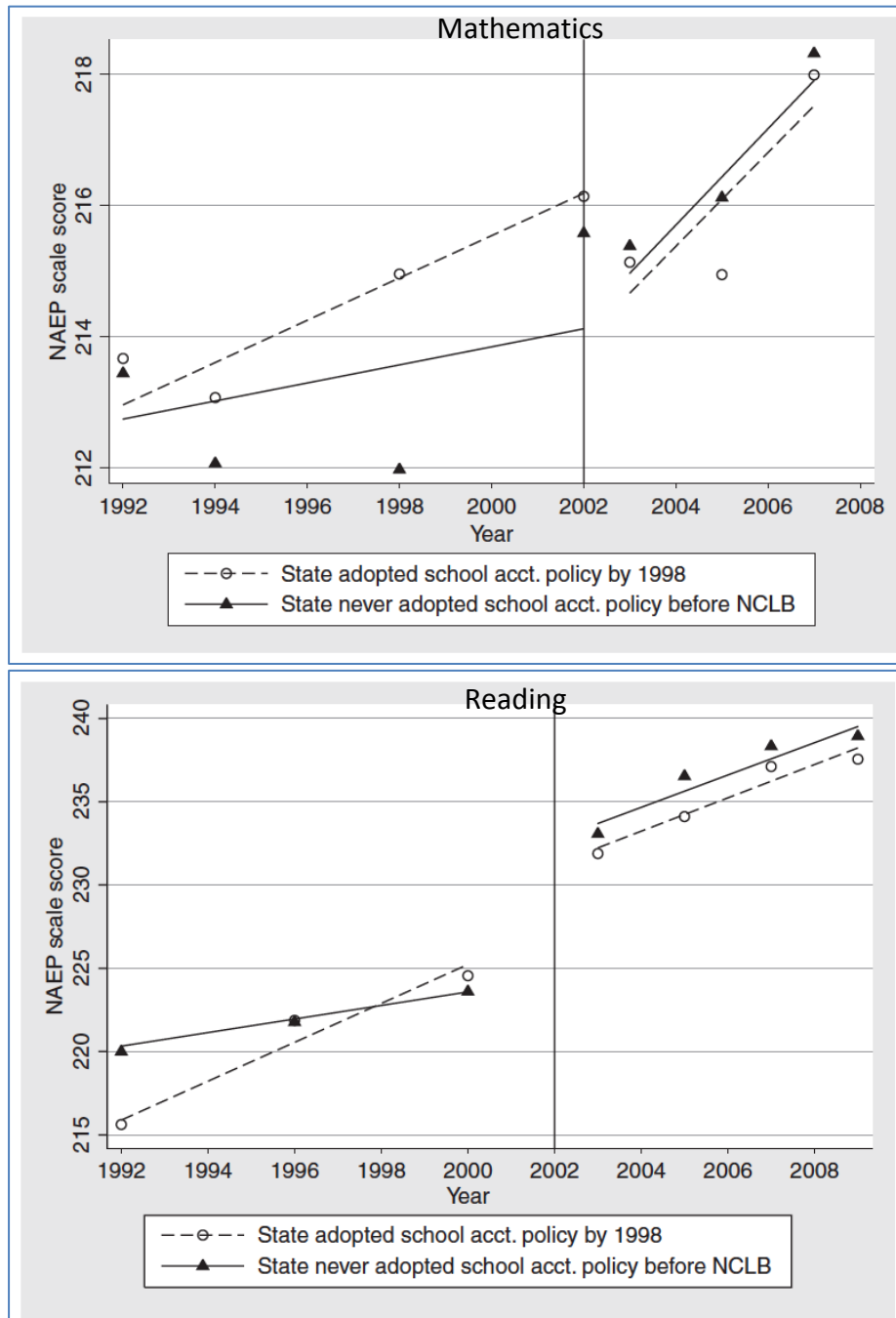


Figure 3.2. Impact of NCLB on NAEP Outcomes in Grade 4

Reproduced from “The impact of No Child Left Behind on student achievement” by T. S. Dee and B. Jacob, 2011, *Journal of Policy Analysis and Management*, 30(3), pp. 418–446.

However, for every study that demonstrates positive academic outcomes in the US since NCLB, there is a study that presents contrasting evidence. One such (J. Lee, 2006) compared NAEP results before and after the implementation of NCLB and concluded that the basic trends in achievement gains were similar in both periods. Namely, modest gains in mathematics, flat achievement in reading, and sustained achievement gaps. Another study (Fuller, Wright, Gesicki, & Kang, 2007), using similar data, corroborated the relatively flat growth trend. Lee's (2008) meta-analysis conducted two years later produced a similar conclusion. Half the studies in the analysis indicated positive improvement; half did not, leaving a modest effect at the macro level. However, Lee's most recent study (J. Lee & Reeves, 2012) analysing NAEP data from 1990–2009 concurred with the positive achievement growth presented by the NCES (2013a) particularly in mathematics, before and after NCLB. Lee and Reeves went one step further to investigate the state characteristics and trends responsible for the gains. Their investigation returned the following findings: (1) progress made by the states was not systematic across grades, subjects, and student subgroups, hence progress is neither sustainable nor generalisable; (2) long-term state-wide efforts in instruction, capacity and teaching resources had more direct impact on progress than short-term NCLB-related implementations, such as raising state standards or using data to inform practice. Another study using Florida's state test data concluded that NCLB had no effect on Black, Hispanic and poor student test scores (Figlio, Rouse, & Schlosser, 2009).

Despite conflicting findings on student achievement, US state and district officials credited the accountability requirement with the positive impact on progress in reading and mathematics. Educators particularly noted NCLB's benefits



for the disadvantaged population. National school surveys conducted on a regular basis by the Center on Education Policy (Jennings & Stark Rentner, 2006) showed that school administrators, “have consistently praised NCLB’s requirement for disaggregation of test data by subgroups of students, because it has shone a light on the poor performance of students who would have gone unnoticed if only general test data were considered” (p. 112). Nonetheless, the same surveys also highlighted that, while state and district officials saw value in reporting the AYP and in disaggregating results for disadvantaged students, they deemed it an unfair, cumbersome and challenging process for schools. Some administrators also protested against the use of disaggregated data as a means to evaluate instructional efforts at schools, or student efforts, particularly in relation to students with disabilities and to ELLs. Others considered the standardised exams inappropriate for disadvantaged students and impractical for teachers, as they lend no instructional value (Jennings & Stark Rentner, 2006). Therefore, policy makers still have a long way to go in designing the appropriate instrument and process to support both students and teachers in closing the achievement gaps.

#### **Academic outcomes of disadvantaged students in Australia.**

The 2014 NAPLAN National Report (ACARA, 2014) indicated statistically positive progress in reading in Years 3 and Year 5. As shown in Table 3.2, mean scale scores increased by 18 and 16 points in Years 3 and 5, respectively, between 2008 and 2014. The increase from year to year was small but steady in Year 3, while more erratic in Year 5. The increase in reading scale scores over time is consistent across jurisdictions, with Queensland (QLD), Western Australia (WA) and the Australian Capital Territory (ACT) making statistically significant change across both Years 3 and

5. In Year 5, Tasmania (TAS) also made a substantial increase. Years 3 and 5 Indigenous students also made a statistically significant gain of 19 points across both years, but the same cannot be said for LBOTE students (Table 3.2). While performance of low-SES students was not reported by ACARA, a COAG report (2013) cited statistically significant gains in both years for this subgroup as well. On aggregate, the six-year cumulative NAPLAN reading growth trend is positive, albeit not consistent across jurisdictions and year levels (ACARA, 2014). An ACER report using NAPLAN data between 2008 and 2012 supported these findings (Ainley & Gebhardt, 2103). Another positive trend is movement within the lower and upper bands. Ainley and Gebhardt noted fewer Year 3 students in the lower band (18% to 14%) and more students in the upper bands (18% to 26%) between 2008 and 2012.

Table 3.2

*NAPLAN Score and Minimum Proficiency Percentage Change between 2008 and 2014*

		<i>Reading</i>		<i>Numeracy</i>	
		<i>Year 3</i>	<i>Year 5</i>	<i>Year 3</i>	<i>Year 5</i>
All students	Scale score	<b>17.8</b>	<b>16.2</b>	4.9	11.7
	Proficiency %	1.4%	1.9%	3.6%	8%
Indigenous	Scale score	<b>19.2</b>	<b>18.7</b>	3.9	9.9
	Proficiency %	6.4%	6.9%	-0.4%	3.5%
LBOTE	Scale score	14.8	14.6	1.6	8
	Proficiency %	1.3%	2.7%	-0.2%	0.9%

*Note:* Bold denotes statistically significant increase.  
(ACARA, 2014)

In contrast to reading, numeracy in Years 3 and 5 demonstrated no statistically significant progress between 2008 and 2014 (ACARA, 2014). This is true at the national level, and by Indigenous and LBOTE status (Table 3.2). Across jurisdictions, only QLD recorded statistically significant growth in both score and proficiency measures across both year levels. While most groups in Year 5 had

slightly larger absolute scale score gains than those in Year 3, which were mainly in the single-digit or negative range, none were statistically significant except for those achieved by WA, ACT and Western Australia (WA). Other reports (Ainley & Gebhardt, 2103; COAG Reform Council, 2013) demonstrated generally similar results between Years 3 and 5. On the whole, performance in numeracy has not increased over the six years since the implementation of NAPLAN.

Overall progress in reading notwithstanding, the NAPLAN National Report (ACARA, 2014) also showed that an increase in scale score did not translate to statistically significant change for the percentage of students at or above minimum standards (QLD was the only exception). Nationally, 94% and 93%, respectively, for Years 3 and 5 met minimum standards in 2014; this was only true for 75% of Indigenous students. More alarmingly, a COAG report (2013) concluded that attendance rate did not improve among the Indigenous population between 2008 and 2012. Among students from a low-SES background, only 78% in 2012 reached the minimum standard (COAG Reform Council, 2013). A reading scale score gap of 90 points existed in 2008 between Indigenous and non-Indigenous students and this gap persisted in 2014 in Year 3. Similar gaps of 86 points in 2008 and 83 points in 2014 were noted in Year 5. The numeracy gaps between these same students dropped slightly to 72–74 points across both years but remained large.

At a 10-point difference in scale score, the smallest achievement gap is that between students of non-English and English-language background of both year levels. However, the small LBOTE gap might not paint the real picture. Lingard, Creagh and Vass (2012) and Creagh (2014) warn against reading too much into this result because the LBOTE category is not differentiated enough to add value. They

argue that the category includes a wide range of children, from those whose parents speak a language other than English (even when the children are English speakers), to recent traumatised refugees arrived from a war-torn country who have never been exposed to English (Multicultural Development Association (MDA), 2010). Lingard and colleagues (2012) declare that the vast disparity in English language skills renders the LBOTE score invalid and unreliable as a diagnostic tool. These arguments reinforced by the fact that 53–62% of students accepted by selected NSW high schools between 2007 to 2011 were students of LBOTE background (NSW Department of Education and Communities, n.d.-b); selection to these schools is informed by placement tests and other evidence of high academic ability.

Since students with disability are invisible on the NAPLAN, no insights can be drawn from the national report. However, two empirical studies (Australian Bureau of Statistics, 2014; Dempsey & Davies, 2013) offer some information. Combining census data with the 2011 NAPLAN results for Years 3, 5, 7 and, 9 for Queensland, the Australian Bureau of Statistics found that one in three students with a disability scored below the national minimum standard in writing. In reading, close to one in four students did not meet standards. In numeracy, the number was close to one in five. Using the Longitudinal Study of Australian Children and 2008–2009 NAPLAN data, Dempsey and Davies (2013) gleaned a significantly lower mean score across all domains for students with disability in Year 3 compared to students without disability.

Furthermore, there is evidence that the 2011 Queensland NAPLAN results were negatively affected by a host of socioeconomic factors, including: low parental education background, young age of mother, single-parent and foster families, low

household income, and parental employment status (Australian Bureau of Statistics, 2014). A separate study (Miller & Voon, 2012), also using NAPLAN data from 2008–2009, further found that non-government schools, particularly independent schools, consistently outperformed government schools across all years and all domains. While ACARA does not identify government schools as economically disadvantaged and independent schools as economically advantaged, these authors pointed out in their literature review that government and non-government schools in Australia have been proven to segregate students along social and academic lines (Lamb et al 2004, cited in Miller & Voon, 2012).

In summary, while average reading outcomes from 2008 to 2014 for Year 3 and Year 5 students have exhibited a modest rise, numeracy outcomes have not. Outcomes for the Indigenous population are also not encouraging and outcomes for other disadvantaged groups are completely unknown, or must be derived by proxy from other data. At present, it is reasonable to conclude that, despite the policy intentions, Australian test-based accountability reform has not provided sufficient attention, or even a 'fair go' (Davies, 2012), to a variety of disadvantaged students.

### **Unintended Consequences of Test-Based Accountability**

Progress in achievement is juxtaposed against mounting evidence supporting a variety of negative consequences of test-based accountability, especially in the US in the early years post-NCLB. By comparison with the modest amount of research evaluating achievement outcomes, there is a large body of work that offers evidence of negative practice. This includes research from influential opponents (Nichols & Berliner, 2005; Wyn, Turnbull, & Grimshaw, 2014). A review of the literature

suggests that the negative consequences outlined in the rest of this section apply in both countries.

### **Cheating to inflate school performance.**

A negative consequence reported in multiple studies concerns the way schools manipulated various data to raise overall school performance. These findings suggest the active creation of proficiency illusions by educators at various levels and by various means: excluding low performers from standardised exams (Dulfer et al., 2012; Kitmitto & Bandeira de Mello, 2008; Sawchuk, 2010, March 9); reclassifying students with special education needs into different categories (Heilig & Darling-Hammond, 2008); holding students back at grade level (Heilig & Darling-Hammond, 2008; Nichols & Berliner, 2005); and adjusting the state proficiency goal (J. Lee, 2010; Nichols & Berliner, 2005). Negative reactions to the test-based accountability system in the US were not only limited to researchers, policy analysts and practitioners; parents and students have joined a growing chorus of disapproval, for example by creating documentary (Gabriel, 2010, December 8) and by expressing their scepticism through the Gallup poll (Rose & Gallup, 2006). In the US, a cheating scandal that was considered the largest in the US education history, led to the 2015 high-profile conviction of 11 public school educators in Atlanta (Blinder, 2015, April 1).

### **Teaching to the test and narrowing the curriculum.**

In the US, because AYP only measures progress in literacy and mathematics, these are, by definition, the only subjects for which schools are accountable. This has led to claims about a narrowing of curriculum (Ravitch, 2010). A meta-analysis (Au, 2007) confirmed evidence of teachers narrowing the curriculum in the US to focus

their instruction on the test itself. Nichols and Berliner (2005) found 15 news reports citing cases of schools and teachers teaching to the test across the US. In addition to spending time preparing for the standardised test, a year-long qualitative study at four US elementary and middle schools observed that data analysis, reflection, and instructional alignment exercises at schools focused mainly on the assessment goals (Halverson, Grigg, Prichett, & Thomas, 2005). These authors also reported that social studies and science were outside the purview of the data-driven analysis process and received minimal program attention. In another study (Louis, Febey, & Schroeder, 2005) involving three high schools, one principal lamented the need to cover more topics in less depth in order to teach the standardised test content. Some policy analysts have claimed that test-based accountability has diluted the definition of schooling. Among them, Hess and Finn Jr. (2007) asserted that the compliance nature of the reform has been “accompanied by a notable creativity vacuum” (p. 312). Similarly, Siegel (2004) wondered whether the goal is to develop critical and creative thinkers, or workers “who [are] able to function successfully in the marketplace” (p. 62). Siegel further argued that the narrowness of test-based accountability measures only ‘minimal competence’ and will produce workers rather than thinkers.

#### **‘Education triage’.**

A quantitative (Krieg, 2008) and a qualitative (Booher-Jennings, 2005) study noted evidence of ‘education triage’, a term originated by Gillborn and Youdell (2000), which Booher-Jennings employed to describe a phenomena observed at a Texas school. In this school, additional attention was given to those students who were close to meeting the benchmark at the expense of those far below and far

above from that benchmark (Booher-Jennings, 2005). Nichols and Berliner (2005) found a large number of journalistic accounts substantiating this practice. According to Krieg (2011), education triage can have an impact on resource and attention allocation in several ways. For example, a principal may assign weak students to strong teachers and strong students to weak teachers. He or she may also adjust class sizes to support those students who are behind. A school may abandon a curriculum with broad appeal for one that helps to raise the skills required in the external assessment. Finally, it may also divert funding away from a low-stakes subject to English and mathematics.

The year-long data collected by Halverson et al. (2005) demonstrated that data discussions among teachers predominately focused on helping struggling students to reach proficiency level. Few discussions were concerned with extending students who already achieved proficiency. Rather than helping every student who was below the proficiency threshold, one of the schools further devoted time and resources to those students with a realistic chance of crossing that threshold. Using a state-wide sample of third- and fourth-grade mathematics data, Krieg (2011) presented evidence supporting the negative effect of education triage on students who have already achieved proficiency. Measuring the growth of proficient students in schools that might fail AYP against peers in schools that were likely to pass AYP, Krieg found that proficient students in the first set of schools scored lower on a subsequent mathematics test compared to their peers at the second set of schools. His earlier analysis (Krieg, 2008) of the same data documented a similar effect of lower gains among students at the higher end of the proficiency scale for students attending schools that were unlikely to meet AYP requirements. Krieg attributed



these findings to the practice of strategic instructional attention, where a teacher strategically allocates attention to students in an attempt to meet AYP.

### **Parallel Evidence in Australia**

Similar concerns have been raised in Australia through a number of qualitative and survey research. One online survey commissioned by the Independent Education Union of Australia (Athanasou, 2010) was administered to 269 teachers and 161 principals at Catholic and independent schools nation-wide. Another conducted by Thompson (2013) among 941 teachers across Western Australia and South Australia examined teachers' perceptions of NAPLAN. The largest study (Dulfer et al., 2012) involved 8,500 educators throughout Australia. The sheer number of the responses to Dulfer et al.'s survey is indicative of the heightened attention and controversy surrounding NAPLAN. The Whitlam Institute also sponsored a qualitative study (Wyn et al., 2014) with interviews among 45 educators, 26 parents and 60 students within 16 schools in Victoria and NSW. All studies explored the impact of NAPLAN on the following areas: testing, pedagogy, curriculum, student health and well-being, and pressures from *My School*. In addition to these studies, submissions to the Senate Inquiries regarding NAPLAN also shed light on NAPLAN's impact on schools, teachers, administrators and students. The Senate held these hearings in response to a chorus of warnings against, and media exposure of the unintended consequences already experienced in the US.

Findings from these studies reaffirmed many issues experienced in the US. For instance, test preparation prior to NAPLAN testing has taken up class time and diverted attention from the syllabus (Athanasou, 2010; Dulfer et al., 2012;

Thompson, 2013). Seventy three percent of teachers agreed that they have taught to the test (Dulfer et al., 2012). Although this practice was generally considered to be negative by respondents, some teachers and principals believed that NAPLAN's focus on reading and mathematics is vital for all aspects of education, including full access to the broader curriculum, to higher education, and to civic participation (Dulfer et al., 2012). Nonetheless, in the same survey, 83% of educators reported experiencing the crowding effect of NAPLAN on the breadth of the curriculum.

The intent for NAPLAN to be a diagnostic tool may have been hindered by the timing of data release (Athanasou, 2010; The Senate Standing Committee on Education and Employment, 2013). For this reason, NAPLAN and *My School* are regarded more as school evaluation or ranking tool by the majority of teachers (Dulfer et al., 2012; Thompson, 2013). School educators reported feeling pressured to raise student achievement (Athanasou, 2010; Dulfer et al., 2012; Thompson, 2013), and to retain students who might move to a better performing school (Athanasou, 2010). In addition, Wyn et al.'s study (2014) called out the overwhelming stress expressed by students who saw absolutely no value in the test. Close to 90% of teachers attested to having witnessed students experiencing stress prior to the exam (Dulfer et al., 2012). Teachers also felt that the stressful testing environment runs counter to what the reform hopes to achieve (Thompson, 2013). However, despite scepticism and concerns, a few parents could see some value in using external assessment to evaluate achievement and attainment (Wyn et al., 2014), and teachers and principals appreciated the idea of "being kept on [their] toes" (Athanasou, 2010, p. 17) as regards accountability.

One notable difference between the Australian and the US experience is the emphasis on the impact of data transparency and market accountability. In Dulfer et al.'s survey (2012), over 90% of educators felt that "lower than expected results on NAPLAN would mean that a school would have trouble attracting and retaining students" (p. 8). Educators in non-government schools articulated the same concern (Athanasou, 2010). A possible explanation could be that over 30% of Australian schools are independent schools, as compared to fewer than 10% in the US. To that end, Australian parents are more likely to exercise their market choice than US parents. For public school parents, the notion of geographical movement for school choice is less common. Since the majority of US students go to their neighbourhood schools, entertaining the idea of school choice is less common. A study (Holme & Wells, 2008) evaluating market accountability in the US concluded that, due to multiple logistical challenges, movement of students has been limited and parents have not considered the option a benefit.

Clearly, these practices reported in the literature resemble none of the productive assessment traits, which many scholars (Conley & Darling-Hammond, 2013; Lingard, Mills, & Hayes, 2006) advocate. Instead, they epitomise the unintentional consequences that many eminent scholars, researchers and experienced practitioners in both countries anticipated at the start of their respective accountability reforms.

### **Contradictory Findings**

Amidst research conveying the negative impact of NCLB in the US on students, teachers, schools and pedagogy, a small but growing body of literature also

found contradictory effects to these unintended consequences. ‘Education triage’, for example, is not observed in some quantitative studies (Lauen & Gaddis, 2012; Reback, 2008). Lauen and Gaddis observed few incidences of triage in North Carolina, where it only occurred after the mathematic proficiency standards were raised. The researchers noticed that the strategy was used to help students who had passed the lower proficiency benchmark but failed the more rigorous new benchmark. They argued that the stability of these standards over time would eliminate triage, because they found no evidence of triage prior to the change in standard or in reading, for which the standard remained constant over the same period. Lauen and Gaddis further concluded, “accountability-induced triage from NCLB is not an automatic consequence of a status-based approach to accountability but rather a risk factor along with the rigor of proficiency levels and the stability of these levels across time” (p. 203).

Other positive effects mentioned in state-wide studies and cross-state surveys include the increase in curriculum rigor and curriculum alignment. In addition to disputing the practice of triage, Lauen and Gaddis (2012) also found an increase in curriculum rigor. However, their data also indicated that, while rigor is important in the long run, it does have a large negative effect on the lowest achieving students, given the added gap between where they were and the higher proficiency standards. Another empirical study (Polikoff, 2012) cited changes in instructional alignment over a seven-year period after NCLB. Analysing the content of state standards and assessments against survey data of more than 27,000 teachers’ instruction in mathematics, literacy and science, Polikoff found evidence of increased alignment across all grades from K–12, and all subjects, particularly with

mathematics. Jennings and Stark Rentner's (2006) survey data corroborated this conclusion and noted that the alignment is more pronounced at schools that have not met AYP requirements for two successive years.

One recent study (Harr-Robins, Song, Garet, & Danielson, 2015) specifically investigated the impact of accountability requirements on the experience of students with disabilities across 12 states in the US. Through survey data, this study compared the experiences of students in schools that are always accountable for students with disabilities to schools that are not. It found that always-accountable primary schools were much more likely than their counterpart to: include students with disability in the general classroom; adopt new instructional programs; implement a tiered instructional intervention; and provide instructional and assistive technology to support students with disabilities. Teachers at always-accountable schools were also more likely to do team teaching and receive professional development and training. While these are encouraging practices spawned from the accountability requirement, the authors warned that the differences could be attributable to differences in school size and other characteristics, since non-accountable schools are by nature smaller.

### **The Promise of Data-Driven Practice and the Reality of Practice**

Be it for external, market or internal accountability, both positive and negative effects of accountability revolve around external testing, where narratives are created by data and numbers from the assessment (Lingard et al., 2012). It is unquestionable that assessment data are the hub that connects all aspects of accountability. As discussed in Chapter 1, despite evidence of unintended

consequences and implementation flaws, researchers and scholars welcome the potential benefits of data. Goldring and Berends (2009), the authors of *Leading with Data, Pathways to Improve Your School* suggest that data-based decision making can lead to continuous improvement through the process of organizational learning,

Organizational learning occurs when knowledge is distributed across individuals and is embedded in the culture, values and routines of the organization. This type of learning is a developmental process that can occur in an organization over time. (p. 15)

However, growing belief in the benefits of data-based decision making has not led to consistent and deep data engagement in many US schools (Wayman, Spring, Lemke, & Lehr, 2012). Nonetheless, evidence is emerging linking schools' high performance to school-wide data-driven practice and teachers' reliance on data to inform instructional decisions (Datnow et al., 2007). Other studies (Feldman & Tung, 2001; Halverson et al., 2005; Louis et al., 2005; Wayman, Midgley, & Stringfield, 2006) noted that teachers have become more data savvy, objective, reflective, and collaborative in their practices. These are all core elements of what Halverson and colleagues (2005) termed 'the new instructional leadership'. Two studies further found that some schools have created accountable learning systems where decisions are based purely on data (Feldman & Tung, 2001; Halverson et al., 2005). Together, these studies offer encouraging evidence that educators are also responding constructively to data-driven reform to make desirable changes in education.

Yet, amidst promising findings on data use, significant challenges persist both internal and external to schools and their faculties (Wayman, Jimerson, & Cho, 2012). The sections that follow examine the role of data in schools and in classrooms

since the enactment of these reforms, and the challenges inhibiting the adoption of data as part of pedagogical practice. It must be noted that studies investigating destructive data use, resulting in the unintended consequences discussed earlier, outnumber those that explore constructive data use. In Australia, with the exception of a pilot study (Pierce & Chick, 2011), systematic data-use investigation has not been an area of research attention. Pertaining to data practice to be reported in the following sections, evidence in Australian schools derived largely from Australian survey research; whereas evidence in the US context is gleaned from in-depth case studies and interviews specific to data-driven practice.

**Perceived trust and benefits in external data.**

The most commonly cited factors influencing the adoption of data-driven practice include: the lack of systematic use of data; teacher perception regarding data usefulness; and overall relevance of external data to teaching (Coburn & Talbert, 2006; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Monpas-Huber, 2010; Young & Kim, 2010). Leadership commitment to building a data-driven organisation also can expedite data-use adoption (Feldman & Tung, 2001; Young, 2006). The cultural conditions of a school, such as a history of data-driven decision-making, and the existence of an internal accountability system, also affect engagement with accountability data. Two studies (Elmore & Fuhrman, 2001; Halverson et al., 2007) indicated that teachers resisted external data accountability because they believed the data directly conflicted with their internal accountability system, which they valued more highly for teaching. When teachers fail to perceive data validity or added-value from external assessment data, they do whatever it takes to retain their existing internal accountability system, and are less keen to align instructional

practices with external assessment data (Kerr et al., 2006; Monpas-Huber, 2010; Young, 2006). Another study (Supovitz & Klein, 2003) at nine US schools supports the view that teachers unanimously value internally-developed assessments over standardised assessments. In this study, teachers rated their own metrics as 'highly useful' and external metrics as only 'useful'.

A longitudinal study at nine schools (Ingram, Louis, & Schroeder, 2004) documented the persistent use of 'personal metrics' by teachers. These metrics tend not to align with those used for external accountability. There are also teachers who still rely on anecdotal information, experience, and intuition (Ingram et al., 2004). Most worrisome among the findings in Ingram et al.'s study is the belief that teachers' responsibility is curriculum delivery rather than data collection and analysis. Halverson and colleagues (2007) further observed that individual student-focused interventions are still the dominant strategy teachers used in data analysis sessions to assist students with special education needs. While focusing on struggling students individually is not negative, Halverson et al. (2007) feared that individually-focused interventions can inhibit a systemic response and overhaul that could benefit more struggling students.

In Australia, NAPLAN's diagnostic value was openly debated at the 2013 Senate Hearing (The Senate Standing Committee on Education and Employment, 2013). Some of the evidence submitted to the hearing included a large-scale survey (Dulfer et al., 2012). This survey found that teachers viewed NAPLAN first and foremost as a school-ranking tool and only secondarily as a diagnostic tool. Fewer than half (42%) of the teachers agreed that NAPLAN has diagnostic value, and among that 42%, only 8% 'strongly' agreed. In a survey of independent school teachers



(Athanasou, 2010), between 47% to 58% of teachers from different year groups considered NAPLAN useful. Teachers' responses (about 55%) to a series of data-usage questions further suggest that NAPLAN results are currently used primarily as an overview of the academic status of the school, and hardly used as a tool to inform student needs or instructional needs (Dulfer et al., 2012). In the same survey, slightly under half reported that their schools were only beginning to figure out ways to implement data use, including forming year- or subject-level teams to analyse data.

Findings from a small pilot study focusing specifically on data use in Australia provided a slightly more promising picture (Pierce & Chick, 2011). In general, more than half of the 74 teachers in this survey held a positive view about the value of NAPLAN. In response to the value of NAPLAN on a string of activities, total 'agree' and 'strongly agree' percentages were as follows: identifying student weaknesses (80%), identifying gaps in curriculum (69%); planning instruction (59%); and grouping students (53%). However, it is also worth mentioning that percentages for 'agree' in all the cases above were significantly larger than 'strongly agree'. Other uses such as identifying students' understanding, misconceptions or knowledge received much lower agreement. On average, over 60% agreed that NAPLAN data are relevant to their teaching but only 39% felt that NAPLAN actually provided information additional to what they already knew. Interestingly, the same survey showed that mathematic teachers were generally more positive about NAPLAN than were English teachers. Mathematic teachers' fluency with numbers may have contributed to their overall positive view.

**The nature of data use.**

Perceptions, however, do not always lead to behavioural change. In Australia, Pierce and Chick (2011) found that only roughly 25% of teachers had access to the NAPLAN data to gauge their students' level of understanding, strengths and weaknesses. Where teachers did have access, Dulfer et al.'s (2012) survey indicated use at a very broad level. Just under 60% used NAPLAN to check for big surprises, and almost half used it for instructional change, collaborating with colleagues, or for assisting individual students. More discouragingly, approaching a third admitted that they glanced at the data but made no changes to their teaching.

Among schools that embrace external assessment data in the US, their uses include: setting and communicating annual and intermediate goals as part of the performance-improvement process; aligning education instructions with state standards; informing school programming and instructional decisions; identifying at-risk students; grouping students; tracking student progress; fostering collaboration; and motivating staff and students (Coburn & Talbert, 2006; Feldman & Tung, 2001; Halverson et al., 2007; Wayman, Spring, et al., 2012). At the state or district level, data engagement is confined to setting goals and directing curriculum planning. At the school, and especially at the teacher level, data usage becomes multipurpose and individualised (Coburn & Talbert, 2006; Halverson et al., 2007; Young, 2006).

**Data-engagement criteria – human factors.**

Despite viewing external data with scepticism, teachers do value data as a useful tool for teaching, and desire to be more fluent in data use. In the US, a district-wide evaluation of data use in Texas (Wayman et al., 2007, p. 21) found that 86% of the respondents agreed to the following statement: "improving my ability to

use data will help me become a better educational professional”, 93% agreed that “data help professionalism and practice”, and 79% agreed that their district should become a data-driven district where decisions are informed by data. In Australia, 60% agreed that their school should engage more with data, and 70% wanted to make more use of data than they currently did (Pierce & Chick, 2011). While teachers and administrators in both countries were equally positive about the prospect of data use, administrators demonstrated more faith in external assessment data (Dulfer et al., 2012; Wayman & Stringfield, 2006a). For example, in Dulfer et al.’s survey, more than twice the number of administrators than teachers strongly agreed that NAPLAN has diagnostic value.

However, many administrators stumble over translating their own enthusiasm for and belief in data use into encouragement to teachers to engage with data and fulfil the promise in actual practice. A meta-analysis (Young & Kim, 2010) reviewing empirical research focusing on data use cited multiple studies mentioning school leaders as the necessary catalyst to start the new data-driven culture. Yet, leadership is the one element that many schools lack (Goldring & Berends, 2009) and some researchers (Wayman, Spring, et al., 2012) suggest that the reason could be that school leaders are not prepared with the skills to implement data-driven practice. For example, skilled leaders set expectations for teachers to engage with data in their core practice (Young, 2006). They build the necessary data infrastructure to store and access data (Young, 2006). In addition, they provide the necessary and appropriate training to help teachers to manipulate and analyse numbers (Goldring & Berends, 2009; Halverson et al., 2007; Young, 2006). Last, but not least, leaders create the right environment, such as grade-level or subject teams,

within which teachers can collaborate and share both positive and negative experiences (Wayman & Stringfield, 2006a). All these practices to encourage data use were observed in an in-depth study combining site-visits and interviews at five US Choice schools identified as having innovative data practice with surveys of school principals elsewhere that implemented similar data-use process (Supovitz & Klein, 2003).

**Data-engagement criteria – systemic factors.**

Organisational leadership is salient because systemic challenges can halt data-driven culture development efforts. A number of studies have cited multiple organisational and technological predictors that can either impede or accelerate a school's adoption of data practice. In the Australian pilot study (Pierce & Chick, 2011), 19% of respondents reported not having received any NAPLAN reports from their school administrators. Only 42% of participants reported having access to data for students in their own class, while 18% said they chose not to access the NAPLAN data. Other reasons for not engaging with data included a lack of guidance on how to utilise NAPLAN data, and a general lack of understanding of the NAPLAN report (Pierce & Chick, 2011). Availability of time for data analysis has been identified in nearly every study on data engagement in the US (Goldring & Berends, 2009; Ingram et al., 2004; Means, Padilla, & Gallagher, 2010; Wayman et al., 2007; Young & Kim, 2010). In the survey by the US Department of Education (2010), under a quarter of the K–12 teachers reported having time during the working day for data analysis. This sentiment was also echoed by their Australian peers in submissions to the 2013 Australian Senate Hearing on the *Effectiveness of the National Assessment Program – Literacy and Numeracy* (The Senate Standing Committee on Education and

Employment, 2013). The submissions also listed similar obstacles mentioned. For these reasons, the Australian Senate Committee concluded that NAPLAN has not lived up to its aspiration as a diagnostic tool for teachers (pp. 11–12).

Equally important is the availability of financial resources to build and collect data and to provide training opportunities for educators to effectively engage with data (Halverson et al., 2007; Monpas-Huber, 2010; Wayman, 2005; Young, 2006). Despite its importance, data training has not been ubiquitous. Only 43% of educators in a US national survey (Means et al., 2010) reported having training on data analysis using standardised assessments, and less than half reported making efforts in designing or purchasing information-management systems for their schools to engage with data in-depth. Furthermore, just over half of teachers reported that they required support in developing classroom assessments, and in adjusting content and instructional approaches in light of results in the data (Means et al., 2010). In general, these findings revealed that teachers have the basic skills to read data, graphs and charts, but not the higher-level skills to analyse and manipulate data to guide instructional change.

All the barriers outlined above also are facilitators of data-driven practice when done correctly, as evidenced in a qualitative study (Datnow et al., 2007) examining two urban school districts and two charter management systems in the US. These organisations have been identified as leaders in data engagement. Together, these four organisations managed 153 schools. According to the study's (Datnow et al., 2007) findings, these schools had leaders who created a data-driven culture by establishing clear and measurable goals to set the foundation and expectations necessary to bring everyone to the table. These leaders invested in

information-management systems to help organise data into comprehensible and user-friendly formats, and to make data access and dissemination easy and timely. They also built capacity for teachers to understand, analyse or manipulate data to gain further insights. According to Datnow et al., building these capacities is what these successful school districts did well. They invested in professional development and training on data, provided time for teacher collaboration based on results from data, and connected teachers to share successes and failures. Their success in raising student outcomes was attributed to these schools' capacity to develop tools, processes and norms for teachers to act and share the data so that everyone – administrators, teachers, students and parents – can work towards a common goal (Datnow et al., 2007).

Successful examples of data use, however, are far and few between. The formidable barriers to changing teacher perceptions about the value of external data and to building technological capacity through professional, infrastructure and cultural development cannot be understated. These changes resemble what Cuban described as 'second-order' change that structurally alters the day-to-day school operation or teacher behaviours (Cuban, 1988). Borrowing the words of Jensen at the Grattan Institute in Australia, "any measure of school performance should not be viewed as an end in itself; they should be a basis of action" (The Senate Standing Committee on Education and Employment, 2010, p. 9), this study intends to explore actions that schools and teachers have taken with the advantage of data to make these second-order changes towards narrowing the achievement gaps.

## Conclusion

The aggregate US and Australian literature on the impact of test-based accountability demonstrates both positive and negative responses at the school level. Test-based accountability has resulted in incidences of unintended consequences, such as narrowing of the curriculum, exclusion, educational triage, and teaching to the test. At the same time, it has encouraged some constructive responses: increasing curriculum rigor, aligning the curriculum with instruction, paying attention to those students most in need of support, and professional collaboration. More importantly, the accumulated evidence from multiple large-scale empirical studies suggests that the reading and mathematics proficiency of disadvantaged students in the US is trending up albeit slowly. At the very least, accountability reform has not led to worse academic outcomes for minority and disadvantaged students. Australian students' outcomes are inconclusive for the general population and obscure for disadvantaged students, due to a host of data-tracking issues, a general lack of external literature to evaluate progress, and the shorter existence of the external assessment implementation. Despite encouraging education progress, particularly in the US, achievement gaps between advantaged and disadvantaged students in both countries remain wide. It is fair to say that, in both countries, the results are not anywhere close to those hoped for by policy makers. Casserly (2007), head of the council for America's large urban school system summed up the current results this way, "it is becoming increasingly evident that this landmark legislation [NCLB] is both living up to many of the promises its strongest proponents hoped for and encountering many of the pitfalls its harshest critics warned against" (p. 43). For disadvantaged students, Casserly concluded that,

“although it has proved complicated to implement and cumbersome to administer, it has helped America’s urban schools direct attention to students who, for too long, were out of sight and out of mind, that alone has made NCLB worth the effort” (p. 65).

It should surprise no one that reforms at this level of complexity are doomed to experience implementation missteps. What is more important is that both governments have listened and responded. In Australia, two Senate Hearings (The Senate Standing Committee on Education and Employment, 2010, 2014) regarding NAPLAN’s impact on various constituencies have already taken place, and recommendations made to improve the experiences of students, educators and schools. Equally important is the Senate’s recommendation to make NAPLAN data available to researchers. In the US, regarding the failure to meet NCLB’s aspiration goal by 2014, President Obama’s administration assumed responsibility for many of the implementation challenges discussed in this chapter. The recognition that NCLB had become a ‘barrier to reform’ instead of an ‘instrument to reform’ (US Department of Education, 2013) led to flexibility for individual states to plan and address their unique educational strengths, challenges, and needs in the reauthorised ESSA (US Department of Education, 2015). In the lead up to the reauthorisation, US Secretary Duncan emphasised that this new direction “isn't simply about compliance—it's about results” (US Department of Education, 2013). What remains unchanged, however, is the reliance on external assessment data to inform decisions at various levels of the education system. If the current climate of civic and business coalitions continues to bring out education reforms that result in producing competitive workers for the global economy (Cuban, 2004), accountability



is not a concept that will disappear overnight. Therefore, it is important that this study contributes to the body of research on constructive data practice to support disadvantaged students.

## Chapter 4 Theoretical Framework

This chapter presents the efficacy theory and the planned behaviour theory, the two constructs chosen to frame the qualitative findings of the present study. The chapter begins by explaining the motivation for identifying these theoretical constructs to guide the qualitative analysis. It follows with an overview of these two theories. Next, it reviews the research that lies behind each theory, discusses their predictive power for behavioural change, and considers their applications within the field of education. Finally, it explains the rationale of why these two constructs have been applied in the present research and of why they have been combined to form a collective construct to frame the data analysis.

### Identifying a Theoretical Framework

The current education reforms in Australia and in the US rely on data to accomplish many bold goals. The intention is for external assessment data to enable governments to monitor school and student progress, and to inform funding allocation. Publically available assessment outcomes also enable parents to exercise choice over school selection for their children. Furthermore, data are expected to permeate all aspects of school management and teaching pedagogy in the effort to close achievement gaps. Policy makers envision teachers using large-scale and comprehensive data, in addition to classroom data, to inform their teaching practice and to diagnose student needs. Of the goals, external accountability and market choice can be legislated for, and in fact are part of the education mandate. The third goal – internal accountability through data practice, by contrast, is more complex and not readily enforceable, particularly given the attitudes and scepticisms of

educators towards external assessment data and also on the host of implementation obstacles discussed in Chapter 3.

The initial purpose of the qualitative phase was to investigate the uses of data and to compare and contrast differences between the US and Australia, given the dissimilar nature of the two regarding accountability policies (punitive in the US; supportive in Australia); and aspiration goals (100% proficiency within 12 years (2002 – 2014) in the US; top five ranking on international measures within 23 years in Australia (2008 – 2025)). However, early results from the field-work indicated that, on the whole, the nature of data use might not differ widely from what is considered innovative data use in the literature. Instead, the difference appeared to be the respondents' perceptions regarding data and data practice. Beyond the mechanics of data utility, the findings also indicated some rather strong beliefs about data practice. The researcher was left wondering why this group of educators managed to effectively use data for continuous improvement while others could not even see value in data as discussed in the literature review. It then became clear that understanding participants' actions necessitated an evaluation of their internal beliefs or 'collective intentionality' (Searle, 1995) of data-driven practice. According to Searle, an American philosopher, through collective intentionality, people engage in cooperative behaviour, share beliefs, knowledge, desires and intentions to create a different set of *social facts*.

The social facts at the participating schools are therefore a product of their individual and collective human agentic function. As Bandura (2001) explained, agency "enable[s] people to play a part in their self-development, adaptation, and self-renewal with changing times" (p. 2). Indeed, education has gone through a

significant change under the current accountability environment. It is the intention of this research to understand the formation of educators' personal and collective beliefs about data, and how those beliefs facilitate or prohibit further behavioural change to embrace and engage with data to support student learning. Two complementary frameworks have been selected to guide the interpretation of the agentic beliefs and intentionality involved in making the switch to data-driven practice. One involves the conceptual framework of personal and collective efficacy (Bandura, 1977, 1995), and the other, the theory of planned behaviour (TPB) (Ajzen, 1991). The combined construct is applied for the purpose of interpreting and finding connections among the themes surfaced and sources of data gathered from field work.

### **Efficacy Theory – An Overview**

Self-efficacy, the foundation of collective efficacy (Bandura, 1977, 1982, 1986b, 1997), is rooted in social cognitive theory. Social cognitive theory (Bandura, 1986b) postulates that behavioural change derives from a common cognitive mechanism. Self-efficacy or, more specifically, perceived self-efficacy refers to the "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1977, p. 3). It refers to people's future-oriented beliefs about their own competence towards a targeted behaviour. These beliefs then influence their feelings and motivations to take on the activity under consideration. Beyond taking on the targeted behaviour, efficacy assessment further determines how much effort they might exert for a specific chosen activity, and how

long they might sustain the effort to achieve expectations when faced with challenges.

When considering the concept of self-efficacy, it is important to heed three distinctions. First, self-efficacy should not be mistaken for other closely related constructs such as self-concept, self-esteem, or self-worth, as it is task-specific (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998; Zimmerman, 2000). Self-efficacy refers to a person's belief in his or her competence for a specific activity. This distinction is important because, as Bandura (1997) pointed out, a person can feel completely inefficacious about an activity, but experiences no loss of self-esteem because he or she does not care about that activity. It has further been argued and proven that self-efficacy contributes to self-concept development (Bong & Skaalvik, 2003; Zimmerman, 2000). Secondly, it is also important to understand that self-efficacy concerns a person's belief about his or her competence as opposed to his or her actual competence (Tschannen-Moran et al., 1998). This distinction is necessary because people may overestimate or underestimate their actual competence, and this has direct implications for the course of action as well as for the efforts they expend on the action (Bandura, 1986a; Tschannen-Moran et al., 1998). Bandura (Bandura, 1977, 1986b, 1997) drew one more important distinction: the difference between self-efficacy belief and outcome expectations or control. The former is concerned with an individual's conviction that he or she is capable of orchestrating a task, while the latter is about whether the action of interest produces expected outcomes. Between the two, some research has shown that the predictive power of outcome expectations is smaller than that of self-efficacy (Shell, Murphy, & Bruning,

1989). The reason, according to Bandura (1986b), is because self-efficacy precedes outcome expectation in the cognitive process.

Collective efficacy is born out of the notion that people do not make behavioural changes in isolation; instead, most challenges and difficulties people ponder are socially connected or motivated. For this reason, having a strong sense of collective efficacy can contribute to desirable change (Bandura, 1995). Bandura defined collective efficacy as a group's shared belief "in their collective power to produce desired results... A group's attainments are the product not only of shared knowledge and skills of its different members, but also of the interactive, coordinative, and synergistic dynamics of their transactions" (Bandura, 2000, p. 75). Efficacy beliefs, whether personal or collective, however, do not form in a vacuum. Instead, they are influenced by four forms of experience: mastery experiences, vicarious experiences, social persuasions, and physiological and emotional states (Bandura, 1982, 1995, 1997; R. D. Goddard & Skrla, 2006)<sup>1</sup>.

### **Mastery experience.**

Bandura (Bandura, 1986b, 1995) theorised mastery experience as the sense of success that people attain when achieving something of significance through giving their very best. Performance accomplishments aid the development of personal efficacy; failures undermine it. This is particularly true if failure occurs before a sense of efficacy has yet to be established. Bandura (1977) emphasised that mastery experience is particularly salient to the strengthening of efficacy when extra effort has been exercised to achieve the expected outcomes. This mode of

---

<sup>1</sup> Bandura used slightly different terminology to convey the same ideas in his earlier scholarly publications in 1977. The terminology used here for the four experiences is from his later publication in 1997.

experience is especially powerful and enduring in its influence on behaviour change because the experience is direct. In academic settings, “teachers who lack a secure sense of instructional efficacy show weak commitment to teaching and spend less time on academic matters” (Bandura, 1993, p. 134). In comparison, teachers with “high perceived coping efficacy manage academic stressors by directing their efforts at resolving problems” (p. 134).

**Vicarious experience.**

Self-efficacy appraisals are influenced in part by perceived internal capability and in part by modelling influences (Bandura, 1986b, 1995). The latter refers to a person’s comparison of his or her own efficacy with the efficacy of people with similar skills. Seeing others with similar skills to oneself succeed in a quest increases a person’s belief in his or her own ability to do the same. Conversely, seeing them fail reduces one’s self-efficacy belief. Bandura stressed (Bandura, 1986b) that this association does more than provide a social standard against which to judge one’s own capabilities. The associated competent social models can “transmit knowledge and teach observers effective skills and strategies for managing environmental demands” (p. 4). This can be helpful in sustaining an observer’s effort when direct experiences fail to boost self-efficacy. Furthermore, people tend to seek out models whose competency they aspire. This leads to the improvement of their personal competency in the long run.

**Social persuasion.**

Encouragement and dissuasion from others influence an individual’s belief that he or she has the necessary skills to perform a particular activity. Often, “people are led, through suggestion, into believing they can cope successfully with what has

overwhelmed them in the past” (Bandura, 1977, p. 198). Social persuasions nudge people to put more effort into their endeavour. Bandura (1986b) noted that the most effective efficacy builders are those that go beyond positive appraisals. For example, appraisers or supporters might also put in place structures that will improve the chances of success or lessen the likelihood of premature failure. It is also important to differentiate between verbal influence that is aimed at enhancing self-efficacy, and that aimed at outcome expectations (Bandura, 1986b). Persuasions that focus on outcomes have a less mediating effect, because simply informing someone of an activity’s benefits does not necessary lead the person to believe what he or she is told. Verbal influence that has the potential to raise self-efficacy belief can result in more enduring effects, because it can contribute to corrective performance, which increases the chances of success.

**Physiological and emotional state.**

A final factor that can alter a person’s belief in self-efficacy is his or her physiological and emotional state of mind (Bandura, 1995). Positive mood raises an individual’s perception of his or her efficacy; negative mood diminishes it. Similarly, exhaustion, stress and pain debilitate belief in one’s capacity; while a healthy and properly functioning body enhances it. However, Bandura (1995) emphasised that it is not always the sheer intensity of affective state that influences efficacy belief; how a person perceives and interprets his or her affective state can also alter efficacy belief.

Together, these four forms of beliefs contribute to either efficacy judgment or outcome judgment. The former reflects a person’s perceived capability to accomplish a particular level of performance; the latter refers to the likely



consequence of the performance (Bandura, 1986b). Self-efficacy expectancy is presumed to create more impact on both the initiation of behaviour and the persistence necessary to push forward when faced with setbacks or failure (Bandura, 1997). Applying these notions to data-driven practice, a teacher's belief that he or she has the requisite skills and time to engage with data to inform instructional practice is an efficacy judgment; the anticipated student achievement as a result of engaging with data to inform instruction constitutes an outcome expectation. Bandura (1986b) noted that it is critical to differentiate the two judgements, because anticipation of outcome does not necessary lead to performance if self-efficacy is weak. This reasoning offers a possible explanation of teachers' general reluctance to engage with data. Despite suggestions that data-driven practice can lead to greater student achievement, which can be considered an outcome expectation, teachers have not unilaterally taken to data-driven practice. Lacking the requisite skills for data analytics, for example, would impact a teacher's efficacy judgment. On the other hand, a person may feel very efficacious for the activity in question, but still choose not to execute it because he or she sees no incentives in doing so. Action can also be constrained by access to the tools or resources necessary to perform the activity adequately. In the case of data engagement, the lack of a good data infrastructure or leadership vision can play a role in teacher motivation.

#### **Efficacy in the education context.**

Self-efficacy has been widely accepted in education as a highly effective predictor of students' motivation and learning (Usher & Pajares, 2008; Zimmerman, 2000) and of teachers' conviction that they can affect student outcomes (R. D. Goddard et al., 2004; Shaughnessy, 2004; Tschannen-Moran et al., 1998). Compelling

empirical results have identified efficacy to be the locus of students' aspiration and motivation to learn to the best of their abilities (Collins cited in Bandura, 1993; Pajares, 1994); of teachers' confidence and effort to deliver effective instruction (Tschannen-Moran et al., 1998); and of faculties' collective ability to create the right environment for students to learn (R. D. Goddard et al., 2000). Working together, these three perspectives of efficacy beliefs reinforce the development of efficacy among students and teachers to strive for positive school outcomes (Bandura, 1997).

A review (Pajares, 1996) of efficacy research provided extensive evidence that self-efficacy can directly predict behaviour as well as indirectly mediate behavioural change through other concepts such as self-concept, self-regulation, goal-setting, and anxiety, to name a few. More notably, some studies in Pajares' review indicated that the effect of self-efficacy was as strong as the effect of ability on student performance ( $\beta = .349$  and  $\beta = .324$ , respectively). In a separate study (1994), Pajares evaluated the predictive power of a series of factors on solving mathematics problems among 350 students. The predictive effect of self-efficacy was significantly higher than other predictors, including gender, prior experience, mathematics self-concept, and the belief in the usefulness of mathematics. Pajares (1994) also noted that self-efficacy's influence on performance was direct, whereas the strength of other attributes' effects was mediated by self-efficacy. As for the four determinants that contribute to students' efficacy development, a comprehensive review of literature (Usher & Pajares, 2008) highlighted the following median correlations: mastery experience,  $r = .58$ ; vicarious experience,  $r = .34$ ; social persuasion,  $r = .39$ ; and affective state,  $r = .33$ . The authors, however, caution that

care must be taken in applying these results, as some of the studies reviewed had operational issues and contexts that could have influenced outcomes.

In regards to teacher efficacy, Tschannen-Moran and colleagues' (1998) review documented the relationship between teacher efficacy and commitment to a series of initiatives. These initiatives included: professional development; progressive approach; clarity and enthusiasm in teaching; willingness to work with struggling students as opposed to referring them to special education; and commitment to student outcomes. Like student efficacy development, mastery experience has also been noted to make a significant contribution to the efficacy development of 74 novice teachers (Tschannen-Moran & Hoy, 2007), findings which corroborated earlier evidence in Goddard's (2001) study on school-level efficacy. Tschannen-Moran and Hoy showed that demographics explained only 2% of novice teachers' efficacy development. When school contextual level settings such as resource support were added to the regression model,  $R^2$  increased to .20. Adding in social persuasions, particularly those from colleagues and the community, further increased  $R^2$  to .31, while including mastery experience took  $R^2$  up to .49.

Collective teacher efficacy is best considered as "the product of the interactive dynamics of the group members" based on the "sum of the individual attributes" (R. D. Goddard et al., 2000, p. 482). It reflects the judgment of the school as a whole, or a grade-level team, regarding the group's ability to organise and execute the courses of action to positively affect student outcomes (R. D. Goddard & Goddard, 2001). In general, across various disciplines including education, there are fewer empirical studies on collective efficacy than on self-efficacy, the former is a more recent construct (R. D. Goddard et al., 2004). However, the few studies

identified (Bandura, 2000; J. C. Lee, Zhang, & Yin, 2011; Tschannen-Moran et al., 1998) yielded conclusions that parallel those found in self-efficacy research: collective efficacy has significant explanatory power for the collective behaviour of a group (Bandura, 2000).

Bandura (1993) was the first to demonstrate a strong relationship between a school's sense of collective efficacy and its academic performance, independent of SES status. His study found that students' SES backgrounds indirectly influenced a school's overall outcomes by altering teachers' beliefs about their collective competency to motivate and to educate students. It is this collective sense of doubt regarding their ability to effect change, as opposed to anything inherent in their students' adverse backgrounds that led to low academic achievement. This study led Bandura to conclude that high collective efficacy enables teachers to approach difficult circumstances and tasks as challenges to be mastered or overcome, rather than as barriers or threats to be avoided.

Other studies corroborated the strong link between collective teacher efficacy and student achievement (R. D. Goddard, 2001; R. D. Goddard & Goddard, 2001). In their study of 47 schools and 438 teachers, Goddard and Goddard (2001) observed that collective efficacy strongly predicted variations in teacher efficacy compared to other contextual factors, such as SES and student performance. Goddard et al. (2000) also concluded that, teachers' feelings of efficacy are very context specific, varying from one situation to the next, from one subject to the next, and from one class to the next. Therefore, in the case of the adoption of data-driven practice, it cannot be assumed that an efficacious staff would see no barriers in implementing data-driven practice. Instead, Goddard et al. (2000) remind us that it is

necessary to assess teacher weaknesses and strengths in each situation. The Australian pilot study (Pierce & Chick, 2011) on NAPLAN data use supports this claim. In this study, mathematics teachers (presumed to be proficient with numbers) were more positive about data use as compared to English teachers.

Therefore, Bandura warned that it is a mistake to consider perceived collective efficacy as a “monolithic group attribute” (Bandura, 1997, p. 479). In general, Bandura’s research noted that early elementary faculties tended to have a higher level of collective efficacy than upper-grade faculties, because academic deficiencies are more glaring in the later years, leading teachers to doubt their personal ability to overcome student challenges.

#### **Justifications for the efficacy construct.**

The concept of efficacy is salient in this study because there is strong evidence (Tschannen-Moran et al., 1998) validating the relationship between teacher efficacy and student achievement across the last three decades. Teachers’ sense of personal efficacy affects their overall attitude toward the educational process and their eventual instructional practice (Bandura, 1993, 1995, 1997; R. D. Goddard et al., 2004). Schools can be a stressful environment due to a variety of stressors including, but not limited to, the wide spectrum of student academic abilities, behavioural tendencies, and demographic backgrounds. Bandura (1997) presented a large body of research demonstrating that teachers with strong perceived self-efficacy focus on finding solutions to overcome these challenges to push on with their academic agendas. On the other hand, teachers with a low sense of efficacy employ punitive approaches to control what they see as barriers and fall prey to assuming the role of student custodian.

Collective efficacy is particularly relevant in the school environment, because “teachers operate collectively within an interactive social system, rather than as isolates” and their collective sense of efficacy or lack thereof “can pervade the entire life of the school” (Bandura, 1995, p. 20). Bandura also suggested that attaining this sense of collective efficacy “requires cogent means of relating factional interests to shared purposes. The unifying purposes must be explicit and attainable through concerted effort” (Bandura, 1986b, p. 145). In the education context, this could be creating a network for teachers, as found in a Dutch study (Moolenaar, Slegers, & Daly, 2012). Moolenaar et al. concluded that well-connected teacher networks are strongly correlated with collective teacher efficacy and can thus strengthen collective efficacy beliefs in schools. Through verbal encouragement and support, these networks raise the confidence of teachers who have a low sense of efficacy.

Bandura’s efficacy construct provides a framework with which to evaluate educators’ sense of agency to engage with large-scale data in the changing education environment. This construct can guide the evaluation and understanding of the underlining sources of influence affecting schools’ collective efficacy belief in relations to data use. Specifically, in what ways have mastery experience, vicarious experience, social persuasion and physical and emotional states affected teachers’ cognitive processes regarding their ability to engage in data-driven practice? However, as Bandura (1986b) noted, having a strong sense of efficacy does not necessarily lead a person or a group to actually take action, because the action might not align with their goals or values. To understanding the rationale behind educators’ goals and intentions, this study turned to the theory of planned behaviour.

## The Theory of Planned Behaviour

In the late 1960s, Wicker (1969), an American social psychologist, published a widely cited article challenging the popular assumption that attitudes and social behaviours are closely related. Wicker's meta-analysis (1969) of empirical studies returned no evidence to support the assumption that feelings are directly translated into actions, and that attitude is the sole determinant of behaviour, as was claimed by researchers of the day. The Theory of Reasoned Action (TRA), an 'integrated model of behaviour' conceptualised by Ajzen and Fishbein (1980), sought to improve or extend the attitude-behaviour relationship (Armitage & Conner, 2001). TRA consists of two constructs: attitude, and subjective norms (Ajzen & Fishbein, 1980; Madden et al., 1992). It suggests that attitude and behavioural norms contribute to people's behavioural intention, and in turn determine whether they will act on that behaviour.

Within the TRA concept (Ajzen & Fishbein, 1980), 'attitude' refers to people's beliefs about the consequences of acting on a behaviour, or taking a particular action. 'Subjective norm' is concerned with people's view of general social pressure to perform or not to perform the action. Working together, attitude and subjective norms influence people's intention, and the strength of their intention further predicts action. Attitude and subjective norms are independent of each other, although both are predictive of intention to act or not act (Ajzen & Madden, 1986). The more favourable the attitude and subjective norms towards a behaviour under consideration, the greater the likelihood of intention to act on the behaviour (Ajzen & Fishbein, 1980) and the reverse is also true. Intention is considered to be an important predictor of behaviour, because intention encapsulates both motivation

to take action and the effort exerted towards that action. Intention is also the precursor of behaviour (Ajzen, 1991). As restated by Orbell and colleagues (Orbell, Hodgkins, & Sheeran, 1997, p. 946), intention is a “summary of the cognitive and affective mechanisms through which attitude, subjective norm, and perceived behavioural control direct future behaviour” (p. 946).

Specific to the current study, TRA offers a meaningful guide for understanding the proliferation, and non-proliferation, of data-engagement. As the surveys on data use presented in the Chapter 2 show, attitude plays a large role in schools’ and teachers’ willingness to engage with data in their daily practice. For example, Australian teachers in general did not believe that NAPLAN data could enhance their knowledge of student needs (Pierce & Chick, 2011). The large percentage of educators not adopting the practice (Dulfer et al., 2012) sends a signal that data-driven practice is not a strategy that most practitioners approve; hence, there is no social pressure for a school, a grade/year level, or an individual teacher to follow suit.

In the mid 1980s, Ajzen (Ajzen, 1991; Madden et al., 1992) extended the TRA by adding an additional construct: perceived behavioural control (PBC). Encompassing attitude, subjective norms and PBC, this extended model came to be known as the theory of planned behaviour (Ajzen, 1985, 1991). According to Ajzen, the additional measure reflects the barriers people perceive could positively or negatively impact the behaviour under consideration. Ajzen (1991) emphasised that *perceived* behavioural control differs from *actual* behavioural control because the latter’s influence on behavioural change is evident. The author argued that, if people have control over resources and opportunities, the decision to take action is clear.



Ajzen (1991) offered PBC to strengthen the predictive power of TRA in situations where the intention to take action is weak, or when people are not in complete control of their collective volition. In other words, intentions are assumed to mediate performance only to the extent that people have behavioural control over other factors affecting performance. In most social situations, Ajzen (1991) contended that, personal and environmental constraints exist that impede actions. These constraints can be both internal and external. Internally related factors include skills, knowledge, and abilities; externally related factors concern resources and opportunities. Ajzen called all of these 'non-motivational factors' (Ajzen, 1991, p. 182), reasoning that,

The importance of actual behavioural control is self-evident: The resources and opportunities available to a person must to some extent dictate the likelihood of behavioural achievement. Of greater psychological interest than actual control, however, is the perception of behavioural control and its impact on intentions and actions. (p. 183)

Under these circumstances, intention alone will not accurately predict behaviour. However, Ajzen (1991) hypothesised that, in their aggregate, attitudes, subjective norms and PBC represent a more valid measure of the underlying behavioural disposition than do the behaviours in the original TRA. Conversely, when the situation affords people complete control over behavioural performance, Ajzen contended that TRA alone will predict behaviour just as effectively. However, when volition is not under control, the level of PBC is raised by a greater belief in resources or opportunities, and a lower perception of barriers or impediments, which in turn, increases the chance of action.

Under TPB, attitude, subjective norms and PBC each can independently determine behaviours (Ajzen, 1991). Ajzen explained that, while all three predictors

may contribute to intentions and subsequently to actions, they do not all necessarily contribute to behaviour change or carry the same predictive weight. For example, PBC's direct influence is evidenced when intentions are weak or when people have low volition (Ajzen, 2002). In these situations, perceived facilitators or inhibitors to the behaviour in question play a stronger role in influencing the likelihood of action (Ajzen & Madden, 1986). Each of the three constructs' individual predictive powers varies across behaviours, situations and populations (Ajzen, 2002). The theory further suggests that any single determinant, or all three in concert, can predict intention (Ajzen, 1991).

If perceived behavioural control sounds similar to self-efficacy, that is because PBC "owes it greatest debt" to Bandura's perceived self-efficacy theory (Ajzen, 2002, p. 667). Like the concept of efficacy, which is about judgment of competency to perform an act, PBC is about judgment of the available resources to perform an act. In Ajzen's view (Ajzen, 2002), they are one and the same in that both refer to the *behaviour* to attain a certain outcome and not control over attainment of an outcome itself. Yet, they are distinct in that self-efficacy measures the "ease or difficulty of performing a behaviour", whereas controllability measures the "beliefs about the extent to which performing the behaviour is up to the actor" (Ajzen, 2002, p. 672).

A considerable amount of evidence from multiple domains (including health behaviours, consumer behaviours, family-planning decisions, and more) has supported TPB's predictive power of behaviour (Ajzen, 1991; Armitage & Conner, 2001; Madden et al., 1992). In their respective aggregates, attitude, subjective norms and PBC deliver a high predictive validity of behaviour, an average correlation

of 0.51 across multiple studies (Ajzen, 1991). As for TPB's predictive power of intentions, the same meta-analysis reported that, on average, TPB explained 71% of the variance of intentions. Another meta-analysis (Armitage & Conner, 2001) consisting of 161 studies across numerous domains confirmed the same trend: that TPB predicts both intentions and behaviours but is more strongly correlated with intentions. It further found that the TPB, with all three constructs, explained 39% of the behavioural variance across the studies; and intentions and PBC together without subjective norms accounted for 27% of the variance. Among the three measures, subjective norm is found to have the weakest predictive power. However, Armitage and Conner suspected that the reason rests with the weakness of the measure. They also found that, just as with efficacy belief, PBC has direct influence over behaviour. Given the strength of the link between PBC and intention, Ajzen and his colleagues postulated that "strategies could be formulated for changing intentions, and subsequently behaviour, by changing perceptions of control" (Madden et al., 1992, p. 9).

#### **TPB in the education context.**

Research application of TPB in school settings has been limited (Crawley, 1990; Davis, Ajzen, Saunders, & Williams, 2002; Haney et al., 1996; J. Lee et al., 2010). In Davis et al.'s (2002) longitudinal study exploring high school completion among 166 African Americans, attitudes, subjective norms and perceived behavioural control accurately predicted 71% of the students' intentions to complete high school. Other studies (Crawley, 1990; Haney et al., 1996; J. Lee et al., 2010) also found a strong link between teachers' intentions and their adoptions of various instructional programs. Using the three determinants of TPB, Haney et al. (1996)

investigated 800 school teachers' intentions to implement four strands of the Ohio State Competency Based Science Model. Their results indicated that attitude towards the behaviour construct had the most significant impact on teachers' intent. Another study (Crawley, 1990) also focused on science education in schools, found similar strength in 50 elementary and secondary teachers' intentions to integrate, and actual integration of, what they learned in a science program into their classroom practice. In a separate study about educational technology adoption, Lee et al.'s (2010) research concluded that all three components of TPB significantly predicted 34 middle school-teachers' adoption of educational technology in their instruction.

#### **Justifications for the Planned Behaviour Theory.**

While self- and collective-efficacy beliefs are an important determinant of motivation to act on a behavioural change, there are other external factors within a school environment that also affect the collective decision by teachers to adopt new practices. In the current era of accountability, social pressure is a lever used by policy makers to effect action. To this end, it is important to consider the concept of subjective norm as part of the investigative process into school and teacher intention to adopt new data practice. Furthermore, given the negativities that surround test-based accountability, it is important to give adequate weight to teacher attitudes to the data generated as part of the testing process. Finally, it is imperative not to overlook educators' actual and perceived behavioural control over the available infrastructure, data and time to collect and to analyse data, as well as the training on how data should be used. TPB offers a strong framework for evaluating these factors, which fall outside the judgment of internal competency.

## Theoretical Issues

Notwithstanding the proven predictive power of both efficacy theory and TPB across multiple disciplines, definitional, operational and measurement challenges remain. Pertaining to the efficacy construct, questions have been raised regarding the distinction between efficacy beliefs and other expectancy constructs (Kirsch, 1986; Maddux & Stanley, 1986; Manstead & Van Eekelen, 1998; Pajares, 1996). Kirsch (1986) argued that many determinants of the efficacy concept had been tested before 1977 (the year Bandura proposed the self-efficacy theory) within the concept of expectancy for success. The two theoretical constructs are “logically and operationally equivalent” (Kirsch, 1986, p. 340). It is generally agreed that, conceptually, efficacy and expectancy beliefs converge; the difference is that the former is task-specific, while the latter is more global in nature (Ajzen, 2002; Kirsch, 1986; Pajares, 1996). Operational issues, such as mismanagement of measurements in studies, also contributed to the lack of success in differentiating between these constructs (Kirsch, 1986; Maddux & Stanley, 1986; Pajares, 1996). Some studies used global assessment statements to predict efficacy instead of task-specific ones (Zimmerman, 1996, April). Other studies employed assessment constructs that were so vague that the responses could be interpreted in multiple ways (Pajares, 1996).

The close connection between efficacy and behavioural control means that, TPB suffers similar measurement and clarification issues (Armitage & Conner, 2001). Furthermore, since “TPB is held to be a complete theory of behaviour” (Conner & Armitage, 1998, p. 1432), as it encompasses multiple influences of behaviour, Conner and Armitage noted that its sufficiency has been subjected to scrutiny. This led to the suggestion of additional predictors for the model including: past

behaviours or habits (Aarts, Verplanken, & Van Knippenberg, 1998), self-identity (Sparks & Guthrie, 1998) and belief salience (Pligt & Vries, 1998), to name a few.

### **Justifications for Combining the Two Theories**

As discussed above, there are more similarities than differences between efficacy theory and TPB. Empirical studies have also demonstrated that both theories predict action equally well, both directly and indirectly, through other concepts such as attitude and intention (Ajzen, 2002; Manstead & Van Eekelen, 1998). However, as noted in the preceding *Theoretical Issues* section, neither theory is complete.

Standing alone, neither addresses all the key elements in the current policy and school environment relating to data-driven practice. In Maddux and Stanley's words (1986), the outcome expectancy and value models "have much in common with self-efficacy theory and need to be viewed as compatible and complementary instead of competing" (p. 253). Because efficacy theory and TPB complement each other in providing a more complete framework to evaluate both the internal and external motivations of behavioural change, these two constructs are combined to guide this study. In the case of schooling, and in the era of accountability, internal and external motivators or pressure both play a role in educators' performance beliefs. It would be incomplete and unrealistic to evaluate teachers' motivations to engage with data from only one angle because "the distinction between internal and external causes of a behaviour can have important implications" (Ajzen, 2002, p. 675).

It is important to highlight that both TPB and the construct of collective efficacy reflect judgment, and not actual competency and control. This is because "people's level of motivation, affective states, and actions are based more on what

they believe than on what is objectively the case” (Bandura, 1995, p. 2). According to Bandura (1997), intention is important, though not the sole proximal determinant of behaviour, for “perceived self-efficacy affects thinking, motivations, and affective states, all of which act upon behaviour” (p. 284). Bandura argued that “the move from intention to action is far from automatic” (p. 285). However, Ajzen’s early empirical work (Ajzen & Fishbein, 1980; Ajzen & Madden, 1986) on the theory of reasoned action which consisted only of the intention construct, demonstrated that intention to action, while not automatic, can happen. It is therefore, reasonable to conclude that efficacy belief, intention, and perceived behavioural control all play a role in regulating decisions to act. For this reason, both constructs are deemed meaningful to frame the findings of the present research. Figure 4.1 demonstrates how these two theoretical constructs are integrated to explain a behavioural change, or lack thereof, in this study.

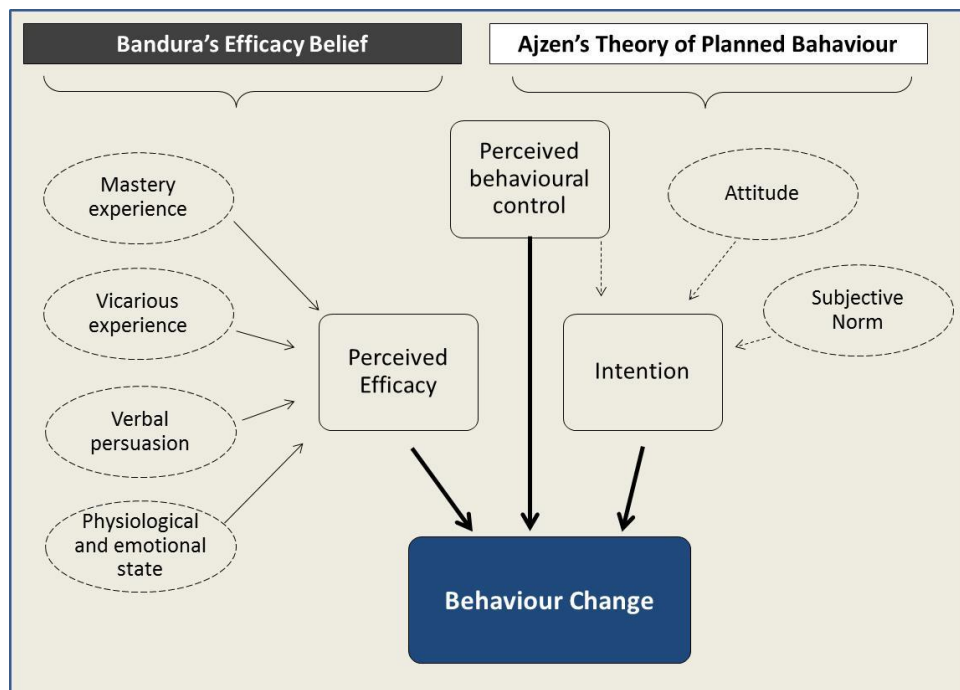


Figure 4.1. Combined Theoretical Framework

In summary, understanding schools' decisions to engage with data requires an exploration of schools' and faculties' collective internal and external beliefs. For most schools, data-driven practice at the scale and depth envisioned by policy makers is a new pedagogical practice. The efficacy construct provides a guide to consider the determinants influencing schools' and faculties' view of their capabilities to take on the reform challenge. TPB offers a framework to understand how educator intentions, along with perceived behavioural controls, positively or negatively impact a school or a grade-level team's decisions to fully engage with data as part of their daily practice. Policy makers put a lot of hope in data-driven practice as a means for educators to support student academic achievement. For this benefit to materialise, it is important to recognise that teachers are both a conduit of, and the actors delivering, the changes envisioned by the policy design. Data-driven practice will not be widely adopted if policy makers fail to recognise that teachers have agency to exercise influence over what they do regardless of policy requirements, and understand the underlying beliefs leading to actual behaviour change. To sum up using Bandura's words (2000), "people are partly the products of their environments, but by selecting, creating, and transforming their environmental circumstances they are producers of environments as well. This agentic capability enables them to influence the course of events and to take a hand in shaping their lives" (p. 75) or, in this case, their schools and the academic achievement of their students. Understanding these antecedents or determinants of behaviours can enhance the adoption of data as part of school practice, because strategies can then be formulated to influence beliefs and ultimately to change behaviours.



## Conclusion

This chapter has argued for the selection of Bandura's (1986b) efficacy theory and Ajzen's (1991) planned behavioural theory as a framework to guide the qualitative findings of this study. The combined strength of these two constructs enables the study to investigate the exercise of forethought and self-regulative standards that motivate and guide school and teacher actions to adopt or to reject data-engagement as a core part of instructional practice. These two theories were chosen for their clarity in explaining the complex cognitive process affecting human behaviour. It is the belief of this researcher that, in rushing to achieve the noble goals of closing achievement gaps through data accountability and data engagement, policy makers have glossed over the need to understand how to motivate educators to share their belief in data. Akin to Bandura's suggestion that "theories that seek to explain human behaviour solely as the product of external influences or the remnants of the past stimulus inputs present a truncated image of human nature" (Bandura, 1989, p. 1179), policy makers' beliefs that teachers will engage with data because of accountability pressure, sanctions and rewards reveals a lack of understanding that teachers have agentic control over their willingness to embrace the data-engagement strategy as envisioned. This study hopes to contribute to a better understanding of these underlining beliefs to encourage educator action.

## Chapter 5 Methodology

This chapter focuses on the research methodology employed to evaluate the three research questions outlined in Chapter 1. It begins by establishing the context and the methodological approach of the present research. This is followed by the philosophical assumptions and justification behind the choice of mixed methods. Next, it recaps the research questions and explains how the independent quantitative study and qualitative study aided the research investigation. The research design, procedure and the instruments are then introduced, including the process for sample selection and data acquisition, preparation and analytical procedures. Lastly, limitations and challenges encountered during this quantitative and qualitative data-gathering process are also discussed.

### Context of the Study

As described in Chapter 1, this research project had three overriding goals: (1) To evaluate the academic progress of disadvantaged students under the current accountability environment; (2) To explore the belief mechanisms behind data-engagement and the relationship between data-use and disadvantaged student achievement; and (3) To compare the impact of the transparency accountability policy on disadvantaged student learning and teaching in both Australia and the US. These research objectives were achieved through a mixed methods approach. The selection of this approach was informed by Fulcher's (1989) interpretation of policy, and Crossley and Vulliamy's (1984) and Broadfoot's (2000) observations about the limitation of policy and comparative research.

Fulcher (1989) interprets policy as the capacity to make decisions and act on them. As such, it is equally possible to consider policy as being any of the following: the written law, a teacher's instructional decision, or instructional style. Fulcher insists that, while government may set policy at the national level, it cannot predict how policy is enacted at subsequent levels within the education system. Judgments and decisions regarding the course of action emerge through debate, persuasion and discourse (Fulcher, 1989). For example, regarding data use, teachers might engage in discourse about how to use assessment results, which aspects of the curriculum to focus on, or which students to target. According to Fulcher, the resulting judgments, decisions and practices, are in effect policies at the classroom level derived from the policy text at the national level, but processed through discourse that is based on contextual uniqueness. This interpretation of policy does not in any way downplay the significance of textual policies or imply that they produce marginal effects. The effects, Fulcher (1989) argues, are contextualised and therefore vary from the intentions of the policy makers. Fulcher's interpretation of policy implies that, to understanding the effects of policy, it is necessary to investigate how it is construed at the local context.

In a similar vein, Crossley and Vulliamy (1984) observed over three decades earlier that "comparisons between schooling in different countries are almost exclusively conducted in terms of educational policies and only rarely... are questions raised as to the relationship of such policies to the realities of schooling" (p. 197). The authors recommended the case-study method as a means to fill that gap. Although it has been three decades since Crossley and Vulliamy's observation, comparison analyses of educational policies and of their actual implementation in

concert are still uncommon. Furthermore, noting the growing influence of quantitatively oriented international studies on policy making, Broadfoot (2000) also suggested that “detailed qualitative data typically complemented by more quantitative data, [can] reveal important insights about the source, the scale and the educational significance of national cultural variations” (p. 362). It is the goal of the present research to connect the accountability policy text with local-level discourse and actual policy implementation to gain a balanced view of policy impact on student outcomes. The mixed methods approach was determined to be appropriate for this study. The quantitative aspect of the design can evaluate the impact on student outcomes, while qualitative case studies can tease out the local context, discourse, and beliefs that influence how schools and teachers support achievement growth.

### **Philosophical Assumption and Implication for Research**

Social science researchers have multiple research methodologies to draw from, and their goal is to find the methods most suitable to answer their research questions, based on the implicit worldviews they bring to their inquiries (Creswell & Plano Clark, 2007). Research methodologists (Corbin & Strauss, 2008; Creswell, 2008) recommend that researchers should consider their own worldviews as “there is an ‘objective’ reality [e.g. the classroom], but there is also a ‘subjective’ reality [e.g. the researcher sees different things as he or she looks at a classroom]” (Creswell, 2008, p. 554). According to Corbin and Strauss (2008), these objective realities are the “great varieties of human action, interaction, and emotional responses that people have to the events and problems they encounter” (p. 6). They contend that there is

no one reality waiting to be discovered by the researcher, instead what he or she will encounter are events, that “are the result of multiple factors coming together and interacting in complex and often unanticipated ways” (p. 8). The researcher brings subjective realities when he or she constructs concepts and theories from these events, which the research participants share (Corbin & Strauss, 2008). The authors further offer that these subjective realities are particularly poignant in qualitative studies.

As the head of an American preschool/primary school at the time of the field-work reported here, this researcher entered the data-collection phase with a worldview that schools are complex environments in which multiple competing and changing contextual factors coexist. The researcher adopted the pragmatist worldview in designing this research, because the goal was to capture the complexity associated with the research participants and their environments, so that knowledge could be constructed (Corbin & Strauss, 2008). This worldview offers a prudent position from which to begin this study for it prioritises the importance of the research questions and the consequences of research over the methods used (Corbin & Strauss, 2008; Creswell & Plano Clark, 2007).

### **Mixed Methods Research Paradigm**

Guided by “a practical and applied research philosophy” of the pragmatism worldview (Tashakkori and Teddlie, 2003 cited in Creswell & Plano Clark, 2007, p. 27), the present study aimed to “draw from ‘what works’, [by] using diverse approaches and valuing both objective and subjective knowledge” (Creswell & Plano Clark, 2007, p. 26). The empirical purpose of this study sought to find

'complementarity' (Ouwueghuzie & Johnson, 2006) between student academic achievement on the external standardised assessment and teacher engagement with external data. According to Chatterji (2005), mixed methods research, through the combination of quantitative and qualitative data, offers the best opportunity to achieve research findings that are contextually grounded and empirically defensible. Being 'a middle ground' (Creswell, 2003, p. 179) approach, mixed methods is deemed a "natural complement to traditional qualitative and quantitative research" (B. Johnson & Onwuegbuzie, 2004, p. 14). Rather than disputing which of the two research traditions – quantitative or qualitative – is superior (as traditional positivists who favour quantitative, and constructivists and interpretivists who favour qualitative have done), this approach enables researchers to embrace the strengths of both traditions and focuses on their compatibility to meet complex research goals (Creswell & Plano Clark, 2007; B. Johnson & Onwuegbuzie, 2004; Phye, 2008; Teddlie & Tashakkori, 2010).

#### **Rationale for mixed methods research.**

Quantitative methods are commonly used to investigate relationships and causes (Wiersma, 2000) based on hard data and statistical analysis, both of which are perceived as having the best chance of achieving reliability and validity (B. Johnson & Onwuegbuzie, 2004; Pratt, 2003). Therefore, a quantitative study is appropriate for analysing standardised assessment results to gauge student progress in the current test-based environment. However, quantitative research lacks context and realism, and these can be supplemented by qualitative research (Guba & Lincoln, 1994). According to Guba and Lincoln, context and realism are necessary components of research, since causes and effects are neither context nor value-free,

but interdependent. This is because behaviours “cannot be understood without reference to the meanings and purposes attached by human actors to their activities” (Guba & Lincoln, 1994, p. 106). As noted at the 2004–2005 Education and Social Research Council’s Research Methods Initiative Workshops (cited in Byrne, Olsen, & Duggan, 2009), even when uniformly applied, national policy will not appear uniform after it has interacted with a local context, because that interaction will generate very different sets of conditions leading to possibly different outcomes.

There are multiple realities within the context of the present study. For example: the dynamics among the federal-, state-, and regional- or district-level policy creation, policy interpretation and implementation; the school environment and school context; and the relationships amongst various players including but not limited to administrators, parents, teachers and students. It is important to note that these realities are not something that research can manipulate (B. Johnson & Christensen, 2008); they are part of the research context that a researcher must take into consideration. Therefore, quantitative method alone would not adequately reflect the social organisation of life in classrooms and schools (Raudenbush and Willms, 1991, p. xi cited in Rowe, 2003), or explain how and why internal or external conditions might have influenced the effects observed (Amrein-Beardsley, 2008; Chatterji, 2005; Day, Sammons, & Gu, 2008). Complementing quantitative with qualitative research can redress the balance (Guba & Lincoln, 1994). Qualitative investigations can provide contextual information to sort out and find meaning from interactions in the research setting.

Since qualitative research offers authenticity and context, and quantitative research provides rigor and objectivity, together these methods make a

complementary pair (B. Johnson & Onwuegbuzie, 2004, p. 193; Phye, 2008; Raudenbush, 2005). Creswell and Garrett (2008) summed up the benefit to researchers as follows:

When researchers bring together both quantitative and qualitative research, the strengths of both approaches are combined, leading, it can be assumed, to a better understanding of research problems than either approach alone. (p. 322)

Furthermore, because the present study aimed to evaluate variables in schools whose “manifestations have already occurred”, as opposed to something that the research actually manipulated (B. Johnson & Christensen, 2008, p. 357), mixed methods was deemed the most appropriate design for this situation. Lastly, mixed methods research is believed to be especially suitable for studies with multiple objectives (B. Johnson & Christensen, 2008; Sammons, 2010) as in the current study.

However, mixed-methods research has a relatively short history, originating fewer than three decades ago (Creswell & Garrett, 2008), which means that the approach is still evolving. Notwithstanding its significant application across multiple fields (including social, behavioural, health, and human sciences) (Heyvaert, Hannes, Maes, & Onghena, 2013), a recent special issue of the *Journal of Mixed Methods Research* (Mertens & Hesse-Biber, 2012) revealed that many of the issues that Creswell and Garrett (2008) had outlined a decade earlier remain unresolved. For example, debate continues about the appropriateness of mixing two distinct research paradigms (Sale, Lohfeld, & Brazil, 2002) that are fraught with tensions (Creswell & Garrett, 2008) and inherently incompatible (Denzin, 2012). Another issue is nomenclature (Teddlie & Tashakkori, 2010): precise terminology and definition for concepts within the mixed-methods approach have not been established, a



challenge that has become greater as the terms used, and variations of them, have multiplied with increased application of this approach (Teddlie & Tashakkori, 2010).

Issues of design and triangulation are of particular relevance to the present study. For example, Teddlie and Tashakkori (2010) raised the question of whether qualitative and quantitative research should be conducted in parallel, or whether one should follow the other. Sale et al. (2002) questioned whether both quantitative and qualitative research should study the same phenomenon or whether they can each explore a separate phenomenon. Finally, there is still a need to clarify the question of how and when is it appropriate to adopt mixed methods research to increase the understanding of a research question (Mertens & Hesse-Biber, 2012). The perspectives of mixed methods authors vary widely regarding both design and triangulation. This study adopts Howe's (2012) view on design and triangulation, and this will be discussed in more detail later in the *Research Design* section of this chapter.

## Research Questions

As stated earlier, the research questions are as follows:

- (1) What have the test-based accountability policies in Australia and California accomplished in the area of assessment inclusion and achievement of disadvantaged students?
  - What are the general trends in academic progress?
  - How have disadvantaged students fared compared to advantaged students?
- (2) How and why have school administrators and teachers chosen to invest time and effort in data-driven practice to support student learning?

- In what ways are schools and teachers taking advantage of the volume of data available to them to advance the learning outcomes of disadvantaged students?
- In what ways have constructive data practice benefited the academic performance of disadvantaged students?
- In what ways does constructive data practice impact school goals, instruction, lessons and expectations of students and of teachers?

(3) How have the different policies in the two countries affected the learning and teaching experiences of disadvantaged students and their teachers?

## Research Design

Adopting Teddlie and Tashakkori's (2010) design proposal, the quantitative and qualitative studies were conducted concurrently in three locations: NSW, Northern California, and Hawaii. The design also followed Howe's (2012) concept of *disjunctive* mixed methods, which the author also refers to as *mixed methods interpretivism*, where quantitative and qualitative each have a distinct role in the investigation of phenomena. Howe states that this view "embraces a division of labor" between the two research paradigms (p. 89). The role of qualitative research is to discover and to interpret, and role of quantitative research is to justify phenomena (Howe, 2012). As Mertens and Hesse-Biber (2012) put it, Howe's interpretivism view is "somewhat controversial" as it assigns quantitative "the role of description" and qualitative "the role of providing causal explanations because they can answer the "'why' question" (p. 76). According to Howe (2012), this distinction is important because social and institutional facts are "human constructions in the sense that they wouldn't exist but for the activities of human

beings” (p. 91). In the present study, the qualitative design helped to discover and interpret the rationale and motivational factors behind data-driven practice adoption leading to the outcomes in the quantitative data. In summary, the design is such that the qualitative and quantitative phases are separate, but complementary, in explaining the same phenomenon.

Disadvantaged students were well represented in the present research design. The student groups included: low socioeconomic background, disability, recent immigrant status, Indigenous students, and English language learners. The first research question aimed to extend the knowledge base regarding the impact of test-based accountability on the achievement of students from disadvantaged backgrounds. This question was achieved through quantitative analysis of large-scale assessment results beyond the few schools where the qualitative research took place. Several factors make evaluation of progress among disadvantaged students particularly pressing in Australia: NAPLAN’s lack of data disaggregation, two poorly designed disadvantage categorisations, an emerging trend of exclusion from the test, and a scarcity of existing empirical research. In California, county-level analysis had not been carried out in previous research and empirical evidence on achievement outcomes in California all relied on NAEP data. It is expected that quantitative findings from the present study will illuminate the progress of disadvantaged students in Australia in a manner not currently offered by the NAPLAN report, and will add to the body of knowledge on this topic in the US.

To explore the second and the third research questions regarding data-use and its connection to trends observed in the quantitative phase, six case studies were conducted altogether, with two cases in each geographical location. Each case

study consisted of one primary school. In California, however, one of the two cases also provided access to a district-level data officer and a district-wide English curriculum coach. In both the quantitative and qualitative phases, results and the learning experience of disadvantaged students were the main focus or dependent variable. Findings from these case studies will contribute to the limited research on the local reality of data practices and on personal and collective belief systems leading to data use.

### **The Quantitative Study**

The quantitative phase of this research consisted of collecting and analysing the NAPLAN assessment results in Australia and the CST results in California to determine the progress of disadvantaged and advantaged student segments over time. In both locations, the unit of analysis was school average, because individual student data were not available to the researcher, a limitation not uncommon in empirical studies using similar data. Despite this limitation, the datasets offered information that was useful for the present study's research objectives. Since multiple studies (Heyneman & Loxley, 1983; Lamb, Rumberger, Jesson, & Tesse, 2004) have indicated that student and school characteristics are the key inputs into the academic outcomes of schools, the present study evaluated the progress of disadvantaged and advantaged students using a combination of school and student characteristics for the Australian sample and student factors for the Californian sample. Figure 5.1 presents the factors used to gauge progress on various achievement and inclusion indicators between various disadvantaged and advantaged populations.

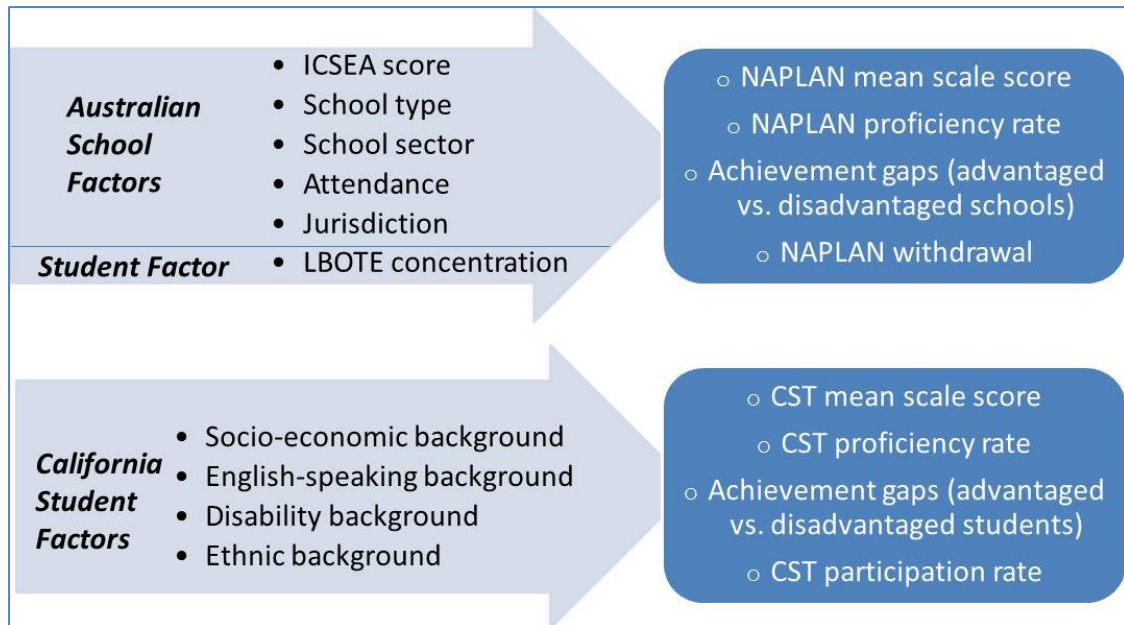


Figure 5.1. Analysis Input and Output Variables

For the Australian NAPLAN data, ACARA provided the dataset directly to the researcher. In California, disaggregated school-level CST data for the two counties where the qualitative case studies took place were drawn directly from the Ed-Data (2011) and DataQuest (California Department of Education, n.d.-c) websites. Hawaii State Assessment data, sought from and provided by the Accountability Resource Center of Hawaii (ARCH), were initially included in the research design. However, after the initial process of data preparation, the quantitative analysis proceeded without Hawaii's assessment data because a significant amount of missing and incomparable data yielded unreliable statistical analysis. The cause of missing data resolves around student subgroup size not meeting ARCH's minimum data-reporting threshold. Incomparable data were a result of the change in subgroup categorisation from one year to the next.

**Data sources.**

NAPLAN and CST vary significantly in design, content, scale score definition and data-tracking intention. To start with, NAPLAN is offered only to students in Years 3 and 5 at the primary level, whereas CST is offered to all primary school students from Grade 2 onwards. As such, each Australian student is tested once every two years, as compared to Californian students who sit the test annually. Secondly, the test contents also diverge. NAPLAN focuses on many strands of literacy as compared to the CST. On the other hand, the CST also tests students in the science domain, which NAPLAN does not. Thirdly, NAPLAN uses a common set of scale scores across all grade levels. This scale ranges from zero to 1000 with younger students, such as Years 3 and 5, falling on the lower end of that scale range (ACARA, 2014). Student proficiency achievement is conceived in the same manner, with students across Years 3 to 9 falling within ten bands. For students in Year 3, a 'band 1' achievement implies performance that is below the national minimum standard. For Year 5 students, band 3 or lower implies the same.

In contrast, the CST adjusts every grade level to a common scale range from 150 to 600. The intention is for the scale score to carry the same meaning regardless of grade, test, or year, so that a reader can compare one calendar year to the next, or one grade to the next, to gauge school progress (California Department of Education, 2013b). Similarly, the CST determines proficiency standards by whether students reached the minimum scale score of 350 and provides the percentage of students reaching that benchmark for each grade level. The last and most significant difference is the manner in which data are disaggregated, tracked and disseminated. NAPLAN is set up to observe the growth of the same cohort from Year 3 to Year 5.

Whereas the CST is designed to track the progress of students in each grade level annually. Hence, every year is a different cohort.

In general, American policy makers and educators focus on the progress that a school makes from year to year at each grade level, or for a particular subgroup of students. For example, did Grade 3 students make progress from calendar year one to calendar year two? Did girls make progress between two calendar years? For this reason, the CST data structure is fairly basic; it provides scores for every group listed in Table 5.1 as well as for students with disability. Between the two datasets, NAPLAN data lacks student-level characteristics but does provide a slightly richer set of school-level characteristics.

It must also be noted that four different assessments are offered to accommodate the needs of different student subgroups in California. They include: the California Standards Test (CST); the modified test (CMA) for students with disability and an individual education plan; the California alternative assessment (CAPA) for students with a significant cognitive disability; and, finally, the STS, which is the equivalent of the CST, but in Spanish. The CST is administered to all students regardless of subgroup designation unless there are specific accommodation requirements that call for one of the alternative tests. Because these are alternative assessments, sample sizes are small by default. Many schools did not meet the minimum ten-student requirement by the state for each subgroup for public data reporting. Hence, there was not enough data to develop meaningful analysis of the alternative assessment results. For this reason, only the general CST assessment results were included in the analysis as this was the only assessment with sufficient observations for most student subgroups.

Due to the differences in scale score design, the current study did not strive to compare the scores between the two countries. Instead, the empirical analysis focused more on the rate of students meeting proficiency standards, and compared and contrasted the general directions of student achievement relative to their individual assessment benchmark over a six-year period from 2008 to 2013. Finally, the year 2008 was chosen as the base year for comparison as that is the year that NAPLAN testing began. The end year of this analysis, 2013, coincided with the discontinuation of the CST. In 2010, California adopted the Common Core Standards which has since been adopted by 43 states (Common Core State Standards Initiative, n.d.-b). This led to the introduction of a new state-wide assessment system to replace the CST in the school year 2014–2015 (California Department of Education, n.d.-a).

**Sample.**

The empirical analysis focused on Years/Grades 3 and 5 reading and numeracy data as these are the consistent domains in both countries' assessments. In California, schools are managed through three layers of education authority. At the very top sits the California State Department of Education; it governs the schools through 58 county-level education offices. County-level offices set further directions for their respective schools through district-level offices. The number of school districts within each county differs from county to county, and so does the number of schools within each district. County-level data were used in this analysis because they offer enough observations for each student subgroup to produce meaningful and credible statistics. For reasons of anonymity, the two counties included in the analysis are referred to as Almond County and Walnut County. Each county served



between<sup>2</sup> 230–250 government-funded primary schools, with Almond educating just over 100,000 and Walnut 130,000 students between 2008 and 2013. Government-funded schools come in the form of public, charter and choice schools. As private school students are not required to take the CST, they are not included in this analysis.

In Australia, the data are from the national sample including all school types: government, Catholic and independent, totalling about 6,300 primary schools and two million students. The dataset for Years 3 and 5 contains 391,684 observations across eight jurisdictions over the six-year period from 2008 to 2013. New South Wales, Victoria (VIC) and Queensland make up three-quarters of these observations at 31.9%, 24.4% and 17.7%, respectively. In contrast, the smallest of three jurisdictions, the Australian Capital Territory, the Northern Territory (NT) and Tasmania (TAS), had, respectively 1.4%, 1.7% and 3.1% share of the observations. Western Australia (WA) and South Australia (SA) had the remaining 20% of the observations. Detailed profiles of both sets of data are presented in the next two chapters prior to the presentation of findings.

#### **Definition of progress.**

This study conceptualised progress through the positive movement of four different measures: (1) assessment participation; (2) means scale scores on the assessment; (3) students meeting proficiency standards; and (4) achievement gap between advantaged and disadvantaged students. In its most basic specification, this study defined progress as statistically significant gains in the first three measures between the calendar years 2008 and 2013, and statistically significant reduction in

---

<sup>2</sup>Exact count is avoided to protect the anonymity of the participating schools.

achievement gap between advantaged and disadvantaged students over the same period.

**Definition of participation.**

For the Californian dataset, the measure for participation is the ratio of two variables: students tested and total number of students enrolled. While the dataset provides the total tested for each student subgroup, it does not provide school enrolment at the subgroup level. Therefore, it was not possible to evaluate subgroup-level participation. In comparison, NAPLAN tracks and disaggregates student inclusion and exclusion in the assessment using four different categories: *assessed*, *exempt*, *absentee*, and *withdrawal*. According to ACARA (2014), *assessed* consists of students who sat for NAPLAN; *exempt* is for students who were officially exempt from NAPLAN due to low English language capability, recent immigrants (less than one year in Australia) and students with severe disabilities. *Absent* refers to students who were not present at school on the day of the test, whereas *withdrawal* represents students who did not sit the test as a result of a conscious and open decision by parents or caregiver to withdraw their children from NAPLAN.

The current analysis focused on the NAPLAN trend for *absent* and *withdrawal* in combination as opposed to the *assessed* and *exempt* trend (ACARA defines it as *participation*), which is the focus of ACARA's summary report. This decision factored in the discussion of negative consequences of accountability and transparency in the literature. Results from recent qualitative studies (Wyn et al., 2014) and submissions to the Senate Committee on Education (Australian Education Union, 2013; Thompson, 2013) suggest that some schools and parents deliberately keep low-performing students from school on the day of NAPLAN testing. The decision also

factored in an interesting trend in the exempt rate, observed by the researcher: it has remained steady since the launch of NAPLAN (ACARA, 2014). Since the NAPLAN participation rate has declined (as discussed in Chapter 3) in spite of an unchanged official exemption rate, the steady exempt rate suggests that deliberate parental decisions to withdraw children must have contributed to the overall decline in participation rate. Therefore, analysing the movement of *absent* and *withdrawal* was determined to provide a more accurate picture of the exclusion claims and the overall impact of external testing on inclusion.

#### **Data analysis and procedure.**

##### ***Preparing data.***

The analysis procedure began with data preparation, which entailed data cleaning and data transformation. The goal of data cleaning is to improve the quality of the data and involves identifying and removing or transforming errors, incompleteness, inconsistencies and outliers from the dataset (Han, Kamber, & Pei, 2012; Rahm & Do, 2000). In addition, tests of normality of data distribution and homogeneity of variance were conducted. This process is important as most statistical procedures are parametric tests and assume a normally distributed dataset whose variances are homogeneous (Field, 2009). Frequency distribution, graphs, histograms and Levene's test were also performed to correct data challenges and to ensure assumptions were met.

##### ***Addressing outliers, non-normality and unequal variances.***

Inconsistencies or outliers were not systematically removed. Where it made sense to change the data using commonly accept practices (Field, 2009), this

approach was applied. For example, if all but one entry for a school showed a school SES score of 1000, then the one value that differed from 1000 was clearly an error as every data entry for the same school should have the same school socioeconomic value. In this case, instead of removing the observation with the 'wrong' value, the value was changed to match the rest of the entries for that school. In situations where outliers were related to scale score and proficiency entries of one school, *mean plus two standard deviations* was the chosen technique (Field, 2009) to correct the problem. Some schools had disproportionately low scores, but these were neither changed nor excluded unless there was strong evidence suggesting that they were outliers (MacDonald & Robinson, 1985). In these cases, school and student factors were used to determine the likelihood of the score in question being true or an outlier. Similarly, the analysis did not exclude small schools, as a recent study (Miller & Voon, 2011) using NAPLAN data demonstrated that excluding them had no effect on the study's statistics. Finally, to correct for distributional challenge and unequal variances, the study executed the square root transformation method (Field, 2009).

### ***Aggregating data.***

To evaluate proficiency rates in Australia, the study underwent a data aggregation process that involved summing two or more numeric variables into one. The NAPLAN data provided percentages for every band from 1 to 10 to indicate whether a school had met the proficiency standard. Based on ACARA's definition of minimal proficiency standards for each year level, bands 2 and above were collapsed into a single variable called *Year 3 proficiency*; and bands 4 and above were

collapsed into a new variable called *Year 5 proficiency*. Similarly, *withdrawal percentage* and *absentee percentage* were computed into a single variable called *exclusion rate* for further analysis.

***Defining advantaged and disadvantaged students.***

Because NCLB mandates assessment data disaggregation in fine detail, it was possible to evaluate the progress of California disadvantaged students without further data manipulation. The CST data disaggregated students by ethnicity, socioeconomic background, native and non-native speaker background, as well as disability background. For the purpose of comparing advantaged and disadvantaged students subgroups, the pairs in Table 5.1 were compared to each other.

Since data disaggregation is a significant weakness of NAPLAN (as discussed in Chapter 3) it was necessary to create proxies for advantaged and disadvantaged subgroups to compare advantaged and disadvantaged school outcomes. A two-step process was followed: (1) the study evaluated the impact of school and student background factors on assessment outcomes using linear multiple regressions; (2) proxy advantaged and disadvantaged categories were created for factors that the regression results indicate an impact on achievement outcomes.

Table 5.1

***Advantaged and Disadvantaged Subgroups for California Data Comparison***

Student background:	Economically disadvantaged/economically advantaged	
	English learners/English only students	
	With disability/no disability	
	Parents with a high school education or less/associate or higher degree	
Student ethnicity:	Black/White	Hispanic/White
	Black/Asian	Hispanic/Asian
	Black/Hispanic	

School and student variables included in the regression model were informed by studies that have evaluated background factors on student outcomes. For example, one study (Miller & Voon, 2012) documented considerable differences between government and non-government schools even after taking ICSEA differences into consideration (Miller & Voon, 2011). Other studies (Marks, McMillan, & Hillman, 2001; Miller & Voon, 2011; Slins & Murray-Harvey, 2000) found that socioeconomic factors explain a significant amount of the variance between government and independent schools. As would be expected, more than any other variables, ICSEA scores explained a large portion of the variance and attendance was found to have a greater effect among government schools than non-government schools (Miller & Voon, 2012). The final list of background variables for the regression model included the following:

- ICSEA score: denotes the school's socioeconomic status,
- school sector: government, independent and religious schools,
- school type: primary, combined primary and secondary, and special school,
- attendance rate,
- LBOTE status: family language background other than English,
- Indigenous status,
- socioeconomic advantage (SEA) score,
- calendar year,
- enrolment size, and
- the combined absentee and withdrawal rate.

Most of the variables are self-explanatory, but some require further clarification. Calendar year was included to capture the effect that might have been related to overall policy changes from year to year. Under *school sector*,

'independent' refers to non-government schools that are not Catholic. With 14% of students in this sector, it is therefore tracked as a separate sector from Catholic schools, the other non-government school sector (Gonski, 2011, December). With regards to *school type*, 'combined schools' serve K-12 students all in one school and 'special schools' are government-sponsored primary, secondary or combined schools that solely serve students with disability. Special schools serve three categories of special needs: disability, sensory, and medical and mental health. In 2011, 14% of all schools were combined schools and 4% were special schools in Australia (Australian Government Productivity Commission, 2013).

The regression results displayed in Table 5.2 indicate that the factors in the model explain 68% and 61% of the variation in NAPLAN scores between 2008 and 2013, respectively, in reading and numeracy in Year 3 and 72% and 66% in Year 5 at the  $p < .05$  significant level. The combined explanatory effects of these factors align with evidence in a recent study (Miller & Voon, 2011) using 2008 NAPLAN data, which found effect sizes in the range of 60% (the current model and Miller and Voon's model shared six similar background factors). Among the background predictors in the regression model in the current model, ICSEA status and attendance rate stand out as variables that had a disproportionately large effect on NAPLAN. Attendance, a variable not tracked in the NAPLAN annual report by ACARA, is found to correlate strongly with SES status and this finding corroborates with evidence found in the US literature (Gottfried, 2010; Ready, 2010). While there are significant differences in school NAPLAN results between calendar years, calendar year has little impact on the other factors' effect on NAPLAN scores, with the exception of SEA status and Indigenous student concentration.

Upon reviewing the significance of each factor's contribution to the model, SEA status and Indigenous concentration were removed from further statistical analyses as these variables are highly correlated with ICSEA scores. Bottom SEA status and Indigenous representation are negatively correlated with ICSEA  $r = -.803$ ,  $p < .05$  and  $r = -.822$ ,  $p < .05$ , respectively. This is not entirely unexpected. ICSEA score, an indicator of a school's socioeconomic status, is calculated based on SEA status and other parental factors (ACARA, 2015) and Indigenous Australians have the lowest socioeconomic status among all Australians (Altman, 2000). Furthermore, while Indigenous status is an important indicator of disadvantaged background, the NAPLAN summary report by ACARA provides extensive information, hence, repetition adds little value. The LBOTE category, however, remained on the list due to the issues surrounding the accuracy of that category as discussed earlier. A finer definition of this category could provide another useful perspective regarding students in this category.

Finally, enrolment size and absentee/withdrawal rate were also excluded in the proxy development process as their impact on outcomes is either slightly weaker or is already captured elsewhere. For example, while students attending remote and very remote schools are likely to be in a small enrolment environment, a quarter of those students are in the NT, and an average of 0.7% are in the remaining seven jurisdictions (Australian Government Productivity Commission, 2013). Therefore, potential differences related to enrolment size would surface through NT's results, which the analysis of NAPLAN progress will explore. Table 5.2 summarises the test variables and the associated  $t$  values and standardised beta coefficients from the



regression model. Causal interpretation of the beta coefficients must be done with caution, as this set of covariates is limited.

Step two involved creating advantage and disadvantage subcategories for each of the variables in the final list. For categorical variables such as *school sector*, *school type*, and *jurisdiction*, which have pre-defined categories, the two subcategories with the lowest and the highest NAPLAN means were used as proxies for disadvantaged and advantaged schools, respectively. For example, for the school sector variable, independent schools were used as a proxy for advantaged while government schools for disadvantaged. Catholic schools were excluded from the comparative or gap analysis because their NAPLAN mean is in between government and independent schools. A similar procedure was used for jurisdiction and school type and a summary is shown in Table 5.3.

Table 5.2

*Standardised Beta Coefficients of Australian School Background Variables*

<i>Variables</i>	<i>Year 3</i>				<i>Year 5</i>			
	<i>Reading</i>		<i>Numeracy</i>		<i>Reading</i>		<i>Numeracy</i>	
	<i>t(16372)</i>	<i>B</i>	<i>t(16363)</i>	<i>B</i>	<i>t(16452)</i>	<i>B</i>	<i>t(16448)</i>	<i>B</i>
ICSEA	33.2***	.67	29.3***	.65	38.0***	.71	33.2***	.69
Attendance	20.9***	.16	18.5***	.15	25.0***	.18	20.5***	.16
Enrolment	-5.2***	-.02	-4.6***	-.02	-6.0***	-.03	0.7	.00
LBOTE	-- <sup>1</sup>	.02	6.3***	.03	-8.3***	-.04	8.9***	.04
Government School	-5.0***	-.03	6.6***	.04	-7.1***	-.04	3.6***	.02
Independent School	3.8***	.02	7.5***	.05	4.7***	.03	7.8***	.05
Special School	-5.2***	-.02	-4.6***	-.02	-6.8***	-.03	-5.1***	-.02
Primary School	-1.9*	-.01	-3.3***	-.02	-1.7*	-.01	-2.2**	-.01
Absentee/Withdrawal	4.8***	.02	3.5***	.02	5.6***	.03	0.2	.00
Indigenous	1.7*	.14	0.44	.10	2.4**	.11	1.9*	.10
Top SEA Status	4.1***	.05	2.1**	.03	1.0	.01	4.6***	.06
Bottom SEA Status	-1.5	-.02	-2.6***	-.03	0.9	.00	1.0	.01

Note: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

<sup>1</sup>Too small to report.

Table 5.3

*Disadvantaged and Advantaged Proxies for NAPLAN Categorical Variables*

<i>NAPLAN variables</i>	<i>Disadvantaged proxy</i>	<i>Advantaged proxy</i>
Jurisdiction	The Northern Territory	Victoria
School sector	Government schools	Independent schools
School type	Special schools	Primary schools

For numeric variables such as ICSEA score, attendance rates, and percentage of LBOTE students, the observations were divided into quartiles following an approach used in a study on school- and student-level factors on tertiary entrance performance (Marks et al., 2001). These quartiles were derived from the full Year 3 and Year 5 dataset. The lowest quartile is the proxy for disadvantaged and the highest quartile for advantaged status. Table 5.4 details the quartile ranges derived from the data.

Table 5.4

*Categorisation of NAPLAN Numeric Variables*

	<i>Bottom quartile</i>	<i>Second quartile</i>	<i>Third quartile</i>	<i>Top quartile</i>
ICSEA score	315–960	961–1007	1008–1063	1064–1282
LBOTE student	0–4%	5–10%	11–29%	30–100%
Attendance rate	0–91%	92–93%	94%	95–100%

In summary, the purpose of creating these proxies was to compare the outcomes of disadvantaged and advantaged students over the six-year period. While useful in providing some indication of the differences between these two student groups, the approach may mask differences not addressed by excluding other subcategories in the statistical models. It is also for this reason that, while comparisons were made using only two subcategories within each variable, results for other subcategories such as ‘Catholic schools’ or those in the second and third quartiles are also displayed and discussed in the findings in Chapter 7. However, for

the purpose of the advantaged and disadvantaged discussion, the focus is placed on the subcategories in Tables 5.3 and 5.4.

***Data analysis.***

The quantitative analysis evaluated Year 3 and Year 5 separately as two independent samples. In addition, within each year level, reading and numeracy were also examined separately. In total, four data samples were evaluated: Year 3 reading data, Year 3 numeracy data, Year 5 reading data, and Year 5 numeracy data. For each dataset, descriptive statistics summarised the participation rate, mean scale scores, and minimum proficiency percentage for each school group. Achievement gaps, as well as associated change between 2008 and 2013, were computed through simple computations, where the 2008 results were subtracted from those of 2013 to obtain the absolute change. While the absolute means scores, proficiency rates and achievement gaps from year to year do provide information on change, they do not necessarily represent the actual differences in the population (Field, 2009). A series of statistical tests followed to ensure that observed differences did not occur by chance. Due to differences in data structure between the two datasets, it was necessary to apply multiple statistical procedures. This section provides an overview of the analytical methods executed. Further details about the statistical procedure related to each dataset are discussed in the next two chapters prior to the presentations of findings.

Quantitative analysis for the California data adopted a similar statistical procedure used by the NCES (National Center for Education Statistics, n.d.). The analysis consisted of a series of one-tailed *t*-tests, which provided statistics to help

determine whether: (1) the observed changes between 2008 and 2013 were statistically significant, and (2) the gap between a disadvantaged and an advantaged subgroup had changed significantly between 2008 and 2013. The analysis of NAPLAN data involved two primary statistical approaches: analysis of variance (ANOVA) and multiple regressions. Both approaches have been employed in prior research evaluating large-scale assessment data (Abedi, Hofstetter, Baker, & Lord, 2001; Clotfelter, Ladd, & Vigdor, 2006; Fryer & Levitt, 2004). The present study performed the ANOVA procedure to measure statistical changes in scale scores, proficiency rates, and participation rates between the base year and the end year. To estimate the relationship between school characteristics and NAPLAN, and to evaluate the changes in score gaps and proficiency gaps over time, multiple regression models using the ordinary least squares were built.

**Limitations.**

One limitation of using school average is that some school-level attributes can change from year to year at any given school. For example, English language learners in the US sample could be in the sample one year and out in another, based on whether they have reached proficiency. In Australia, among the attributes listed above, school sector and school type do not change. ICSEA scores, Indigenous and LBOTE student distributions, and attendance rates could change as a result of student mobility. However, the changes observed in descriptive statistics were modest and did not occur at every school or every year. Therefore, it is reasonable to believe these changes were endogenous and random.

Another limitation is the representation of students with disabilities in both countries' datasets. In the California data, the breakdown of the student subgroup is

detailed and makes for meaningful subgroup data analysis. However, some of the subgroups, particularly the students with disability subgroup, often have fewer than the required ten-student reporting minimum, which is applied by the California State of Education to protect student confidentiality. This limitation created a challenge to observing growth of some schools in the present study. For example, at a small school, students with disability might have two years' data, only for this to discontinue the following three, when some students had left and the minimum reporting requirement is no longer met. As a result of this ten-student minimum data reporting parameter, larger schools are better represented in this analysis by default.

This particular limitation is even more acute in Australia because the NAPLAN data do not reveal the enrolment or NAPLAN outcomes of students with disabilities in mainstream schools. In the present study, the proxy for students with disability is the student population enrolled in special schools. However, among all jurisdictions, only NSW and WA consistently included special schools in the NAPLAN data across all years. Queensland included special schools in 2008 only, and Victoria included special schools in some years but not others. Special schools were not represented in any year in the remaining jurisdictions. However, it is also important to note that even if those jurisdictions had reported their data, they make up only 8% of all special schools (Australian Government Productivity Commission, 2013).

Without the disaggregation, the performance of students with disabilities outside of special schools was not represented in this analysis and special schools represent only 4.5% of all schools in Australia (Australian Government Productivity Commission, 2013). Based on the same report, 5% of the total student population in

other school sectors are students with disability, whose outcome the present study could not evaluate due to the lack of NAPLAN identification. A large part of these 'missing' students could be Indigenous students, as the rate of their enrolment in special settings has been found to increase faster than in the mainstream setting in NSW (Sweller, Graham, & Van Bergen, 2012). Despite this limitation, it is still valuable to include the special school sector data to gain some understanding of students with disabilities. Nonetheless, caution must be taken when interpreting the results of special schools.

The third limitation applies only to the California dataset. Alternative assessments such as the CMA, CAPA and STS have a significant amount of missing data due to a subgroup size of fewer than ten students per school. This limitation leaves a gap in understanding the progress of students with disability who took the CMA and CAPA, or English Language learners who took the STS. While lack of visibility indicates that fewer than ten students per school take each of these alternative assessments, and that could be good news, it also prevented the investigation of progress for the most disadvantaged of all students, by the very nature of their need for alternative assessments.

### **The Qualitative Study**

As this is a comparative study of three distinct cultural and geographical settings, the qualitative approach offers "valuable insights into how people construct meaning in various social settings" (Neuman, 2006, p. 308). The present study adopted the case study approach to explore research questions two and three regarding data-driven practice at the local level and its impact on learning and

teaching. Within the qualitative method, case study is a preferred approach for examining “contemporary events, when the relevant behaviours cannot be manipulated” (Yin, 1994, p. 8). Another strength of the case study method lies in its “intensive, holistic description and analysis of a single instance, phenomenon or social unit” (Merriam, 1998, p. 17). In employing the multiple-case study, the research focused on exploring the narratives behind the assessment data-engagement process “in context, rather than a specific variable” (Merriam, 1998, p. 19) to make connections with the observations in the quantitative study. It is precisely because the social reality at each school level is complex, conflict-ridden, and ‘messy’ (Byrne et al., 2009), that it necessitates “narratives of difference, which include – necessarily – narratives of differential, indeed differentiated outcomes” (p. 520). It is this complexity and the indirect influence of the policy pressure on teaching practice and student outcomes that this research sought to explore by linking student achievement trends to changing practices through the availability of assessment data.

Furthermore, the case study method offers a unique strength to deal with multiple sources of evidence (Yin, 1994). Indeed, that was precisely the context of the qualitative components of this study which included: (1) a combination of semi-structured one-on-one and group interviews with teachers, resource specialists and school administrators; (2) site visits; (3) school-level and grade-level team meeting observations; and (4) school documentation collection. Finally, the multiple-case study offers the researcher an opportunity to explore within each school setting, state setting and across country settings (Yin, 1994).

**The role of the researcher.**

In qualitative research, “the researcher is the primary instrument for data collection and analysis” (Merriam, 1998, p. 7). As such, a researcher’s “social relationships and personal feelings and personal, subjective experiences are field data” (Neuman, 2009, p. 276). For this reason, Creswell (2003) reminds researchers of the importance to reflect on their own biography and be sensitive to the role it plays in the inquiry. Because the “personal-self becomes inseparable from the researcher-self”, Creswell suggests that researchers should reflect and acknowledge any biases, values and interests at the onset of the study (2003, p. 182).

Having lived, worked and raised two children in all three geographical locations in the study, the researcher has a personal understanding and experience of the educational environmental, social and cultural nuances of these places. These experiences and understanding enabled the researcher to pay close attention to the nuances of each social and cultural context for the comparative aspect of this study. Furthermore, having spent many years conducting intervention work with at-risk youth, whose background mirrors the disadvantaged student subgroup in this study, the researcher came to this study with some understanding of their academic and social struggles in and outside of school. Thus the researcher was sensitised to the need for good listening and unbiased and non-judgmental qualitative interviews (Corbin & Strauss, 2008) during the case study interviews. Recognising that “all research is advocative”, because the nature of research is that it “should make things work better” (Stake, 2010, p. 200), the researcher took extra care to ensure ethics were followed and biases were controlled by employing recommended procedures in the data-analysis phase.



**Ethical considerations.**

Neuman (2009) encourages researchers to balance the priority of “gaining knowledge and finding a clear answer to a research question”, and “protecting research participants and upholding broader human rights” (p. 62). Approvals from all appropriate departments within the respective education authorities at each of the geographical locations were sought and obtained prior to contacting schools and inviting educators to participate in the qualitative study. First, this involved seeking approval from the University of Sydney Human Research Ethics Committee for research in all three geographical locations. Next, in two of the three locations – Hawaii and Sydney – additional ethics applications to the NSW DET State Education Research Approval Process (SERAP) and Hawaii State Department of Education were lodged. Anticipating a low response rate of 10–20%, 23 government primary schools meeting the sampling criteria were submitted to the NSW SERAP for approval. The same procedure was followed in Hawaii, where an application to recruit from 24 schools in four different school complexes was submitted to the Hawaii State Department of Education. The California State Department of Education does not require approval from its office to conduct school research. However, adhering to the University of Sydney ethics protocol, approval was sought from local school district leaders to recruit individual schools.

Once approval from the necessary higher authority was secured, recruiting efforts began with the school principal, who Creswell and Plano Clark (2007) consider the “gatekeeper, an individual in the organization supportive of the proposed research who will, essentially ‘open up’ the organization” (p. 113). Even when a school principal gave permission and support for the research, the

researcher did not assume that every teacher would be willing to participate.

Therefore, each teacher and specialist was recruited following the same protocol for principal recruitment. The protocol included sharing the research goals, research instruments, and participant consent forms with potential participants. Samples of the following instruments can be found in Appendices A–D:

- research invitation letter to the principal,
- participation information statement,
- school and individual educator participation consent forms, and
- interview guide.

#### **Sampling procedure.**

Since school administrators and teachers are two key actors in the socially constructed environment we call school (Raudenbush and Willms, 1991, p. xi cited in Rowe, 2003), they were the necessary participants for the qualitative phase. As Raudenbush (2005) suggests, policy levers such as resources, incentives or punitive measures cannot directly intervene to improve student outcomes, just as giving the medical community resources and incentives cannot itself save lives. In education, school leaders and teachers provide the vital link between policy measures and schooling. In the case of assessment-based accountability, the greatest degree of variability can be expected at the school level as evidenced in educator responses to data use in the literature review. Their differing responses would then produce varied effects on student learning. Therefore, it is important to explore their objectives, perceptions, expectations and decisions to affect student achievement through the use of assessment data.

The six case studies were neither experimental nor randomised. Instead, a purposeful sampling strategy was followed to recruit schools and participants. According to Merriam, (1998) “purposeful sampling is based on the assumption that the investigator wants to discover, understand, and gain insight and therefore must select a sample from which the most can be learned...” (p. 61). These samples are considered information-rich cases (Creswell, 2008), from which “one can learn a great deal about issues of central importance to the purpose of the research...” (Patton (1990, p. 169) cited in Merriam, 1998, p. 61). In the current study, all six cases were selected because they had earned public recognitions, or made significant progress, in providing support to raise the outcomes of disadvantaged student populations.

To identify candidate schools for the case studies, the researcher applied a two-step sampling process. The first step involved using a combination of government accountability data (ACARA, n.d.-a; Ed-Data, 2011; State of Hawaii Department of Education, 2011), superintendent reports (State of Hawaii Department of Education, 2010), and school accountability reports retrieved either from the government accountability databases or from school websites to create a list of school districts/region/complex and schools that met the following sampling criteria:

- low socioeconomic background,
- an ethnically, culturally or socially diverse student body, such as minority students, non-English speakers, transient students, or a larger than average number of students with disabilities,

- progress on the national or state assessment, particularly for disadvantaged students, and
- a reputation for being innovative in raising student outcomes.

In America, low socioeconomic schools are Title I schools where 40% or more of the students are participating in the reduced or free lunch program. In Australia and for the purpose of this study, low-SES schools are those falling below the average ICSEA index on *My School*.

The second step of the sampling process followed the network sampling strategy, a sub-set of the purposeful sampling strategy (Merriam, 1998, p. 63). Local educators, researchers, school board members, and grant funders were approached for recommendations and referrals to candidate schools that best fit the research criteria. Short lists from both steps became the final lists of schools included in the ethics applications to the education authorities. Teacher sampling was solely based on recommendations from the site administrator or principal who accepted the invitation to participate in the research. While this sampling method clearly creates a bias, according to Supovitz and Kline (2003) who conducted a large-scale data-use study and used a similar sampling method, this is not a deficit of the study so long as the purpose is to identify best practice.

Upon receiving ethics approval from the appropriate education authorities, email invitation letters were sent to either the district leaders or principals to explain the research project and scope. A week following the initial email invitation, a follow-up email and a phone call were employed to secure participation. Once a school district superintendent or a school principal agreed to participate in the research, the same process was repeated to recruit teachers, special resource

teachers, grade-level leaders and other instructional leaders who had been recommended by their respective administrators.

**Participating schools and participants.**

Altogether, the NSW and Hawaii education authority approved 23 and 24 schools, respectively, for recruitment. Of these, 9% responded positively in NSW and 8% in Hawaii. Many schools did not reply to the email invitations or return the follow-up phone calls. The few schools that formally declined the invitation indicated that “too many initiatives” were already going on at the school, or that it was “not the right time for the school to participate” despite their having interest in the research. Recruitment was more successful in Northern California, since most of the schools/districts in the California sampling pool came through recommendations; the response rate in California reached 40% with five school districts being approached.

The final sample, outlined in Table 5.5, consists of six schools, two in each geographical location, and a total of 50 educators with a broad range of roles and responsibilities. Of the 50 educators, 43 were directly interviewed either individually or collectively in small groups of three to four, and eight were observed through grade-level meetings. Among the interviewees, the two roles of ‘instructional coach’ and ‘accountability officer’ exist exclusively in the US sample. Individuals who hold these roles are responsible for supporting teachers in meeting accountability requirements through data analytics and curriculum development. The external data consultant is not affiliated with the school but is a member of the company providing the curriculum to the school. Although not randomly selected, the schools (whose names have been altered in this report to protect their anonymity) represent very

diverse school backgrounds and contexts. One school contains a highly mobile student population; another serves a large recent immigrant population; two schools predominantly support students from extremely low socioeconomic neighbourhoods; two schools have a combination of students from all backgrounds. Given their local contexts, the needs of the schools, teachers and students vary significantly. The limited sample of only six schools in the study makes generalisability more difficult, but certainly not invalid. As the literature on data use suggests (Means, Padilla, DeBarger, & Bakia, 2009), many schools struggle with the same issues to maximise or realise the benefits of data.

Table 5.5

*Summary of Case Study Participant*

	<i>New South Wales</i>		<i>Northern California</i>		<i>Hawaii</i>	
	<i>Bilby</i>	<i>Koala</i>	<i>Almond</i>	<i>Walnut</i>	<i>Kukui</i>	<i>Hibiscus</i>
Classroom teacher	6	4	1	8 <sup>1</sup>		3
Learning support staff	4	1	1			3
Instructional coach				1	1	3
Grade-level leader	1					
Reading/mathematics leader				1		
Accountability officer				1		
Assistant principal	2	3				
Deputy principal	1					
Principal		1	1		1	1
External data consultant					1	
<b>Total participants</b>	<b>14</b>	<b>9</b>	<b>3</b>	<b>11</b>	<b>3</b>	<b>10</b>

*Note.* <sup>1</sup>These teachers were not directly interviewed; they were participants in meetings where the researcher was present to observe.

***California.***

The California public school system is the largest in the US (National Center for Education Statistics, 2013b) with close to 10,000 schools and over six million students (Ed-Data, 2012). It also enrolls the largest number of ELL students in the

nation and 54% of its students live below the poverty line (National Center for Education Statistics, 2013b). The present study's sample involved two urban schools located in Northern California, which will be referred to as Almond School and Walnut School. Neither school was in a Program Improvement plan at the time of this study. Walnut School County operates about 250 elementary schools and close to 140,000 elementary students and it belongs to one of the smaller districts within the Walnut County. District offices are the regional education authorities between the school county and the school. Walnut District operates a total of ten schools from preschool to Grade 8. Together, these schools served roughly 6,500 students in the spring of 2011, of which 4,500 students were in the elementary program. Walnut School district provided the researcher with opportunities to interview faculty and administrators beyond Walnut School, but recruitment beyond Walnut School was less successful. Therefore, most of the data was gathered at Walnut School.

In contrast, the district in which Almond School lies is large with over 70 elementary schools serving about 25,000 students. The County operates over 200 elementary schools and close to 110,000 elementary students. Almond School differs from Walnut School in various aspects. First, unlike Walnut, which is a traditional public school, Almond is a charter school. Charter schools are government-funded school and free; they differ from regular public schools primarily in that they are granted operational flexibility in exchange for public funding and greater accountability (Uncommon Schools, n.d.). Secondly, like private schools, charter schools are free choice schools (Uncommon Schools, n.d.), whereas traditional public schools only serve students within official geographical boundaries. Thirdly, Almond

operates a duo-language-immersion program which prepares students to be bi-lingual and bi-literate.

Table 5.6 provides a summary of the demographics at both schools and their respective counties. Both are K-5 elementary schools and as indicated in the table, both schools serve a predominantly disadvantaged student population compared to their respective counties. White students typically considered the more advantaged group is under-represented while Hispanic students and ELLs are over represented at both schools. Student teacher ratios are both under 20 though Almond has a slight advantage as noted in Table 5.7. Both ratios are just under the county ratios of 21 and 22 at Almond and Walnut, respectively.

Table 5.6

*Student Profiles at California Participating Schools and Districts (2011–2012)*

<i>Student subgroups</i>	<i>Almond</i>		<i>Walnut</i>		<i>State profile</i>
	<i>School profile</i>	<i>County profile</i>	<i>School profile</i>	<i>County profile</i>	
<i>Ethnic groups</i>					
Black or African American	15%	14%	6%	3%	7%
American Indian/Alaskan Native	0%	0%	2%	0%	1%
Asian	7%	22%	15%	27%	9%
Filipino	0%	5%	11%	5%	3%
Hispanic/Latino	68%	32%	50%	39%	52%
Native Hawaiian/Pacific Island	0%	1%	2%	1%	1%
White	6%	22%	11%	23%	26%
Two or more Races	3%	3%	3%	3%	2%
No identification	2%	1%	0%	1%	1%
<i>Other designations</i>					
Free and reduced lunch	96%	44%	63%	37%	56%
English language learners	44%	39%	50%	24%	22%
Students with disability	Information not available				

*Note.* (Ed-Data, 2012)



Table 5.7

*Staff Information at Almond and Walnut (2011–2012)*

	<i>Almond school</i>	<i>Walnut school</i>
Administrator	1	1
Teachers	14	25
Office staff	2	2
Paraprofessional	2	8
Other	4	6
Total	23	42
Teacher to student Ratio	20	21

*Note.* (Ed-Data, 2012)

***Hawaii.***

The school system in Hawaii is small in comparison to California. There are 167 elementary schools serving slightly over 100,000 students throughout the whole state. The complexes where participating schools are located serve close to 8,000 students each. For the purpose of this study, the two participating schools are referred to as Kukui School and Hibiscus School. Both are large elementary schools serving over 500 students. At the time of the research, enrolment at both schools had been increasing for three straight years. These schools' student subgroup profiles and staff counts are as indicated in Tables 5.8 and 5.9. Kukui School is much more economically disadvantaged and serves more students who are identified as English learners compared to Hibiscus School. Hibiscus School, however, serves a more ethnically diverse community. Compared to the Californian samples, both Hawaiian samples enjoy a relatively low teacher-to-student ratio.

Table 5.8

*Student Profiles at Hawaiian Participating Schools and Complexes (2011–2012)*

<i>Student subgroups<sup>1</sup></i>	<i>Kukui</i>		<i>Hibiscus</i>		<i>State profile</i>
	<i>School profile</i>	<i>Complex profile</i>	<i>School profile</i>	<i>Complex profile</i>	
Free and reduced lunch	85%	73%	46%	54%	50%
English learners	34%	25%	3%	6%	9%
Students with disability	6%	8%	11%	11%	10%
Black or African American	0%	- <sup>2</sup>	13%	- <sup>2</sup>	- <sup>2</sup>
Native Hawaii/Pacific Islanders	52%	-	8%	-	-
Asian/East Asian	43%	-	9%	-	-
Hispanic or Latino	1%	-	14%	-	-
White	1%	-	53%	-	-
Two or more races	4%	-	2%	-	-

*Notes.* (Hawaii Department of Education, 2011a, 2011b, 2011c, 2011d)

<sup>1</sup>To protect the anonymity of the school, some ethnic categories have been collapsed in this table.

<sup>2</sup>No information available at the complex level.

Table 5.9

*Staff Information at Kukui and Hibiscus (2011-2012)*

	<i>Kukui school</i>	<i>Hibiscus school</i>
Administrator	3	3
Teachers	30	37
Paraprofessional & special instructors	4	14
Other administrators	1	3
Total	38	57
Teacher to student ratio	17	16

*Note.* (Hawaii Department of Education, 2011a, 2011b, 2011c, 2011d)

At the time of this research, all four schools met AYP under the NCLB accountability system and their respective state accountability measures. California has had its own accountability system in place since 1999. It manages schools through an API index. This index measures the academic growth of a school based on a growth target until the school reaches the state target of 800, it must then maintain or continue to grow from there (California Department of Education, n.d.-b). Hawaii also has an accountability measure called the *Strive HI Index* (Hawaii

Department of Education, n.d.-a). However, unlike California, Hawaii created and implemented this accountability system in 2013 as part of the amendment to NCLB by taking advantage of the one-time US federal NCLB waiver program (Systems Accountability Office, n.d.). This program gives states the opportunity to create accountability measures that are more appropriate for their student population, rather than adhering to the one-size-fits-all AYP requirement imposed by NCLB. The *Strive HI Index* takes into consider multiple measures: assessment scores; readiness (which includes absenteeism, graduation rate and college enrolment rate); and achievement gaps. Goals are customised to meet the needs of each school complex and school. Based on the index score, schools are placed into one of five different categories: 'recognition', top 5% of the schools; 'continuous improvement', 75% - 85% of the schools; 'focus', next lowest 10%; 'priority', lowest 5%; and 'superintendent's zone'. The last two categories are reserved for schools that require extremely high state interventions. Both schools in the sample fell into the 'continuous improvement' category at the time of the research.

### ***New South Wales.***

The participating schools in Australia were located in a metropolitan area. They are referred to as Bilby School and Koala School. Both are government primary schools serving students from Kindergarten to Year 6. Enrolment at Bilby was around 900 and it was over 250 at Koala. With an ICSEA index below the average value of 1000, both schools are considered socioeconomically disadvantaged. A third of the students at both schools fall into the bottom quarter of the socioeconomic advantage indicator (SEA). As a result, both schools were partners in the National

Partnership for Low Socioeconomic Schools funding program at the time of the present study. Both schools are also among the 1,500 government schools in NSW that have support units catering to students with special education needs from around the geographical area (New South Wales Department of Education, n.d.).

Table 5.10

*Australian Participating School Profiles (2012)*

	<i>Bilby school average</i>	<i>Koala school average</i>
ICSEA value	996	936
SEA bottom quarter distribution	13%	25%
SEA top quarter distribution	13%	25%
Indigenous students	1%	9%
LBOTE students	92%	20%
Recurring income per student	~A\$ 9,300	~A\$14,000

*Note.* (ACARA, n.d.-a)

However, student populations at the two participating schools differed considerably. As shown in Table 5.10, Bilby's student population is largely identified by NAPLAN as LBOTE students, whereas only one in five of Koala's students share the same designation. In contrast, Koala School serves more Indigenous students. Where the economic distribution of Koala is evenly split across all spectrums, over 70% of Bilby's students are in the middle two quartiles. While Bilby's teacher-to-student ratio is similar to the Hawaiian sample at 16:1, Koala is significantly lower at 10:1 and its government funding level is also higher.

Table 5.11

*NAPLAN Mean Scale Scores (2012)*

	Year 3				Year 5			
	Reading		Numeracy		Reading		Numeracy	
	School	<sup>1</sup> Similar schools	School	Similar schools	School	Similar schools	School	Similar schools
Bilby	411	409	417	395	469	492	493	478
Koala	362	387	349	376	449	471	461	461

*Note.* <sup>1</sup>ACARA provides average group statistics for up to 60 schools that share statistically similar backgrounds using the ICSEA value. Schools in each group are not limited to neighbourhood schools but can be found nation-wide. They also include both government and non-government schools. Source: (ACARA, n.d.-a)

The two schools' NAPLAN achievements also differed during the year in which the research was conducted. In reading, Year 3 and 5 students at Bilby performed 7 and 32 points, respectively, below the national average. However, in numeracy, they exceeded the national average by 15 and 5 points, respectively. In contrast, Koala's outcomes were significantly below the national average in both domains for both year levels. In reading, there is a large gap of 56 points in Year 3 and 52 points in Year 5. In numeracy, Year 3 trailed by 53 points and Year 5 by 27 points. Compared to similar schools, Bilby outperformed schools that shared the same characteristics with the exception of Year 5 reading. In comparison, Koala underperformed in every area but in Year 5 numeracy, where they matched schools of similar background (Table 5.11).

#### **Data collection and sources of data.**

Data collection for the qualitative phase involved two stages that took place concurrently in the first half of 2012. In both phases, the researcher was the primary instrument of data collection and analysis. The first stage involved gathering factual information from each participating school's website and respective education accountability office. Where applicable, the following documentation was collected:

national- or state-level assessment results; individual school accountability reports; annual accountability reports from each state's department of education; program improvement plans; and newsletters to parents. Establishing an early understanding of the complex nature of the students at each school, its finances and staffing situation and accountability status helped the researcher to gauge the accountability pressure at each school prior to visiting. However, it is important to note that, because the focus of the research questions is on the belief mechanism and not on effects, public data collected in this stage serve only a descriptive function in the reporting.

The second phase of data collection involved site visits at all six locations. Field work at three out of the four sites in the US and one site in NSW was conducted within a single day at each school. At the remaining two schools, research took place over two days at one school and over six different days at another school. During the site visits, multiple sources of data collected were as follows: (1) semi-structured interviews with teachers, instructional coaches and administrators; (2) real-time observation recordings and notes at grade-level data meetings, or reading and mathematics instructional training sessions; and (3) site-level formal and working documents, and site principal or district leaders' general communications to families. Many research methodologists (Corbin & Strauss, 2008; Creswell & Plano Clark, 2007) consider that having access to extensive data source is a key benefit of qualitative research. Collecting these different data sources occurred in a "recursive and interactive process in which engaging in one strategy incorporates or may lead to subsequent sources of data" (Merriam, 1998, p. 134).

During field work, data collection generally began and ended with an informal meeting with a site administrator or principal. Upon arrival, logistics and interview schedules were discussed and finalised. These discussions sometime resulted in an invitation to join a team meeting or to drill further into the school's data-use process and data documents. The opportunities to observe and gather data in a live setting enabled the researchers to reflect on contexts, decisions and nuances that might not have surfaced during the interviews (Cohen, Manion, & Morrison, 2002). These reflections generally led to deeper questions by the time the wrap-up meeting came around with the site administrator or principal. This holistic process of collecting multiple sources of evidence was intended to create opportunities for data triangulation and verification of the same phenomenon to increase internal research validity (Yin, 1994). The following sections describe each data source in more detail and the data-collection process.

### ***Semi-structured interviews.***

The present study employed semi-structured interviews because this format was suitable for addressing the research questions and issues (Crossley & Vulliamy, 1984). Using current research on data use presented in Chapter 3, key themes that were likely to affect data use formed the basis of the interview guide (Appendix D). These themes (student goals and definition of success; data use motivation; data system; organisation support; and collaboration) were woven into four open-ended questions with sub-questions in the guide to explore data-engagement decisions and practice. A pilot of two interviews was conducted with the original guide, which led

to slight modifications to capture the different views of a principal versus a teacher and an instructional coach, given their differing responsibilities in school.

The standard interview guide provided the necessary framework for the interviews. However, the wording used and the order of the questions presented were not always exactly the same from one respondent to another to provide freedom for “the researcher to respond to the situation at hand, to the emergent worldview of the respondent, and to new ideas on the topic” (Merriam, 1998, p. 74). In addition, the interviews were conducted using an informal and conversational tone designed to keep the participants at ease. In the first two interviews, the participants asked whether the researcher was a teacher or an educator, their question aimed to gauge the level of knowledge the researcher had about school education and schooling. After some reflection, the researcher began to inform participants at the beginning of each interview that the researcher’s knowledge pertained only to the private school sector, where the researcher was working at the time. The intent was to communicate that the researcher was unfamiliar with the government school sector and was therefore eager to learn more from the participants. This declaration also served to assure the participants that they could speak without fear of judgment, as the researcher was outside the government school environment.

While one-on-one interviews were the norm for participant interviews at each site, due to the challenge in scheduling teachers’ time, interviews of small groups of three to four were also conducted at two schools. Interviews ranged from 30 minutes to 120 minutes depending on participants’ availability and enthusiasm for sharing information. On average, the interviews lasted 50 to 60 minutes. All



participants viewed and signed the participant information statement and the participant consent form, which included permission to audio record the interview. In addition, each site principal provided consent for the school to take part in the study. All interviews were conducted at the participants' site during, or at the end of the school day and were audio-recorded using an MP3 player.

### ***Observations.***

In two of the six cases, after discussing the research intent in the pre-interview meetings, a principal at a Hawaiian school and a district data accountability officer at a California district invited the researcher to join their data meetings. At the California school, the researcher was a silent observer all the meetings. At the Hawaiian school, the researcher was invited to ask questions during the grade-level meeting. These opportunities were accepted with gratitude as observation "offers a firsthand account of the situation under study and, when combined with interviewing and document analysis, allows for a holistic interpretation of the phenomenon being investigated" (Merriam, 1998, p. 111). These observations were of significant value because they took place in a setting where data discussions naturally occurred; thus, they provided live evidence (Merriam & Tisdell, 2016) for practices mentioned in the semi-structured interviews. While every effort was made to gain access to similar meetings at other sites, they were simply not part of the normal schedules on the days on which site visits took place. Just as in the semi-structured interviews, the observations focused on finding evidence of data-engagement and its impact on teaching and on student outcomes. Guided by the theoretical framework, the researcher focused on patterns of behaviour that could

affect the depth of data engagement. A code sheet based loosely on the nature of data use, and facilitators and inhibitors of data use discussed in Chapter 3, was used to record evidence that had been proven in prior empirical studies and new behaviours.

Altogether, the researcher was present at five data meetings and one instructional training meeting. Four of the five data meetings were called grade-level meetings and they covered Kindergarten, third and fifth grades. The fifth meeting was a full-school data meeting with the district officials. The grade-level meetings were attended by all teachers for the specific grade and one or more instructional coaches, depending on the school. These meetings occurred during instructional hours and lasted for an hour at one school and an hour and forty-five minutes at another. The goals of these meetings were the same: to review student progress using data; to evaluate progress against goals; to determine the next milestones; to discuss the effectiveness of the ongoing programs and initiatives to support academic growth; and to share success, challenges and strategies. The formats differed between the two schools and they are discussed in Chapter 7 alongside the findings.

The full-school data meeting was part of the district review process. The principal and the entire teaching staff were present at this meeting, which was held after school and lasted roughly an hour and a half. Other meeting attendants included the district superintendent, deputy superintendent, district data accountability officer, and a principal from another school within the district. In this meeting, each grade-level team presented its respective student goals, progress, successes at that time and strategies for moving forward to meet the annual goals.

Finally, the two-hour reading instructional training meeting was a district-wide meeting for all third-grade teachers within the same school district. This meeting was held after school and facilitated by two teachers within the district who also took on the role of reading coaches for the district. Audio recording was used at these observations and descriptive field notes were taken.

### ***Documents.***

According to Yin (1994), “the most important use of documents is to corroborate and augment evidence from other sources” (p. 81). Documents could provide ‘clues’ (Yin, 1994, p. 81) to inferences made from other sources, and “the advantage of being in the language and words of the participants” (Creswell, 2008, p. 231). To understand the depth of data-engagement at each school site, relevant artefacts were either collected or examined during site visits. Examples of the primary documents gathered included: data-meeting agendas; data organisation and analysis methods; data analysis and discussion forms; monthly and quarterly check-in forms; data-engagement meeting protocols; and professional development plans, among other materials. Where permitted, copies of the documents were made, where this was not possible, permission to review the document on site was sought. A significant portion of the data gathered from the field visits was a result of “accidental uncovering of valuable data” (Merriam & Tisdell, 2016, p. 174) through the interviews and observations. All the documents gathered were nonreactive, grounded in the context of this study (Merriam & Tisdell, 2016), and have existed and were in use prior to the researcher’s visit. While every effort was made to access similar documentation from every participating school to minimise variation in the

data (Creswell & Plano Clark, 2007), each site's data management and representation systems, as well as data priorities, differed. Therefore, it was not always possible to achieve the data standardisation ideal. Finally, after each site visit, field notes were created to capture noteworthy aspects of the interviews or observations and they became part of the documentation artefacts.

### **Data analysis.**

The qualitative analytical approach was inductive and comparative, inferred from the rich dataset containing the transcribed interviews, transcribed data meetings in their natural settings, field notes, and documents collected during field work. Not only were comparisons a goal of this overall study, but they are "implicit rather than explicit" (Stake, 2000, p. 24) of the case study approach. The overriding analytical technique focused on discovery and explanation-building (Howe, 2012; Yin, 1994). The multiple sources of data were never considered independently but converged in an attempt to understand the overall trends that explained the impact on student achievement. Through this analytical approach, similarities and differences within and across cases can be drawn regarding the policy impact on teachers' intentions to carry out data practice and the impact of the practice or lack thereof on student outcomes. While themes were identified during the analysis process, the focus was on understanding the case holistically as they are 'subordinate' to the case (Stake, 2000).

All three data sources were analysed using QSR Nvivo software and coded using best practice recommended by multiple researchers (Merriam & Tisdell, 2016; Yin, 1994). This began with data consolidation, followed by data reduction, then data triangulation, and finally data interpretation. Each case was first analysed as a stand-

alone case for within-case analysis. This enabled contextual factors to surface (Merriam & Tisdell, 2016). When all individual cases were analysed, cross-case analysis began. In this process, similarities and differences were identified and mapped following the same process as in the individual cases.

***Analysis of the semi-structured interviews.***

Transcription of the interviews was performed by TranscriptionHub.com and the transcriptions were done verbatim. On completion, the researcher reviewed every transcription against the audio recording to ensure accuracy. Inaccurate transcriptions were sent back to TranscriptionHub.com and the process repeated. The researcher listened to all audio recordings and reviewed them against the transcriptions twice. Through the transcription review process, the researcher also became familiar with the content and the key and unique attributes of each interview and observation.

Coding was divided into to three strands. The first related to the nature of data use. This was guided by existing literature discussed in Chapter 3. The second strand pertained to personal and collective beliefs regarding data engagement. This aspect of coding was grounded by the combined theoretical framework of efficacy and the theory of planned behaviour. The final strand concerned the resulting benefits of participants' decisions, and these codes were created inductively. Initial broad categories were created to consolidate and organise common themes that captured recurring patterns across the interviews. For example, the first strand included codes such as 'diagnostic purpose', 'resource allocation', and 'gaps identification', etc.; the second strand 'subjective norm', 'attitude', and

'intentionality', etc.; and the third strand 'collaboration', 'goal-oriented', 'high expectation', and more. Where appropriate, sub-codes were created to accommodate emergent insights that did not fit into the earlier coding schemes. Altogether, 82 codes and sub-codes emerged. Where possible, codes were created using terms that the participants had used. Each interview was analysed as an individual case but labelled as one of these categories: teacher, coach or administrator. The intent was to capture similarities and differences that might have arisen from these respective roles and responsibilities. Finally, data reduction began with identifying patterns, similarities and differences from the coding scheme to form concept maps. Codes that did not cut across too many interviews were either recoded into an existing category or became a subcategory.

***Analysis of the grade/year-level and school-level meeting observations.***

Since the meetings were audio recorded, a similar procedure described in the previous section for transcription, review, theme identification and coding applied to these observations. For these observations, the field notes played an important complementary role. Technically, there is one overarching process and one set of goals for each grade-level meeting within a school, as it is a meeting process adopted by the school. However, just as policy implementations differ from school to school depending on the team involved (Fulcher, 1989), grade-level meetings could also differ significantly from grade to grade for the same reason. The field notes were particularly sensitive to these contextual nuances, which included: how teachers in a grade level engaged with each other and with the instructional coach and vice versa,

their body language, and their level of enthusiasm and collaborative spirit – aspects that the audio tapes of spoken dialogue might not have captured.

#### **Validity and reliability.**

The study followed a few recommended strategies to ensure the validity and reliability of the analysis. First, by organising the field visit schedules to begin and end with the site administrator or principal, the researcher had the opportunity to clarify and verify events, observations, goals, and claims made during the teacher interviews. Because participant observation is, by nature, subjective (Merriam & Tisdell, 2016) and research is advocative (Stake, 2010), extra care was taken to weed out researcher bias. For example, the final interview at each site was intentionally structured with the site administrator or principal to validate the researcher's interpretations derived during the data-collection process. Secondly, having multiple data sources collected through three different means – interview, observation, and documentation – in and of itself is an internal validity strategy (Merriam & Tisdell, 2016). These authors state that this strategy enables the research to triangulate interpretations beyond a single data source and this increases the credibility and quality of the findings. Furthermore, interpretations were further checked for corroboration with existing literature. Finally, during coding and theme identification and categorisation, an audit trail was created to ensure accuracy and clarity of the coding and recoding process.

#### **Limitations.**

The first limitation in this qualitative study lies in the reality that the findings are not generalisable. This limitation arises from the small number of cases, as well as the nature of case studies itself. Because qualitative studies are not context-free,

they fail to meet the definition of generalisation described by Lincoln and Guba (2000), “generalisations are assertions of enduring value that are context-free” (p. 27). Nonetheless, the findings still offer valuable insights, precisely because they are context-driven. One of the strongest critiques of the data-driven accountability policies in both countries is their one-size-fits-all perspective without regards to local contexts (Australian Education Union, 2013; US Department of Education, 2013). The cases in this study offer the much-needed local contexts, albeit not satisfying the generalisation requirement. The second limitation pertains to the lack of permission to interview or to observe teachers in every grade/year at each site. This would have provided a more holistic view of the school’s success in implementing data-driven instruction. In spite of that, the data collected through various means was sufficient for the researcher to explore the current data practice, explain the belief mechanism and subsequent decisions regarding data use.

## **Conclusion**

This chapter has presented the mixed methods approach employed to explore the three overriding research questions. Quantitative analysis of large-scale student assessment data to evaluate achievement progress was conducted concurrently with semi-structured qualitative interviews and observations at six primary school sites in three locations across two countries. The analysis of assessment data evaluated variability in achievement by student subgroups with additional explanatory variables where data were available. The qualitative case studies, on the other hand, explored the macro-level perspective along with a micro-level view of the dynamics of teacher and school responses to the transparency and



accountability policies vis-à-vis assessment data. Mixed methods research enables this study to meet the data-driven research standard required by policy makers for education research (Raudenbush, 2005); to explore the complex relationships between policy and practices (Ball, 2005b; Crossley & Vulliamy, 1984); and to draw out the cross-cultural context valued in comparative studies (Crossley & Watson, 2003). When both quantitative and qualitative techniques work together, there is increased potential for explanation-building (Howe, 2012) that may reflect critical insights into the public policy process (Yin, 1994) without sacrificing realism (B. Johnson & Christensen, 2008). Realism is a particularly important element in research relating to disadvantaged students due to their multiple challenges (Council for Learning Disabilities, n.d.). Finally, mixed research methods afford the opportunity to triangulate, converge and corroborate findings (Hesse-Biber, 2010; B. Johnson & Christensen, 2008; B. Johnson & Onwuegbuzie, 2004), helping to link one element of the policy implementation, data practices, to the learning experience of students and their academic achievements.

## **Chapter 6 Quantitative Findings: Academic Progress in California**

Chapters 6 and 7 report the quantitative findings of disadvantaged students' progress on external assessments between 2008 and 2013. As the datasets and analytical procedures differ for NAPLAN and the CST, their findings will be presented separately. However, for both datasets, analyses focused on finding answers to the following research question.

- (1) What have the test-based accountability policies in Australia and California accomplished in the area of assessment inclusion and achievement of disadvantaged students?
  - What are the general trends in academic progress?
  - How have disadvantaged students fared compared to advantaged students?

This chapter presents reading and mathematics progress on the CST for students in Grades 3 and 5 in the two California counties where qualitative research for the present study took place. The quantitative analysis examined progress between 2008 and 2013 by student subgroups, achievement gaps between student subgroups, and, lastly, progress towards assessment inclusion. This chapter begins with an explanation of the analytical methods and continues with descriptive information about the dataset before presenting the analytical findings and conclusion. To capture and present the variations in the findings, descriptive information, charts and tables are used throughout this chapter. The Australian findings follow in Chapter 7.

## Data and Statistical Procedures

The analysis aimed to explore whether disadvantaged students in two Californian counties experienced similar trends to those reported at the national level, as discussed in the literature in Chapter 3. These key trends included: (1) moderate growth for mathematics; (2) near-negligible growth for reading; (3) some disadvantaged groups (namely Black and Hispanic students) improved at a faster rate than White students, while other disadvantaged groups trailed; (4) little movement in the achievement gap; and (5) a general improvement in the inclusion of disadvantaged students in standardised assessment. The following student subgroups were evaluated in the current analysis: economically advantaged and disadvantaged students; English learners and English only students; students with disability and without disability; students whose parents only had high-school or lower education and those whose parents had an associate or higher degree; and students who are ethnically Asian, Black, Hispanic and White.

### **Statistical procedures.**

The quantitative analysis for this chapter adopts a similar statistical procedure to that used by the NCES (National Center for Education Statistics, n.d.). The analysis consisted of a series of one-tailed *t*-tests, which provide statistics to help determine whether: (1) students in each subgroup progressed between 2008 and 2013; and (2) whether the gaps between disadvantaged and advantaged subgroups changed significantly between 2008 and 2013. To evaluate progress between 2008 and 2013 for each independent student subgroup, for example Hispanic students, with respect to the mean scale scores and proficiency rate, one sample *t*-tests were conducted. These tests compared the means of all schools for

one subgroup in 2008 to the means in 2013. To compare the academic gap between two student subgroups outlined in Table 6.1, for example, White versus Black students, *t*-tests for partially overlapping groups were employed. In these tests, the difference between the means of the comparative groups in 2013 is compared with the difference of their means in 2008. Only the schools that had data in both 2008 and 2013 were included in the achievement gap analysis, hence, a smaller population than the analysis for a single subgroup described above. Due to the reduction in observations, results from the *t*-tests for students with disability are omitted from the gap analysis as the remaining observations were too small to provide reliable statistics. However, the absolute difference between students with disability and students without disability is left in the achievement gap tables to illustrate the size of the gap between these two groups. Nonetheless, conclusions should not be drawn about its statistical significance.

To control for the proportion of falsely rejected hypotheses due to multiple comparisons for the subgroups previously described, the ‘false discovery rate’ procedure (National Center for Education Statistics, n.d.) was applied in this analysis to increase the power of the statistical tests. In this procedure, adjustments were made to account for the large number of *t*-tests while still holding the statistical significance level for the set of comparisons at  $p < 0.05$ . Through this procedure, the *p* values get larger as the number of *t*-tests increases to control for falsely rejected hypotheses. Separately, the number of observations in each subgroup also impacts the statistical significance of the absolute difference being observed because the *t* critical value gets larger when the number of observations decreases. To be statistically significant, the absolute margin or difference between two comparative

means or proficiency rates must produce a large enough  $t$  statistic to overcome the smaller number of observations. Therefore, while the observed or absolute mean score and percentile differences between groups, or between calendar years, might appear significant, only statistically significant results indicated by an alpha level of less than 0.05 can accurately evaluate the progress made by an independent student subgroup, or gap movement between two subgroups over time.

**Data summary.**

The analysis consists of data from two Californian counties, referred to as Almond County and Walnut County in the report. The unit of analysis is the school, as only school averages were accessible in the department of education database. In 2008, the base year for the dataset, both counties served between 230–250 schools. By 2013, Almond had two fewer schools and Walnut had added 16 more schools. The total number of schools differed slightly from year to year due to school shut down as part of the NCLB mandate when a school failed to meet AYP repeatedly. Tables 6.1 and 6.2 provide an overview of school-reported data for each of the subgroups in the analysis.

In subcategories such as ‘with disability’, there are more schools serving students with disability than are represented by the figures. However, the minimum ten-student reporting threshold, discussed previously, means that schools with fewer than ten students with disability in a particular grade do not report data for this particular subgroup in that grade. The same is true for the ‘Black’ subcategory in Walnut County due to the small percentage of Black students it serves. Lastly, the changes between the two calendar years reflect student mobility from or to the counties.

Table 6.1

*Percentage of Schools Represented by Subgroup in Grade 3*

<i>Student subgroups</i>	<i>Almond county</i>		<i>Walnut county</i>	
	<i>2008</i>	<i>2013</i>	<i>2008</i>	<i>2013</i>
<i>Ethnic groups</i>				
Black	37%	28%	0%	0%
White	48%	46%	54%	47%
Asian	35%	39%	59%	59%
Hispanic	60%	62%	70%	72%
<i>Other designations</i>				
Economically disadvantaged	71%	77%	71%	72%
Economically advantaged	82%	73%	86%	76%
English language learners	64%	59%	75%	74%
English Only	92%	91%	91%	90%
With disability	7%	4%	8%	5%
No disability	98%	98%	98%	97%
High-school degree or lower	40%	55%	59%	54%
Associate degree or higher	58%	77%	83%	85%

Table 6.2

*Percentage of Schools Represented by Subgroup in Grade 5*

<i>Student subgroups</i>	<i>Almond county</i>		<i>Walnut county</i>	
	<i>2008</i>	<i>2013</i>	<i>2008</i>	<i>2013</i>
<i>Ethnic groups</i>				
Black	37%	29%	0%	0%
White	45%	41%	56%	48%
Asian	39%	38%	57%	57%
Hispanic	59%	65%	73%	71%
<i>Other designations</i>				
Economically disadvantaged	75%	79%	73%	71%
Economically advantaged	79%	75%	88%	75%
English language learners	50%	36%	64%	53%
English only	90%	89%	92%	61%
With disability	8%	4%	14%	3%
No disability	96%	98%	99%	97%
High-school degree or lower	52%	43%	61%	57%
Associate degree or higher	67%	63%	85%	81%

## Reading Achievement Trends

In Almond County, all student subgroups in Grade 5 (Table 6.4) made larger absolute gains in scale scores and proficiency rates between 2013 and 2008 than did students in Grade 3 (Table 6.3). The majority of the groups' gains in Grade 5 were statistically significant, with an average gain of 17 mean scale points and 9% increase in proficiency standard in 2013. In comparison, only the following groups in Grade 3 demonstrated statistically significant gains in proficiency rate: Hispanic, economically advantaged and disadvantaged, students without disability, and students from families with higher education. While all advantaged subgroups in Grade 3 showed significant growth in scale scores, Black students, ELLs, students with disability, and students from a family with less education made no significant progress. Among disadvantaged subgroups, economically disadvantaged and Hispanic students made statistically significant gains in both score and proficiency rate. Economically disadvantaged students produced one of the largest proficiency growths in both grades. In Grade 5, students from a less-educated family also saw significant gains of 15% in proficiency rate.

In both grades at Almond County, a majority of White and Asian students achieved proficiency and statistically significant gains in scale scores. Among other background characteristics, economically advantaged students, and students from better-educated families were the highest performers. Over three-quarters of these four advantaged groups in Grade 5 achieved proficiency in reading; in Grade 3, 60–70% of the same groups reached proficiency.

Table 6.3

*Grade 3 Reading Means and Proficiency Rates at Almond County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008</i>	<i>2013</i>	<i>Diff.</i>	<i>2008</i>	<i>2013</i>	<i>Diff.</i>
Black	311.6	317.6	6.0	23.0	28.1	5.1
White	369.9	383.9	14.0*	64.2	71.3	7.1
Asian	371.8	387.0	15.2*	66.0	70.4	4.4
Hispanic	311.9	324.4	12.5*	22.5	31.6	9.0*
Economically disadvantaged	311.0	326.5	15.5*	21.8	32.3	10.5*
Economically advantaged	354.2	374.3	20.1*	52.6	64.6	12.0*
English language learners	307.7	303.2	-4.5	18.9	15.3	-3.6
English only	343.9	357.4	13.5*	45.7	53.5	7.8
With disability	334.7	338.9	4.2	39.0	35.9	-3.1
No disability	339.8	352.8	13.1*	42.3	49.9	7.6*
High-school degree or lower	319.6	324.3	4.6	27.1	31.3	4.2
Associate degree or higher	356.0	368.5	12.5*	54.3	61.1	6.8*

Note. \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Table 6.4

*Grade 5 Reading Means and Proficiency Rates at Almond County*

<i>Student subgroups</i>	<i>Mean (Scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008</i>	<i>2013</i>	<i>Diff.</i>	<i>2008</i>	<i>2013</i>	<i>Diff.</i>
Black	329.5	339.8	10.3	33.5	42.5	9.0
White	380.2	402.9	22.7*	73.7	83.9	10.2*
Asian	385.8	402.6	16.8*	76.5	82.6	6.1
Hispanic	332.6	351.9	19.3*	35.5	50.9	15.4*
Economically disadvantaged	331.5	351.4	19.9*	34.6	51.6	16.9*
Economically advantaged	368.1	390.9	22.8*	63.7	76.0	12.3*
English language learners	316.1	312.3	-3.8	20.5	15.4	-5.1
English only	356.6	375.4	18.7*	54.5	65.8	11.3*
With disability	339.0	360.7	21.7	39.5	58.8	19.3
No disability	356.4	373.7	17.2*	53.8	64.5	10.7*
High-school degree or lower	332.6	352.0	19.4*	35.7	50.9	15.2*
Associate degree or higher	366.6	388.2	21.7*	62.2	75.3	13.1*

Note. \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Between the two grades, a larger percentage of students in Grade 5 achieved proficiency standard, and this is true across every student subgroup (Figure 6.1). The exception is for ELLs, of whom only 15% achieved proficiency in both grades in 2013.



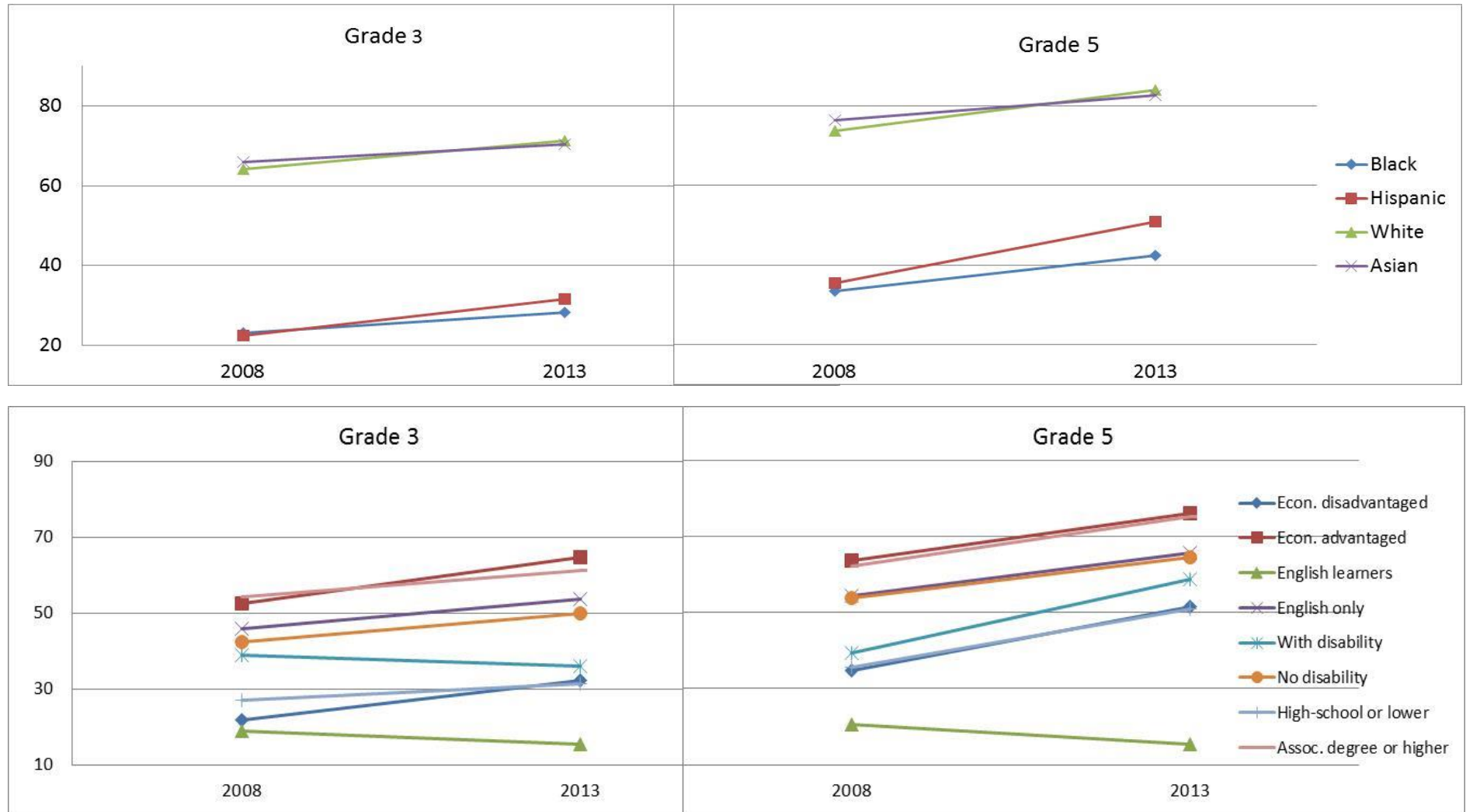


Figure 6.1. Reading Proficiency between Grade Levels at Almond County

Their growth also trended down, although the change is statistically insignificant. Nonetheless, the declining trend is alarming. Excluding ELLs, an average of 16% more students reached proficiency in Grade 5 than in Grade 3. The groups with the largest difference between the grades were: students with disability (23%), students from a less-educated family (20%), economically disadvantaged students (19%), and Hispanic students (19%). This trend suggests that disadvantaged groups' achievement not only grew significantly but had continued to make sizable progress as students aged.

At Walnut County, most of the subgroups in both grades made significant progress on scale score and proficiency rate as noted in Tables 6.5 and 6.6. The exceptions were ELLs across both grades, and students with disability in Grade 3. Both groups were also the lowest-performing subgroups. Similar to Almond County, the absolute size of growth in Grade 5 at Walnut County also exceeded Grade 3's growth. On average, Grade 5 has 12% more students meeting proficiency when excluding ELLs whose achievement actually slipped between Grade 3 and 5. In comparison, Grade 3 students lodged an average 9% gain in proficiency rate.

Fifth-grade disadvantaged student groups at Walnut County demonstrated the strongest progress, namely: Hispanic students, students with disability, economically disadvantaged students and students with less-educated parents. All increases were statistically significant. Also worth noting is the magnitude of growth achieved by students with disability, which outpaced all other groups by a minimum of two-fold in score and proficiency rate. In contrast, the results of students with disability in Grade 3 remained unchanged between the two calendar years. Hispanic

students in Grade 5 also made twice the proficiency gains of other ethnic groups. In

Grade 3, all ethnic groups made similar significant progress in proficiency.

Table 6.5

*Grade 3 Reading Means and Proficiency Rates at Walnut County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008</i>	<i>2013</i>	<i>Diff.</i>	<i>2008</i>	<i>2013</i>	<i>Diff.</i>
Hispanic	315.1	331.8	16.7*	25.5	35.5	9.9*
White	374.0	389.4	15.4*	67.1	75.1	8.0*
Asian	378.9	399.2	20.4*	69.9	79.8	9.9*
Economically disadvantaged	315.6	332.8	17.1*	25.9	36.6	10.7*
Economically advantaged	365.4	387.4	22.0*	60.8	73.2	12.4*
English language learners	315.2	325.2	10.0*	24.8	31.0	6.2
English only	356.3	371.9	15.5*	53.9	63.1	9.2*
With disability	321.8	345.0	23.2	32.6	36.2	3.6
No disability	350.9	367.1	16.2*	50.1	59.5	9.3*
High-school degree or lower	315.0	327.6	12.5*	25.3	33.1	7.8*
Associate degree or higher	362.7	376.9	14.2*	58.5	66.4	7.9*

Note. \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Table 6.6

*Grade 5 Reading Means and Proficiency Rates at Walnut County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008</i>	<i>2013</i>	<i>Diff.</i>	<i>2008</i>	<i>2013</i>	<i>Diff.</i>
Hispanic	330.8	351.3	20.5*	35.4	51.2	15.8*
White	385.8	403.4	17.6*	78.3	85.2	6.9*
Asian	390.7	412.1	21.3*	78.3	86.1	7.8*
Economically disadvantaged	332.2	353.5	21.3*	36.1	52.3	16.2*
Economically advantaged	378.3	401.9	23.6*	71.5	82.9	11.4*
English language learners	317.4	320.2	2.8	22.1	23.6	1.5
English only	367.1	386.3	19.3*	62.8	73.6	10.7*
With disability	317.3	367.0	49.7*	25.8	58.4	32.7*
No disability	366.0	383.3	17.3*	61.4	70.8	9.3*
High-school degree or lower	332.8	351.4	18.6*	37.2	50.7	13.5*
Associate degree or higher	375.6	393.8	18.2*	69.3	78.5	9.2*

Note. \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

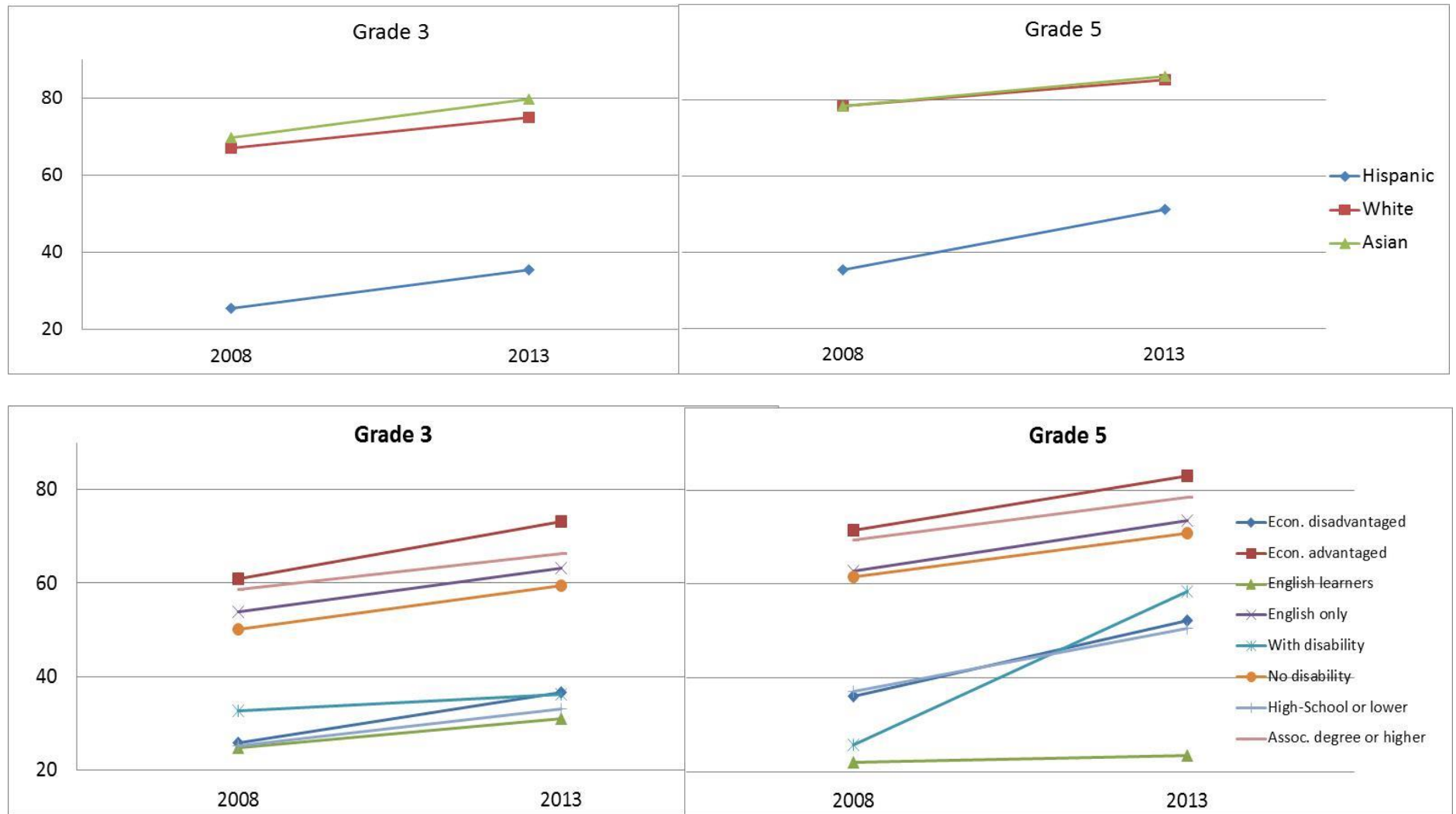


Figure 6.2. Reading Proficiency between Grade Levels at Walnut County

## Mathematics Achievement Trends

As can be seen in Tables 6.7–6.10, third graders appeared to have a stronger foundation in mathematics than in reading, based on the higher percentage of students reaching proficiency standards in 2008 across counties. Average scale score gains were larger in mathematics than in reading across counties and grades.

However, increases in proficiency rate were relatively similar between reading and mathematics. For the most part, subgroup-related patterns in mathematics followed those found in reading in both counties. Fifth graders achieved larger gains than third graders, and Walnut County enjoyed larger gains than Almond County. ELLs were also the lowest-performing group in mathematics and students with disability followed closely behind. A difference worth noting is the proficiency rate drop of 11% at Almond and 20% at Walnut between Grades 3 and 5 among ELLs students.

Table 6.7

*Grade 3 Mathematics Means and Proficiency Rates at Almond County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008</i>	<i>2013</i>	<i>Diff.</i>	<i>2008</i>	<i>2013</i>	<i>Diff.</i>
Black	337.4	343.2	5.7	40.4	41.5	1.0
White	419.9	442.0	22.1*	80.5	84.8	4.3
Asian	447.3	469.8	22.5*	87.1	88.8	1.7
Hispanic	353.6	369.4	15.9*	49.9	55.4	5.5
Economically disadvantaged	350.1	371.5	21.4*	47.2	55.5	8.3*
Economically advantaged	405.2	437.3	32.1*	71.8	80.4	8.6*
English language learners	356.7	350.0	-6.7	50.0	45.1	-4.9
English only	384.8	406.3	21.6*	62.2	68.3	6.0
With disability	370.0	383.9	13.9	58.3	56.7	-1.6
No disability	387.8	408.4	20.6*	64.0	68.9	4.9
High-school degree or lower	355.6	364.6	9.0	48.4	51.7	3.3
Associate degree or higher	410.3	428.0	17.8*	74.5	77.0	2.6

*Note.* \*  $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Table 6.8

*Grade 5 Mathematics Means and Proficiency Rates at Almond County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008</i>	<i>2013</i>	<i>Diff.</i>	<i>2008</i>	<i>2013</i>	<i>Diff.</i>
Black	330.3	343.8	13.5	34.3	43.6	9.2
White	406.3	437.9	31.6*	71.2	82.3	11.2*
Asian	442.3	466.9	24.6*	82.2	88.2	6.0
Hispanic	341.2	367.8	26.6*	40.5	56.0	15.4*
Economically disadvantaged	343.4	368.6	25.2*	41.0	56.1	15.2*
Economically advantaged	395.8	429.0	33.2*	65.1	77.5	12.3*
English language learners	332.0	325.4	-6.7	35.6	33.5	-2.1
English only	372.6	398.1	25.5*	54.3	65.8	11.5*
With disability	354.3	384.5	30.2	43.6	63.4	19.9
No disability	380.6	404.1	23.5*	57.9	68.5	10.6*
High-school degree or lower	342.1	365.4	23.3*	40.9	55.6	14.7*
Associate degree or higher	392.7	423.6	30.9*	63.0	75.0	12.0*

*Note.* \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

On the whole, the mathematics proficiency rate at Almond (Tables 6.7–6.8) remained flat in Grade 3 despite statistically significant growth in scale scores across a few subgroups. Only those achieved by economically advantaged and disadvantaged students translated to significant proficiency growth in Grade 3. Grade 5 disadvantaged student groups saw double-digit growths of 15% compared to an average of 11% for the advantaged groups. As for Walnut County, the mathematics trends (Tables 6.9–6.10) are nearly identical to their trends for reading, the only difference being larger absolute scale score change between the two calendar years as compared to those found for reading.

Table 6.9

*Grade 3 Mathematics Means and Proficiency Rates at Walnut County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008</i>	<i>2013</i>	<i>Diff.</i>	<i>2008</i>	<i>2013</i>	<i>Diff.</i>
Hispanic	352.4	378.6	26.3*	48.8	59.8	11.0*
White	425.1	451.6	26.5*	81.8	87.2	5.4*
Asian	457.2	488.9	31.7*	89.6	93.8	4.2*
Economically disadvantaged	356.9	385.3	28.4*	51.1	61.9	10.8*
Economically advantaged	423.2	457.8	34.6*	78.8	86.2	7.4*
English language learners	365.3	384.1	18.7*	54.6	60.3	5.8
English only	402.7	430.3	27.6*	70.6	77.2	6.6*
With disability	351.8	395.4	43.6	46.4	63.5	17.1
No disability	404.0	432.6	28.6*	70.4	77.9	7.5*
High-school degree or lower	357.0	381.7	24.7*	51.1	60.7	9.6*
Associate degree or higher	419.5	443.8	24.3*	78.0	82.4	4.4*

Note. \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Table 6.10

*Grade 5 Mathematics Means and Proficiency Rates at Walnut County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008</i>	<i>2013</i>	<i>Diff.</i>	<i>2008</i>	<i>2013</i>	<i>Diff.</i>
Hispanic	335.4	366.2	30.9*	37.3	53.1	15.8*
White	416.4	446.2	29.8*	75.5	83.6	8.1*
Asian	448.5	487.5	38.9*	83.0	91.9	8.9*
Economically disadvantaged	341.9	375.4	33.5*	40.1	57.2	17.1*
Economically advantaged	413.7	453.2	39.5*	71.9	83.2	11.3*
English language learners	329.2	341.7	12.5	33.3	40.3	7.0
English only	387.6	419.3	31.7*	60.7	72.6	11.9*
With disability	319.0	394.9	75.9*	28.9	63.5	34.6*
No disability	395.6	424.9	29.3*	63.4	73.6	10.2*
High-school degree or lower	344.4	373.2	28.9*	40.6	55.5	14.9*
Associate degree or higher	408.3	438.6	30.3*	70.4	79.8	9.4*

Note. \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

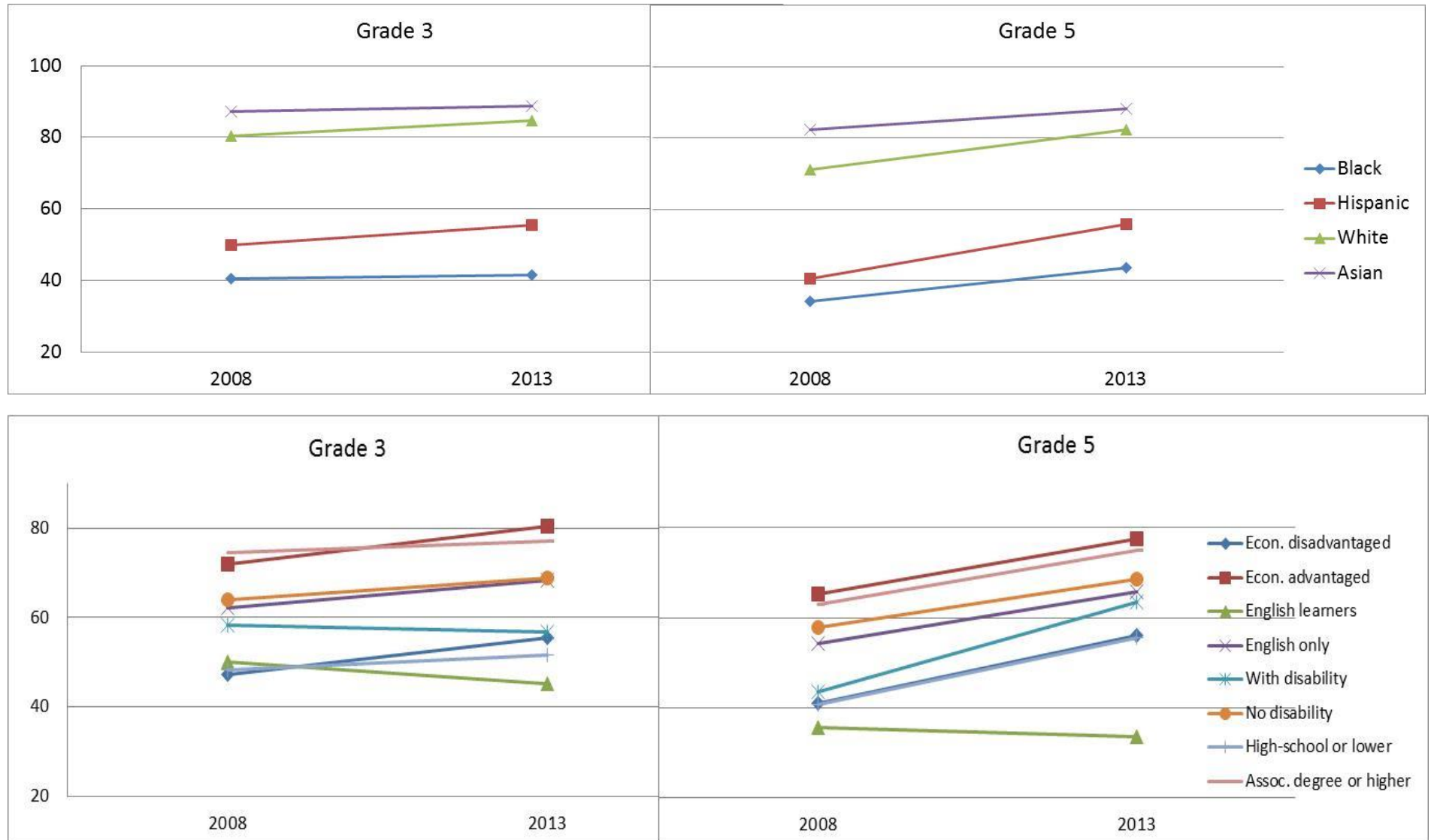


Figure 6.3. Mathematics Proficiency between Grade Levels at Almond County



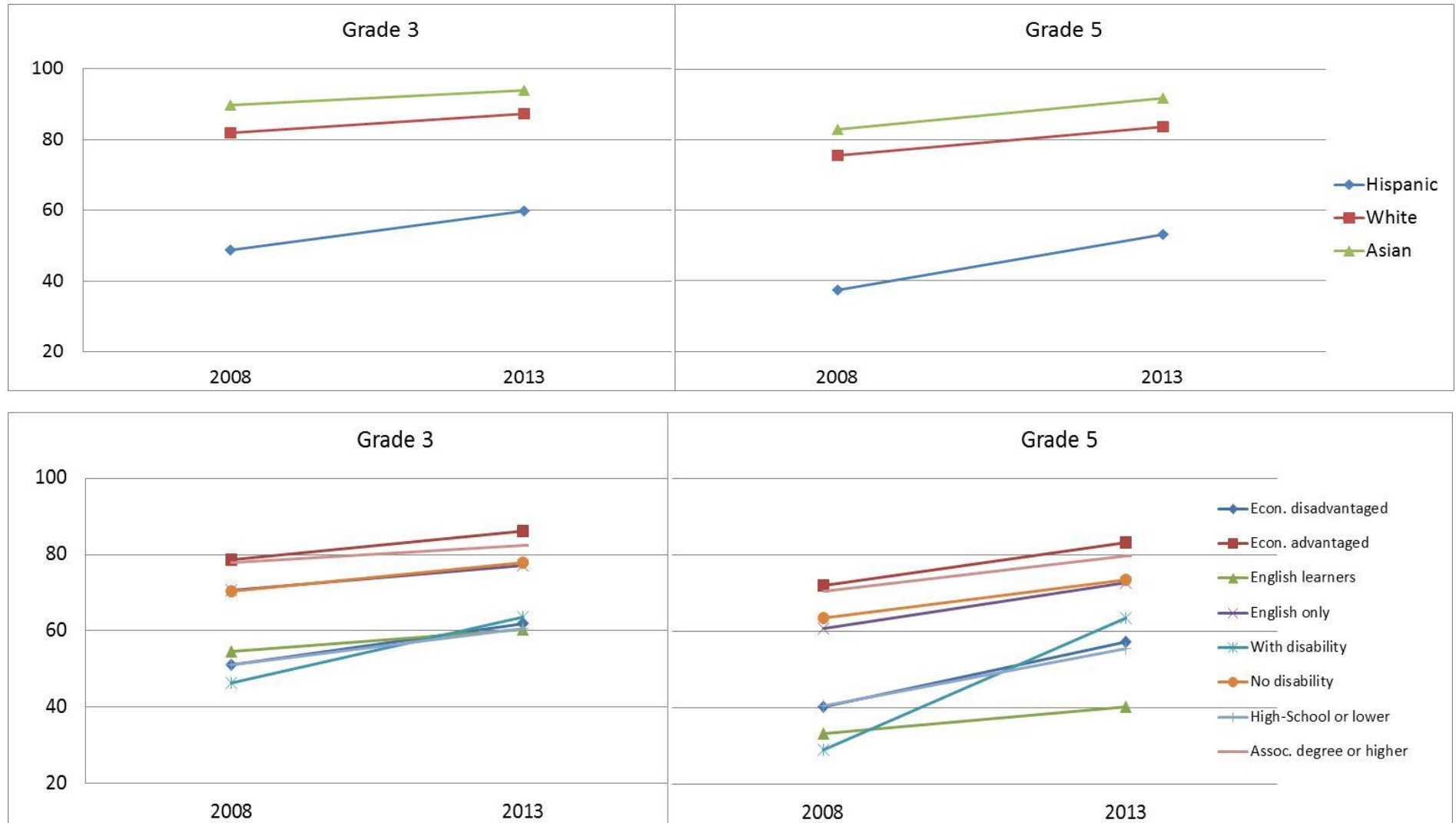


Figure 6.4. Mathematics Proficiency between Grade Levels at Walnut County

In summary, absolute progress based on scale scores has been made across counties, grade levels and domains. However, growth was consistently more modest in Grade 3 than in Grade 5, weaker in reading than in mathematics, and slower in Almond County than in Walnut County. In reading, average gains of 11 and 17 points in third and fifth grades, respectively, across all groups were achieved at Almond. At Walnut, the average gain was 17 points in third grade and 21 points in fifth grade. In mathematics, the average increases in scale score at Almond were 16 and 24 points in third and fifth grades, respectively. At Walnut, third grade gained an average of 29 points and fifth grade an average of 35 points. Most of the subgroups' scale score increases between 2008 and 2013 were statistically significant, except those for three disadvantaged subgroups: Black students, students with disability, and ELLs. In Grade 3, positive changes in scale scores translated to small single-digit growth in proficiency rates for most student subgroups across domains and counties. Grade 5 students fared better with higher single-digit to mid-teen increases in proficiency rate across both domains and counties. Amongst ELLs, however, a large majority, sometimes over 80% had still not met proficiency requirements. They suffered the lowest proficiency rates amongst students of all backgrounds.

Two disadvantaged groups stood out in particular: Hispanic and economically disadvantaged students. They made some of the largest and most significant gains amongst advantaged and disadvantaged subgroups. These general findings corroborate the national trends found in the NAEP discussed Chapter 3. The significance of Hispanic students' gains in the current study has been demonstrated in other studies and so too has the lack of progress by ELLs and students with disability. The only difference between the current findings and other studies is the

results of Black students, whose progress was insignificant in the current study. This difference might have been affected by the overrepresentation of ELLs in these two counties, which could have changed the dynamic of resource allocations in these counties compared to states where Black students are the predominant population.

## **Progress on Achievement Gap**

### **Reading.**

Despite some significant gains among a few disadvantaged student groups in mean scale scores and proficiency rates, achievement gaps between advantaged and disadvantaged students remained wide and unyielding since 2008, as highlighted in Figures 6.1–6.4. In Almond County (Tables 6.11 – 6.12), some proficiency gaps have widened, while others narrowed in absolute terms. However, only one change is statistically significant: the gap between ELLs and English speakers. The proficiency gap between these groups widened significantly by 11% and 16% in Grades 3 and 5, respectively, between 2008 and 2013. This is not surprising given the significant drop in ELL performance presented in the last section. Between these two subgroups, the proficiency gap in 2013 stood at 38% in Grade 3 and 50% in Grade 5.

Other large proficiency gaps at Almond are found between ethnic groups, with the highest at 43% in Grade 3 and 41% in Grade 5 between White and Black students. One bright spot is the proficiency gap between students with and without disability in 2013. Standing at 13% and 6% in Grades 3 and 5, respectively, these gaps are relatively small compared to most other group differences. Hispanic and Black students achieved relatively similar reading outcomes in 2008 but the gap between

them widened by 2013 as a result of the progress made by Hispanic students.

However, the change was not large enough to be statistically significant.

Table 6.11

*Grade 3 Reading Achievement Gaps between Groups at Almond County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>
White–Black	58.3	66.3	8.0	41.2	43.2	2.0
Asian–Black	60.2	69.4	9.3	43.0	42.2	-0.8
Hispanic–Black	0.3	6.9	6.5	-0.5	3.4	3.9
White–Hispanic	58.0	59.5	1.5	41.7	39.8	-1.9
Asian–Hispanic	59.8	62.9	2.7	43.5	38.8	-4.7
Economically advantaged–Disadvantaged	43.2	47.9	4.7	30.7	32.3	1.6
English learner–English only	36.2	54.1	17.9*	26.8	38.2	11.4*
With disability–No disability	5.1	13.9	NA <sup>1</sup>	3.3	14.0	NA
High-School/lower–Associate/higher degree	36.4	44.3	7.9	27.1	29.8	2.6

*Note.* <sup>1</sup>As stated in the *statistical procedures* sub-section in this chapter, the statistical significance for this group is not available due to the very small number of observations when samples between the two calendar years were matched.

\* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Table 6.12

*Grade 5 Reading Achievement Gaps between Groups at Almond County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>
White–Black	50.7	63.1	12.5	40.1	41.4	1.2
Asian–Black	56.3	62.8	6.5	42.9	40.0	-2.9
Hispanic–Black	3.1	12.1	9.0	2.0	8.4	6.4
White–Hispanic	47.5	51.0	3.5	38.1	33.0	-5.1
Asian–Hispanic	53.1	50.7	-2.5	40.9	31.7	-9.3
Economically advantaged–Disadvantaged	36.6	39.4	2.8	29.0	24.4	-4.7
English learner–English only	40.5	63.1	22.9*	34.0	50.4	16.4*
With disability–No disability	17.4	13.0	NA	14.3	5.7	NA
High-school/lower–Associate/higher degree	33.9	36.2	2.3	26.4	24.4	-2.1

*Note.* \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

In Walnut County (Tables 6.13 and 6.14), gap movements were all insignificant except for the widening gap between ELLs and English speaker in Grade 5. In 2013, 32% and 50% fewer ELLs achieved proficiency compared to English

speaking students in Grades 3 and 5, respectively. Similar to Almond County, the larger achievement gaps were those between ethnic groups; the smallest were those between students with and without disability. In regards to gap sizes, close to one in three fewer Hispanic fifth graders reached proficiency compared to White or Asian students. In Grade 3, the difference between Asian and Hispanic students was 44%. The sizes of the proficiency gap between Hispanic students and their White and Asian peers in Walnut County were relatively similar to those found at Almond County.

Table 6.13

*Grade 3 Reading Achievement Gaps between Groups at Walnut County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>
White–Hispanic	58.7	57.6	-1.3	41.6	39.7	-1.9
Asian–Hispanic	63.8	67.4	3.7	44.3	44.3	0
Economically advantaged–Disadvantaged	49.8	54.7	4.9	35.0	36.7	1.7
English learner–English only	41.1	46.7	5.6	29.1	32.1	3.0
With disability–No disability	29.1	22.1	NA	17.6	23.3	NA
High-school/lower–Associate/higher degree	47.7	49.4	1.7	33.3	33.3	0.1

*Note.* \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Table 6.14

*Grade 5 Reading Achievement Gaps between Groups at Walnut County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>
White–Hispanic	55.0	52.1	-3.0	42.9	34.0	-8.9
Asian–Hispanic	59.9	60.7	0.8	42.9	34.9	-8.0
Economically advantaged–Disadvantaged	46.1	48.4	2.3	35.5	30.7	-4.8
English learner–English only	49.6	66.1	16.5*	40.7	50.0	9.3*
With disability–No disability	48.7	16.3	NA	35.6	12.3	NA
High-school/lower–Associate/higher degree	42.8	42.3	-0.4	32.1	27.8	-4.3

*Note.* \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

**Mathematics.**

The achievement gaps between advantaged and disadvantaged students in mathematics reflect the unrelenting gaps observed in reading. As shown in Tables 6.15–6.16, the gap between ELLs and English only students in Almond County widened statistically by 11% in Grade 3 and 14% in Grade 5. All other movements are statistically insignificant, except the gap between Hispanic and Asian students in Grade 5, which narrowed by 9%. Similar to reading, ethnic group differences continued to be where the largest gaps were found. Between Asian and Black students, the proficiency gaps were 47% and 45% in Grades 3 and 5, respectively. Furthermore, the White–Black proficiency gaps in both grades are 14% and 12% wider than the White–Hispanic gap in Grades 3 and 5, respectively. The smallest gaps remained with students with and without disability. In Grade 5, this gap dropped from 13% in 2008 to 5% in 2013; however, the trend reversed in Grade 3.

Table 6.15

*Grade 3 Mathematics Achievement Gaps between Groups at Almond County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>
White–Black	82.4	98.8	16.4	40.1	43.3	3.3
Asian–Black	109.9	126.6	16.7	46.7	47.4	0.7
Hispanic–Black	16.1	26.2	10.1	9.5	13.9	4.4
White–Hispanic	66.3	72.6	6.3	30.5	29.4	-1.1
Asian–Hispanic	93.8	100.4	6.6	37.2	33.5	-3.7
Economically advantaged–Disadvantaged	55.1	65.8	10.7	24.6	24.9	0.3
English learner–English only	28.1	56.3	28.2*	12.2	23.1	10.9*
With disability–No disability	17.8	24.5	NA	5.7	12.2	NA
High-school/lower–Associate/higher degree	54.7	63.5	8.8	26.1	25.3	-0.7

*Note.* \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Table 6.16

*Grade 5 Mathematics Achievement Gaps between Groups at Almond County*

	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>
White–Black	75.9	94.1	18.2	36.9	38.8	1.9
Asian–Black	111.9	123.1	11.2	47.9	44.7	-3.2
Hispanic–Black	10.8	24.0	13.2	6.2	12.4	6.2
White–Hispanic	65.1	70.1	5.0	30.6	26.4	-4.3
Asian–Hispanic	101.1	99.1	-2.0	41.7	32.2	-9.4*
Economically advantaged–Disadvantaged	52.4	60.4	7.9	24.2	21.3	-2.9
English learner–English only	40.6	72.7	32.2*	18.7	32.3	13.6*
With disability–No disability	26.2	19.5	NA	14.3	5.1	NA
High-school/lower–Associate/higher degree	50.6	58.2	7.6	22.2	19.4	-2.8

*Note.* \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

At Walnut County (Tables 6.17 – 6.18), there was absolutely no statistically significant change in achievement gap across subgroups and grades whether measured by scale scores or proficiency standards. The gaps in proficiency rates however, had narrowed in absolute terms albeit in single digits. The proficiency gaps between disability–no disability students dropped by 10% and 24%, respectively in Grades 3 and 5 between 2008 and 2013. Similar to Almond County, the gaps between disability and no disability subgroups in both grades are also the smallest amongst other subgroup gaps. The gaps sizes, such as those between ethnic groups, and economically disadvantaged–advantaged gaps are relatively similar in both counties. As a matter of fact, the gap sizes at Walnut between subgroups have widened between Grades 3 and 5, whereas at Almond, they have narrowed. The most alarming being the proficiency gap between ELL and English only students. While they had not grown significantly between 2008 and 2013, they nearly doubled between Grades 3 and 5. This is true in both 2008 and 2013.

Table 6.17

*Grade 3 Mathematics Achievement Gaps between Groups at Walnut County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>
White–Hispanic	72.7	72.9	0.2	33.0	27.4	-5.6
Asian–Hispanic	104.8	110.2	5.4	40.9	34.0	-6.8
Economically advantaged–Disadvantaged	66.3	72.5	6.2	27.7	24.3	-3.4
English learner–English only	37.4	46.2	8.8	16.0	16.9	0.9
With disability–No disability	52.2	37.2	NA	24.0	14.4	NA
High-school/lower–Associate/higher degree	62.5	62.1	-0.4	27.0	21.7	-5.2

*Note.* \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

Table 6.18

*Grade 5 Mathematics Achievement Gaps between Groups at Walnut County*

<i>Student subgroups</i>	<i>Mean (scale score)</i>			<i>Proficiency and above (%)</i>		
	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>	<i>2008 gap</i>	<i>2013 gap</i>	<i>Diff.</i>
White–Hispanic	81.0	80.0	-1.1	38.1	30.5	-7.6
Asian–Hispanic	113.2	121.3	8.1	45.7	38.8	-6.9
Economically advantaged–Disadvantaged	71.8	77.8	6.0	31.8	26.0	-5.8
English learner–English only	58.4	77.7	19.2	27.5	32.3	4.9
With disability–No disability	76.6	30.0	NA	34.5	10.1	NA
High-school/lower–Associate/higher degree	64.0	65.4	1.4	29.9	24.4	-5.5

*Note.* \* $p < 0.05$  indicates that the change between the two calendar years is statistically significant.

In summary, most achievement gaps between advantaged and disadvantaged students did not budge between 2008 and 2013, whether the comparisons were between ethnic groups or other student backgrounds. The gap between English learners and English only students has increased significantly across counties, grade levels and domains. The exceptions are the gaps in mathematics at Walnut County where they remained unchanged. The proficiency difference between ELLs and English speakers varied, being up to 50% in reading in Grade 5. Other disadvantaged groups, such as Hispanic and Black students, trailed their advantaged Asian and White peers by about 40% on average. Among all gaps, the



gap between students with and without disability is the smallest across counties, domains and grades.

### **Inclusion in Assessment**

Inclusion of all students in external assessment does not appear to be an issue for either county. Both counties assessed an average of 99% of students eligible for the general and alternative external assessments between 2008 and 2013, with the majority of schools demonstrating 100% inclusion. While it was not possible to analyse inclusion by subgroup due to the lack of subgroup enrolment data, despite the availability of subgroup tested data, the total inclusion average of 99% per school suggests that few students were left out of the assessments.

However, it was also not possible to validate this total figure, because a significant part of the alternative assessment enrolment and results were not reported in the dataset, given the ten-student minimal reporting threshold discussed previously.

Nonetheless, the high overall inclusion rate appears to have confirmed the prediction made by Hanushek and Raymond (2005) that, “many of the adverse effects that involve “gaming” the system come from short run incentives that are unlikely to be present over time” (p. 33). Additionally, the mandate to exempt no more than 2% per subgroup (O'Day et al., 2009) as part of the criteria to meet the annual API could also have attributed to the near-total inclusion status.

### **Discussion**

The quantitative analyses of Almond and Walnut Counties were conducted using CST data and *t*-tests. Results from these counties tracked what has been found

in studies based on NAEP data. No empirical evidence based on CST data has been identified in the literature. On average, some progress has been made in both domains and grade levels, but reading progress trailed mathematics and Grade 3 made smaller gains than Grade 5. While not all disadvantaged subgroups achieved better outcomes in 2013, Hispanic students and students from economically disadvantaged background did show greater growth than their advantaged peers. Nonetheless, it is important not to forget that, fewer than 40% of any disadvantaged subgroup achieved reading proficiency in Grade 3 and fewer than 60% in Grade 5 as of 2013. Most discouragingly, fewer than 15% and 25% of ELLs reached proficiency status as of 2013 at Almond and Walnut, respectively, in each grade. The results were slightly better in mathematics. At the current proficiency level, the achievement of disadvantaged students will need to continue to grow at significant rates before a majority reaches proficiency. Nonetheless, it is encouraging that an average of 11–14% more students reached reading proficiency between Grades 3 and 5 in both counties. The largest between-grade increases were all evidenced amongst disadvantaged student groups. Does this suggest that once disadvantaged students build a strong reading foundation, they can achieve progress at a faster rate? Perhaps evidence of growth contributed positively to students' efficacy beliefs, in turn affecting their personal efforts, a cognitive process that has been associated with successful school performance (Usher & Pajares, 2008).

Notwithstanding some encouraging growth demonstrated by a few disadvantaged groups, progress over time has not translated to the closing or narrowing of achievement gaps between advantaged and disadvantaged students. Although most achievement gaps were trending down between 2008 and 2013 in

absolute terms, the narrowing has not been large enough to be statistically significant. On the positive side, these gaps remained because both advantaged and disadvantaged subgroups made progress under the accountability policy. To that end, disadvantaged students' progress has not come at the expense of advantaged students. Another encouraging sign is the gap between students with and without disability, which was the smallest of all advantaged/disadvantaged gaps. Albeit only a very small sample in the analysis compared to all other subgroups, students with disability have made progress. Perhaps inclusive education and higher expectations under the accountability environment have enabled students with disability to demonstrate that disability and performance are not necessarily negatively correlated.

The most troubling of all trends is the outcomes of ELLs, who made no progress and continued to be the lowest performers in both domains. Compounding this already-challenging trend is the further drop in Grade 5 mathematics proficiency gap compared to the gap in Grade 3. While finding the root cause is outside the scope of this research, could this trend suggest that the challenge ELLs face in reading is affecting their ability to keep up with their mathematics knowledge development? Is this drop a true indication of ELLs' mathematics proficiency or is it confounded by other challenges? Regardless of the measure's accuracy, this lack of progress signals trouble for California as it educates the largest ELL population in the US. While it is encouraging that the California State Assembly had passed a bill in 2011 mandating revision of the English Language Development Standards aimed at supporting ELLs (California Department of Education, 2013a), the revision process

has been slow and lengthy. At the time of this writing, various pieces of the program were still being developed and finalised.

Lastly, results from these two counties also confirmed that student outcomes are highly correlated with socioeconomic factors and that socioeconomic advantage is associated with parental education background, as has been found in past research (Marks, Cresswell, & Ainley, 2006; OECD, 2012a). As seen in Figures 6.1 to 6.4, students from economically advantaged and better-educated family backgrounds achieved the highest outcomes. On the low end of the scale, outcomes of students from economically disadvantaged and less-educated families stacked closely together.

## **Summary**

The quantitative findings using data from two California counties produced positive and negative outcomes of test-based accountability on student achievement. On the negative end, Figures 6.1 to 6.4 illustrate the magnitude of how test-based accountability has not supported the needs of ELLs in these counties. However, ELLs' results were the exception and not the norm, as multiple subgroups achieved statistically significant scale score growth, albeit at different rates. For the most part, those groups that benefited most under NCLB were the disadvantaged students, as other research has also observed. For example, Hispanic students and students with disability experienced the steepest growth. However, because the majority of the subgroups also made statistically significant progress, the achievement gaps have not budged in these two counties. Given the size of some of these gaps, particularly in reading, where Black and Hispanic students trailed their

cohorts by more than 30%, closing the achievement gap has proven to be an elusive goal, as some critics have claimed. The evidence of overall achievement progress, however, suggests that the US federal government is heading in the right direction by moving the accountability policy towards measuring growth as opposed to universal proficiency in its latest reiteration of the Education Act. This may also reduce unintended consequences stemming from the pressure to reach 100% proficiency under a tight deadline, leaving educators with more flexibility to build on the growth achieved thus far.

## Chapter 7 Quantitative Findings: Academic Progress in Australia

This chapter continues with the quantitative findings for the Australian sample using NAPLAN reading and numeracy data for students in Years 3 and 5. The presentation follows similar format to that of Chapter 6. It begins with descriptive data followed by an explanation of the proxy advantage and disadvantage school category development. To enable statistical comparisons between advantaged and disadvantaged groups, proxy category development was an additional step to those taken in the CST data analysis, due to insufficient data disaggregation in the NAPLAN data structure. The NAPLAN data structure also necessitated the use of different statistical techniques from those used for the CST data analysis; these will be described further in the *Statistical analysis* section within this chapter. Finally, the statistical results are presented and discussed.

### Data and Statistical Procedures

The ex post facto quantitative design examined the effects of school and student characteristics on NAPLAN performance and test inclusion. The research project neither introduced nor manipulated any variables in relation to the NAPLAN data. Instead, it analysed school-level variables captured at the time of official NAPLAN testing as part of the background statistics. As such, these factors were already present prior to the study, and the analysis aimed to explore their effects on the NAPLAN results. School background factors include variables such as: socioeconomic status (ICSEA score), school sector, school type, jurisdiction, attendance rate and enrolment. Student-level factors cover language spoken and

Indigenous identification. Beyond each factor's contribution to student outcomes, the analysis aimed to evaluate how disadvantaged schools fared against their advantaged peers. The determination of 'advantaged' and 'disadvantaged' was the first step of the analysis which used multiple regression. Analysis of variance (ANOVA) followed to explore NAPLAN growth over time and regression models were built to evaluate achievement gap movements.

**Data summary.**

ACARA provides cross-sectional measures of NAPLAN results aggregated by grade and school from 2008 to 2013. As in the CST data, the outcome measures are scale score and the percentage of students reaching or achieving beyond the minimum proficiency standards. Proficiency measures are represented by two new variables created in this study: 'Year 3 proficiency' and 'Year 5 proficiency', which combine results in different proficient bands into one variable, as discussed in Chapter 5. The dataset includes a large number of background variables. Of interest to the present study were variables that indicate Indigenous status, disability status, English language ability, socioeconomic status, student population, attendance trend, and school sector and type. These factors have been identified in the literature as strong predictors of outcomes. In addition to academic results, the analysis also evaluated NAPLAN absentee and withdrawal rates to gauge exclusion trends in the national assessment.

Altogether, the Years 3 and 5 dataset contained 391,684 observations over the six-year period from 2008 to 2013. The average ICSEA score across all schools was 1007 and average attendance rate was 93%. The Indigenous student population was 9%, and 22% students were of a non-English speaking background. Among the

schools, 63% were identified as government, 20% Catholic and 12% independent schools. In regards to school type, over 80% were primary schools, 17% combined schools (K–12), and only 0.1% special schools for students with disability.

### **Statistical analysis.**

#### ***Proxy categories for comparative analysis.***

As will be recalled in Chapter 5 *Defining advantaged and disadvantaged students*, proxy categories were created to enable a comparison of advantaged and disadvantaged schools. The following disadvantaged and advantaged categories (Table 7.1) were defined, and they will form the basis for the comparative analysis in this chapter.

Table 7.1

*Disadvantaged and Advantaged Proxies for NAPLAN Variables*

<i>NAPLAN variables</i>	<i>Disadvantaged proxy</i>	<i>Advantaged proxy</i>
Jurisdiction	The Northern Territory	Victoria
School sector	Government schools	Independent schools
School type	Special schools	Primary schools
ICSEA score	Bottom quartile	Top quartile
LBOTE student	Bottom quartile	Top quartile
Attendance rate	Bottom quartile	Top quartile

#### ***NAPLAN progress over time.***

The present study conducted analyses of variance (ANOVA) to evaluate progress over time on scale score, proficiency achievement, and test inclusion status, and multiple regressions to evaluate trends on achievement gaps. Both approaches have been used in prior research evaluating growth in large-scale external test data (Abedi et al., 2001; Clotfelter et al., 2006; Fryer & Levitt, 2004). A series of two-way analyses of variance (ANOVAs) were employed to evaluate changes between 2008 and 2013 on reading, numeracy and assessment exclusion. ANOVA is deemed most



appropriate to test group differences when “the independent variable is defined as having two or more categories and the dependent is quantitative” (Mertler & Vannatta, 2002, p. 15). Variables representing school and student characteristics all have multiple subcategories; some are predefined, such as school sector which has ‘independent’, ‘Catholic’, and ‘government’ as subcategories; others are a product of proxy category creation which generated four quartiles.

Separate ANOVAs were performed for each year level (3 and 5), school characteristic (jurisdiction, ICSEA score, attendance, LBOTE concentration, school sector, and school type), and achievement outcomes (reading and numeracy) and rate of exclusion. For example, a two-way ANOVA was conducted with time (2008 and 2013) and jurisdiction (eight jurisdictions) as the independent variables and Year 3 reading scores as the dependent variable. A total of 24 two-way ANOVAs were performed (12 for each year level of which six were for reading and six for numeracy). The process was repeated for exclusion rate. To address the research question of change over time amongst the subgroups within each school background characteristic and to minimize type I error with multiple comparisons, post hoc tests were conducted on significant interactions between time and a characteristic variable using the Bonferroni correction. Given its conservative approach, the Bonferroni correction is an appropriate test to employ for the multiple post hoc comparisons in this series of analyses (Field, 2009). For the post hoc tests, the dependent variable was compared between two calendar years for each category within a school attribute. For example, following a significant interaction between jurisdiction and time on Year 3 reading scores, eight post hoc tests were performed to compare the Year 3 reading scores between the two time points for each

jurisdiction category. To minimise type I error, the study did not follow up with pairwise comparisons between subcategories following a significant interaction as that is also not the focus of the study.

***Trends on achievement gap.***

To measure changes in achievement gap over the six-year period from 2008 to 2013 between the advantaged and disadvantaged school groups, regression models using ordinary least squares were built. The modelling in the present study draws heavily from studies that have examined achievement gaps using large-scale test data (Clotfelter et al., 2006; Fryer & Levitt, 2004; Schnepf, 2007). Since group differences are generally mediated by a host of background factors that are systematically associated with achievement outcomes as reported in Chapter 5 (Table 5.2), regressions offer a better chance of isolating achievement effects that are more directly related to the factor being measured by isolating other factors (Clotfelter et al., 2006). In the current study, the models evaluated the linear time trend of the disadvantaged group relative to the advantaged group; for example, it compared the outcomes of schools in the top ICSEA quartile against the bottom ICSEA quartile, or government schools against independent schools while controlling for other factors. The estimated positive or negative coefficients for the pair of interacting terms provide an overview of the trends in achievement gap. A positive coefficient indicates that the disadvantaged group has improved relative to the advantaged group of schools, hence narrowing the achievement gap between the two. A negative coefficient suggests a widening of achievement gap. Tables 7.5 to 7.6 summarise the trends in NAPLAN scores, proficiency rate, and exclusion rate. An

example of the model specifications that predict NAPLAN trends by looking at the ‘attendance’ factor is as follows:

$$NAPYR3_i = \beta_0 + \beta_1 trend_t + \beta_2 Trend_t\_AttendQ4 + \beta_3 AttendQ4_i + \beta_3 AttendQ3_i + \beta_3 AttendQ2_i + \beta_3 Z_{i,t} + \varphi_t + \mu_i + \varepsilon_{i,t}$$

Where

$$NAPYR3_i =$$

*the measure of Year 3’s average achievement outcomes at year i*

$\beta_0, \beta_1, \beta_2, \beta_3 =$  *the estimators to minimise the sum of squared errors*

$Z_{i,t} = (ICSEA\_RD_i, Attend_i, LBOTE_i, Govt_i, Catholic_i, Combined_i, Special_i)$

$\varphi_t =$  *Calendar Year*

$\mu_i =$  *Jurisdiction*

$\varepsilon_{i,t} =$  *Error term*

As the study design is ex post facto, it is important to note that the analysis in this research could not capture or explain the possible effects of non-school characteristics on achievement. For example, grants from the federal partnership programs, which existed in some disadvantaged schools, additional support and funding from either the state or federal government, or new programs that a specific category of schools might have implemented. Therefore, findings must be read with caution.

## Reading and Numeracy Achievement Trends

The outcome of the analyses using school average NAPLAN data shows that disadvantaged schools performed below the national average in mean scale scores and in percentage of students reaching minimal national standards across domains and year levels (Figures 7.1–7.2). While not on the chart, all advantaged subgroups’ achievements exceeded the national average. The exception was the LBOTE

penetration at schools. Having more LBOTE students did not impact NAPLAN outcomes. Schools in the Northern Territory and special schools stand out as two school groups that consistently performed far below the national average. As presented in Table 7.4, there is a consistent pattern of progress in reading across jurisdictions and different school characteristics in both year levels. Much of the gains are significant at the  $p < .001$  level. In contrast, the same table shows that in Year 3, schools experienced a general decline in numeracy and the disadvantaged school groups lodged the sharpest decreases. In Year 5, progress was modest and varied among subgroups and jurisdictions, and where there were gains, the size of absolute score gains lagged reading.

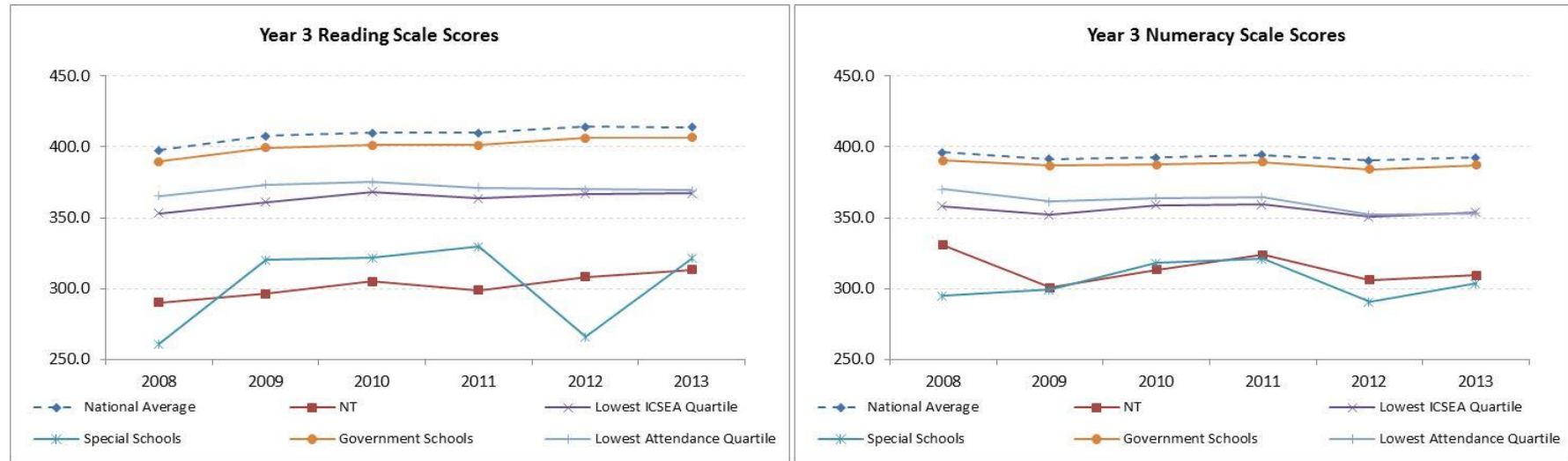


Figure 7.1. Year 3 Reading and Numeracy Average Score amongst School Groups

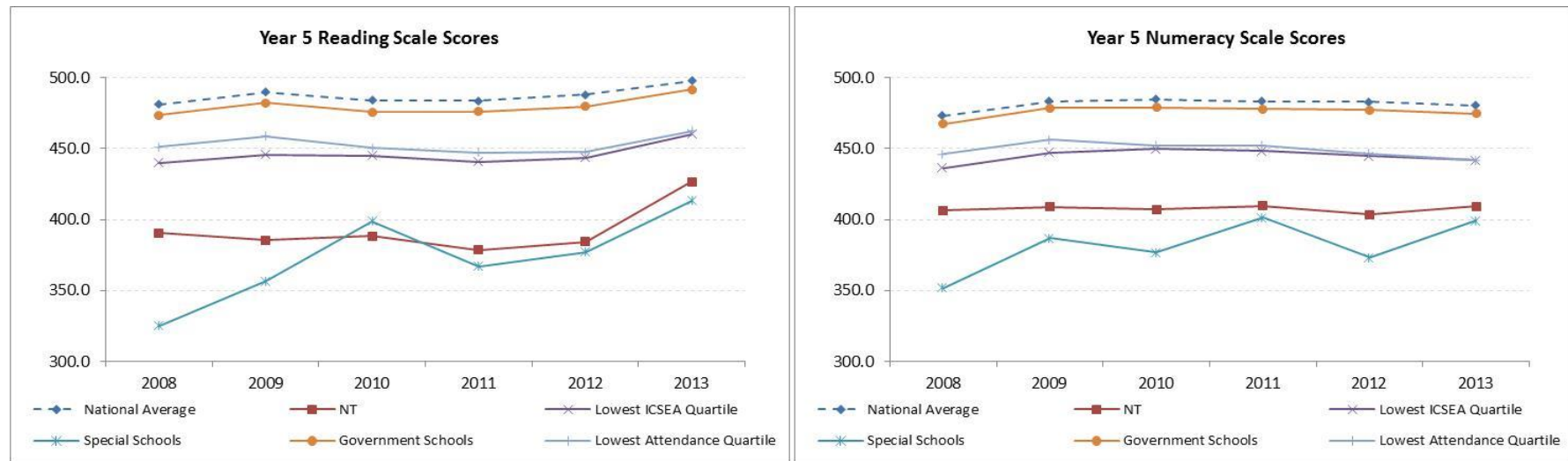


Figure 7.2. Year 5 Reading and Numeracy Average Score amongst School Groups

**Subgroup differences in reading and numeracy.**

As shown in Table 7.2, schools that serve only special needs students, have a low socioeconomic score and are government schools consistently scored at the bottom of the scale regardless of subject domain or year level. The reverse was true for independent schools and schools with the highest ICSEA score. Admittedly, not all government schools are disadvantaged schools, however, the results show that, as a group, independent schools consistently outperformed Catholic schools, who in turn out-performed government schools. This finding is consistent with prior studies (Marks et al., 2001; Miller & Voon, 2012). Furthermore, student attendance, another proxy for equity, had a large impact on NAPLAN performance. Similar to ICSEA background, the higher a school's overall attendance rate, the higher the NAPLAN score, with the opposite also true. Across jurisdictions, NT, followed by QLD and WA, had the lowest means in scale score across domains and years. In Year 3, NT trailed the national average by as much as 107 points. Similarly, special schools differed from the national average by over 100 points across year levels and domains.

The size of a school's LBOTE student population did not affect NAPLAN outcomes. In fact, Figures 7.3 and 7.4 demonstrate the opposite effect. Schools enrolling the smallest percentage of students whose language background is not English consistently earned the lowest average scores; those enrolling the second largest percentage of LBOTE students (third highest quartile) earned the highest scores. This aberration corroborates other evaluations (Creagh, 2014; Lingard et al., 2012).

Table 7.2

*Average NAPLAN Scale Score Means and Standard Deviations (2008–2013)*

<i>School subgroups</i>	<i>Year 3</i>		<i>Year 5</i>	
	<i>Reading</i> 408.8 (47.8)	<i>Numeracy</i> 392.9 (41.2)	<i>Reading</i> 481.3 (43.0)	<i>Numeracy</i> 487.3 (39.7)
<i>Jurisdiction</i>				
NT	302.2 (101.1)	313.8 (72.6)	392.0 (99.6)	407.5 (67.2)
QLD	389.2 (41.9)	374.9 (36.3)	471.8 (37.8)	467.3 (34.3)
WA	394.6 (49.6)	380.5 (41.9)	476.7 (46.2)	470.4 (41.5)
SA	400.3 (41.2)	379.0 (36.0)	478.5 (37.4)	466.5 (34.0)
TAS	407.3 (41.1)	390.8 (35.8)	483.4 (37.9)	472.2 (33.5)
NSW	417.2 (42.0)	401.1 (37.3)	495.3 (38.3)	490.9 (38.4)
VIC	427.6 (36.9)	410.4 (33.3)	501.8 (32.0)	494.4 (31.9)
ACT	434.3 (34.2)	409.7 (28.3)	510.0 (30.9)	493.9 (28.8)
<i>School type</i>				
Special	305.1 (64.2)	306.0 (42.3)	375.2 (54.0)	382.7 (52.6)
Combined	405.6 (68.5)	390.6 (55.7)	484.6 (62.7)	480.5 (52.6)
Primary	409.6 (42.2)	393.5 (37.4)	488.0 (37.4)	481.6 (36.2)
<i>School sector</i>				
Government	400.6 (48.5)	387.4 (42.1)	479.7 (44.1)	475.7 (40.2)
Catholic	422.1 (36.8)	399.0 (33.0)	499.3 (32.4)	487.2 (30.4)
Independent	434.7 (44.9)	414.6 (39.0)	511.2 (32.7)	503.3 (37.8)
<i>ICSEA quartile</i>				
Bottom	363.4 (49.6)	355.5 (41.3)	445.9 (46.7)	444.8 (38.3)
Q2	399.9 (30.9)	385.9 (28.8)	479.9 (26.7)	473.9 (26.2)
Q3	418.9 (29.0)	400.9 (27.8)	496.4 (25.0)	487.9 (25.3)
Top	450.7 (29.9)	427.2 (28.2)	524.3 (26.5)	515.9 (39.5)
<i>LBOTE quartile<sup>1</sup></i>				
Bottom	404.0 (37.9)	384.9 (32.8)	483.6 (32.8)	475.7 (29.5)
Q2	412.6 (40.7)	392.3 (34.8)	491.1 (36.1)	482.2 (32.3)
Q3	422.8 (45.0)	400.7 (39.1)	499.0 (33.4)	490.7 (37.9)
Top	408.8 (60.1)	392.3 (51.3)	483.7 (56.3)	485.9 (51.6)
<i>Attendance rate</i>				
Bottom	370.8 (61.9)	361.1 (50.7)	452.8 (57.9)	449.5 (46.5)
Q2	402.4 (37.9)	387.2 (29.4)	481.9 (32.8)	475.0 (30.6)
Q3	415.2 (35.6)	398.1 (31.9)	493.7 (31.0)	485.9 (28.7)
Top	433.0 (37.5)	413.4 (33.5)	508.9 (32.9)	503.0 (34.4)

*Note.* Numbers without brackets represent scale score. Numbers in brackets are the standard deviations.

<sup>1</sup>Data was unavailable for 2008. The numbers in this category represent the average for 2009–2013.

Table 7.2 also records the standard deviation for each subgroup. It measures the variability in NAPLAN scale scores of the specified group. A larger standard deviation indicates a wider spread of scores. On the whole, in Years 3 and 5, there was a wide dispersion of scores among the subgroups within each background characteristic across domains. The dispersions amongst disadvantaged schools were wider compared to advantaged schools. Among eight jurisdictions, NT's dispersions across year levels and domains were three times larger than the two best-performing jurisdictions, VIC and ACT. NT's dispersions were also the largest among all subgroups regardless of background. Western Australia also varied widely from the rest of the jurisdictions. In comparison, ACT and the top two quartiles of SES subgroups had some of the smallest dispersions across year levels and domains.

In addition to the scale scores, NAPLAN also measures the percentage of students meeting grade-level national minimum standards. According to ACARA (2014) students not meeting the national minimum standard will have difficulty progressing at school. Table 7.3 presents a summary of how school groups of various background characteristics met standards over the six-year period. Across Australia, 90–93% of the schools met reading and numeracy minimum standards across both years between 2008 and 2013. The exceptions were schools in the NT jurisdiction, special schools, the lowest ICSEA quartile and lowest attendance quartile. While schools with the lowest ICSEA score and attendance rate trailed other subgroups, their results were not as discouraging as the results of NT school and special schools.



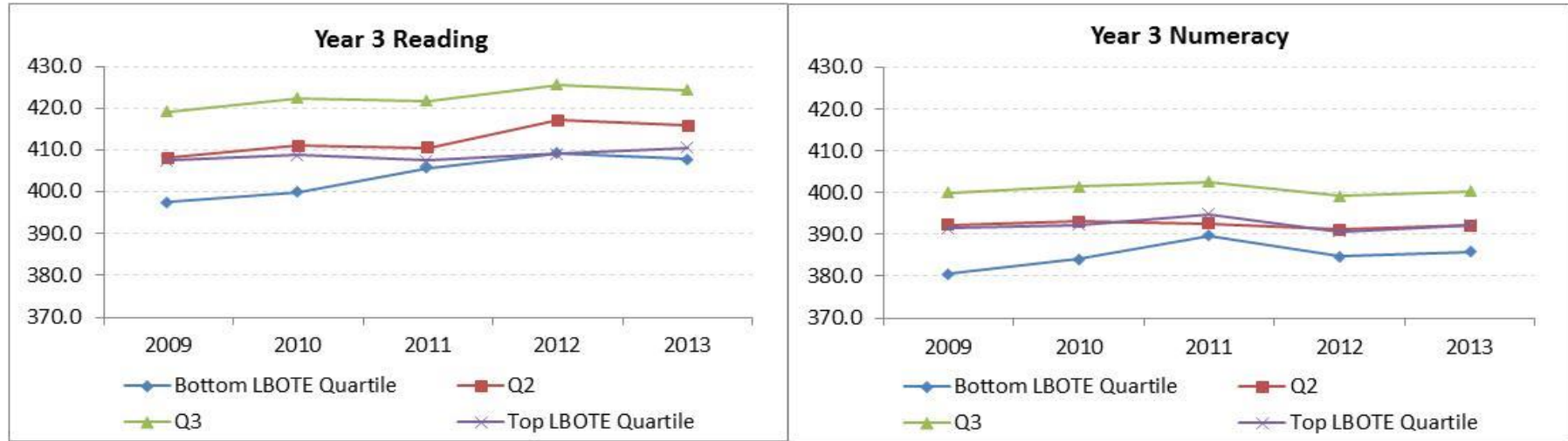


Figure 7.3. Year 3 NAPLAN Mean Score by LBOTE Quartile

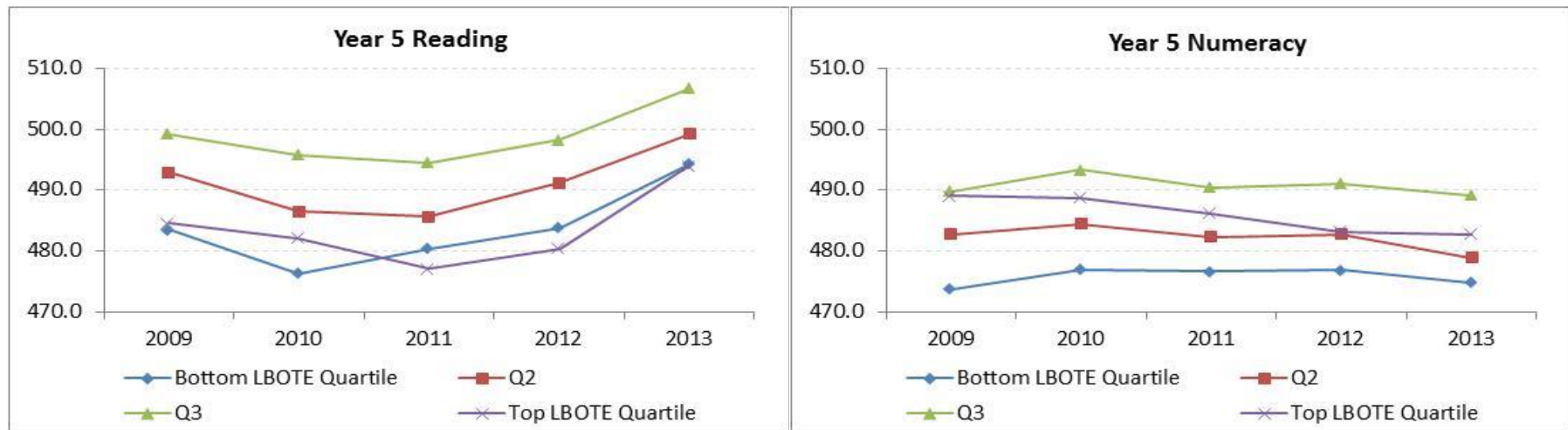


Figure 7.4. Year 5 NAPLAN Mean Score by LBOTE Quartile

Table 7.3

*Average NAPLAN Proficiency Rates and Standard Deviations (2008–2013)*

<i>School subgroups</i>	<i>Year 3</i>		<i>Year 5</i>	
	<i>Reading</i> 92.5 (11.1)	<i>Numeracy</i> 93.2 (10.9)	<i>Reading</i> 89.6 (14.0)	<i>Numeracy</i> 92.0 (12.1)
<i>Jurisdiction</i>				
NT	63.2 (32.4)	66.1 (32.5)	55.3 (36.8)	61.5 (34.6)
QLD	90.6 (10.5)	91.4 (11.0)	86.3 (14.4)	90.2 (12.2)
WA	90.0 (13.1)	91.6 (12.8)	86.7 (17.2)	89.9 (14.9)
SA	91.5 (11.2)	91.4 (11.7)	88.3 (14.2)	90.1 (12.6)
TAS	92.0 (9.0)	93.2 (8.5)	88.4 (12.0)	91.1 (9.6)
NSW	94.2 (8.2)	94.6 (8.1)	91.6 (10.9)	93.6 (9.2)
VIC	95.1 (6.8)	95.6 (6.4)	93.4 (8.1)	94.2 (6.8)
ACT	94.4 (7.2)	95.1 (6.6)	93.0 (8.3)	94.3 (7.7)
<i>School type</i>				
Special	56.1 (23.0)	54.5 (24.6)	34.2 (31.1)	41.3 (32.3)
Combined	90.2 (17.7)	91.0 (17.6)	86.9 (21.8)	89.8 (19.5)
Primary	93.0 (8.9)	93.6 (8.8)	90.2 (11.5)	92.6 (9.6)
<i>School sector</i>				
Government	91.0 (11.9)	91.9 (11.8)	87.5 (15.1)	90.5 (13.0)
Catholic	95.6 (7.6)	95.7 (7.9)	93.8 (9.7)	95.1 (8.8)
Independent	95.7 (8.9)	96.4 (8.4)	94.5 (10.3)	95.9 (8.9)
<i>ICSEA quartile</i>				
Bottom	84.5 (16.4)	85.4 (16.7)	77.9 (20.6)	82.3 (18.4)
Q2	92.8 (7.9)	93.6 (7.6)	89.9 (9.4)	92.7 (7.7)
Q3	95.2 (6.4)	95.8 (5.9)	93.4 (7.5)	95.2 (6.1)
Top	97.2 (4.9)	97.7 (4.2)	96.5 (5.2)	97.4 (4.5)
<i>LBOTE quartile<sup>1</sup></i>				
Bottom	93.1 (8.2)	93.5 (8.4)	90.2 (10.4)	92.9 (8.9)
Q2	93.8 (8.1)	94.2 (8.4)	91.3 (10.8)	93.6 (8.8)
Q3	94.2 (8.2)	94.6 (8.3)	92.1 (10.5)	93.8 (9.1)
Top	91.1 (15.1)	90.6 (15.6)	87.5 (19.1)	89.2 (17.6)
<i>Attendance rate</i>				
Bottom	84.4 (18.4)	85.3 (18.6)	78.3 (23.1)	82.4 (20.6)
Q2	91.3 (8.9)	92.2 (8.9)	87.6 (11.3)	90.9 (9.3)
Q3	93.8 (7.5)	94.4 (7.2)	91.4 (9.2)	93.7 (7.7)
Top	96.0 (6.4)	96.6 (5.9)	94.6 (7.5)	96.0 (6.2)

*Note.* Numbers without brackets represent proficiency rate. Numbers in brackets are the standard deviations.

<sup>1</sup>Data was unavailable for 2008. The numbers in this category represent the average for 2009–2013

In the NT jurisdiction, no more than three in five students reached proficiency level in either reading or numeracy. Among special schools, just over 50% achieved minimum standard in Year 3 and only 34% and 41% met proficiency standards in Year 5 reading and numeracy respectively. The dispersions are wider under this measure than scale score and the dispersion difference between ACT and NT is about five-fold. In contrast, nearly 100% of the schools in the highest ICSEA quartile and schools with the highest attendance rate reached proficiency standards.

In Year 3, results in reading and numeracy are nearly identical across every subgroup. In Year 5, schools were marginally more successful at reaching minimum national standards in reading than in numeracy. As with NAPLAN scale scores, disadvantaged schools had the lowest percentage of students meeting minimum proficiency standards. Disadvantaged schools' proficiency rates also varied more dramatically than advantaged schools as indicated by the larger standard deviations. One notable difference from the NAPLAN scores was observed in the LBOTE characteristics. Unlike NAPLAN scores, where schools with the highest percentage of LBOTE students did not always show the worst results, they did persistently have the lowest percentage of students not meeting minimum standards and wider dispersions when compared against schools in the other quartiles.

#### **Progress on NAPLAN.**

The ANOVA results show that the interactions between jurisdiction and time were significant across year levels and domains: Year 3 reading,  $F(7, 13046) = 26.2$ , Year 3 numeracy,  $F(7, 13045) = 38.6$ , Year 5 reading,  $F(7, 13096) = 24.0$ , and Year 5 numeracy,  $F(7, 13103) = 21.3$  all at  $p = .000$ . These statistics suggest that the change between 2008 and 2013 was different depending on jurisdiction and on year. Other

background variables that had significant interactions included attendance: Year 3 reading,  $F(7, 13024) = 10.0$ , Year 3 numeracy,  $F(3, 13022) = 24.9$ , Year 5 numeracy,  $F(3, 13074) = 16.0$  all at  $p = .000$ . Interactions between school type and time were also significant in Year 3 reading  $F(3, 13049) = 4.3$  and numeracy  $F(3, 13100) = 9.3$ ,  $p = .000$ . Similar results were found under ICSEA reading score  $F(3, 13076) = 16.5$ ,  $p = .000$  and school sector reading score  $F(2, 13106) = 8.4$ ,  $p = .000$  in Year 5.

### ***Reading.***

Between 2008 and 2013, schools in Years 3 and 5 made a statistically significant gain of 16 points on average in reading. This growth has also been relatively steady as demonstrated in Figures 7.1–7.2. However, the absolute difference in scale scores between the two calendar years varied substantially amongst subgroups as recorded in Table 7.4. Nonetheless, most of the changes were statistically significant except for the LBOTE subgroups. The two lowest-performing groups, NT schools and special schools achieved some of the larger and statistically significant growths. Special schools gained 60 points in Year 3 and 88 points in Year 5, almost a six-fold difference in Year 5 compared to primary schools' 16-point increase. Their gains were four to five times the national average. However, in comparison to most groups, their growth has been inconsistent from year to year as noted in Figures 7.1–7.2. Amongst jurisdictions, NT schools gained an average of 23 points in Year 3. Only QLD did better with a gain of 36 point and ACT was on par with a 24-point increase. In Year 5, NT had the largest increase of 36 points followed by QLD at 30 points. Among ICSEA quartiles, lower SES schools in Year 5 grew by 21 points compared to the 14-point gain made by their advantaged peers. Differences

among other background characteristics were less pronounced. While absolute score increase measured by attendance rate was not high across all four quartiles, schools with the highest attendance rate increased their scores by more than 2.5 times than schools at the bottom quartile in both years.

### ***Numeracy.***

The growth patterns in numeracy are mixed and less impressive in terms of absolute growth. In fact, Year 3's scores declined nationally by four points while Year 5 increased modestly by seven points. Both changes are statistically significant. In Year 3, across all background characteristics and jurisdictions, only schools in Queensland and WA made statistically significant gains. Not only did Queensland buck the downward trend, it grew significantly by 16 points. While special schools did not make significant gains, the scores of this group of schools did not decline, as did the majority of the school groups analysed. Among the declining scores, NT had the biggest drop of 21 points followed by Tasmania at 13 points. While growth was relatively modest in Year 5, most school groups recorded statistically significant increases. Queensland schools and special schools nationally made the largest statistically significant gains of 21 and 48 point, respectively, while most other groups achieved only single-digit growth.

Table 7.4

*Difference in NAPLAN Means (2013 less 2008)*

School subgroups	Year 3		Year 5	
	Reading	Numeracy	Reading	Numeracy
National	↑ (16.3) <sup>***</sup>	↓ (-3.8) <sup>***</sup>	↑ (16.3) <sup>***</sup>	↑ (7.1) <sup>***</sup>
Jurisdiction	↑	↑	↑	↑
	ACT (23.6) <sup>***</sup>	ACT (2.4)	ACT (16.7) <sup>**</sup>	ACT (12.5) <sup>*</sup>
	NSW (9.7) <sup>***</sup>	QLD (15.8) <sup>***</sup>	NSW (9.8) <sup>***</sup>	NSW (2.4) <sup>*</sup>
	NT (23.1) <sup>***</sup>	WA (3.7) <sup>*</sup>	NT (35.9) <sup>***</sup>	NT (2.7)
	QLD (35.7) <sup>***</sup>	↓	QLD (29.8) <sup>***</sup>	QLD (21.1) <sup>***</sup>
	SA (7.1) <sup>*</sup>	NSW (-8.8) <sup>***</sup>	SA (14.1) <sup>***</sup>	SA (5.2) <sup>*</sup>
	TAS (12.3) <sup>**</sup>	NT (-21.4) <sup>***</sup>	TAS (18.6) <sup>**</sup>	TAS (4.3)
	VIC (13.0) <sup>***</sup>	SA (-10.8) <sup>***</sup>	VIC (12.5) <sup>***</sup>	VIC (0.9)
	WA (20.0) <sup>***</sup>	TAS (-12.9) <sup>***</sup>	WA (21.3) <sup>***</sup>	WA (15.2) <sup>***</sup>
		VIC (-10.4) <sup>***</sup>		
ICSEA score	↑	↓	↑	↑
	Bottom 25 <sup>th</sup> % (14.2) <sup>***</sup>	Bottom 25 <sup>th</sup> % (-4.0) <sup>***</sup>	Bottom 25 <sup>th</sup> % (20.5) <sup>***</sup>	Bottom 25 <sup>th</sup> % (5.3) <sup>***</sup>
	25 <sup>th</sup> % (15.2) <sup>***</sup>	25 <sup>th</sup> % (-3.8) <sup>***</sup>	25 <sup>th</sup> % (15.2) <sup>***</sup>	25 <sup>th</sup> % (6.1) <sup>***</sup>
	50 <sup>th</sup> % (13.4) <sup>***</sup>	50 <sup>th</sup> % (-7.3) <sup>***</sup>	50 <sup>th</sup> % (12.3) <sup>***</sup>	50 <sup>th</sup> % (4.3) <sup>***</sup>
	75 <sup>th</sup> % (17.7) <sup>***</sup>	75 <sup>th</sup> % (-4.4) <sup>***</sup>	75 <sup>th</sup> % (14.4) <sup>***</sup>	75 <sup>th</sup> % (8.5) <sup>***</sup>
Attendance	↑	↓	↑	↓
	Bottom 25 <sup>th</sup> % (5.5) <sup>**</sup>	Bottom 25 <sup>th</sup> % (-9.0) <sup>***</sup>	Bottom 25 <sup>th</sup> % (1.6) <sup>***</sup>	Bottom 25 <sup>th</sup> % (3.6) <sup>**</sup>
	25 <sup>th</sup> % (10.7) <sup>***</sup>	25 <sup>th</sup> % (-3.5) <sup>***</sup>	25 <sup>th</sup> % (6.0) <sup>***</sup>	↑ 25 <sup>th</sup> % (8.5) <sup>*</sup>
	50 <sup>th</sup> % (11.6) <sup>***</sup>	50 <sup>th</sup> % (-2.6) <sup>***</sup>	50 <sup>th</sup> % (6.4) <sup>***</sup>	50 <sup>th</sup> % (7.9) <sup>***</sup>
	75 <sup>th</sup> % (14.2) <sup>***</sup>	75 <sup>th</sup> % (-0.7)	75 <sup>th</sup> % (8.6) <sup>***</sup>	75 <sup>th</sup> % (10.1) <sup>***</sup>
LBOTE concentration <sup>1</sup>	↑	↑	↑	↑
	Bottom 25 <sup>th</sup> % (6.4) <sup>***</sup>	Bottom 25 <sup>th</sup> % (4.3)	Bottom 25 <sup>th</sup> % (0.1)	Bottom 25 <sup>th</sup> % (2.1)
	25 <sup>th</sup> % (4.4) <sup>***</sup>	25 <sup>th</sup> % (0.0)	↓ 25 <sup>th</sup> % (-1.8)	↓ 50 <sup>th</sup> % (-0.9)
	50 <sup>th</sup> % (3.5) <sup>**</sup>	50 <sup>th</sup> % (0.8)	50 <sup>th</sup> % (-0.2)	↓ 25 <sup>th</sup> % (-0.6) <sup>**</sup>
	75 <sup>th</sup> % (1.2)	75 <sup>th</sup> % (0.8)	75 <sup>th</sup> % (-0.9)	75 <sup>th</sup> % (-3.1) <sup>***</sup>
School sector	↑	↓	↑	↑
	Government (17.0) <sup>***</sup>	Government (-3.4) <sup>***</sup>	Government (17.8) <sup>***</sup>	Government (7.2) <sup>***</sup>
	Independent (12.2) <sup>***</sup>	Independent (-3.8)	Independent (13.8) <sup>***</sup>	Independent (9.6) <sup>***</sup>
	Catholic (17.1) <sup>***</sup>	Catholic (-6.4) <sup>***</sup>	Catholic (11.6) <sup>***</sup>	Catholic (4.2) <sup>**</sup>
School type	↑	↑	↑	↑
	Primary (16.2) <sup>***</sup>	Special (8.9)	Primary (16.0) <sup>***</sup>	Primary (6.9) <sup>***</sup>
	Combined (17.6) <sup>***</sup>	↓ Primary (-3.8) <sup>***</sup>	Combined (17.9) <sup>***</sup>	Combined (8.1) <sup>***</sup>
	Special (60.0) <sup>***</sup>	Combined (-3.6) <sup>*</sup>	Special (88.2) <sup>***</sup>	Special (47.5) <sup>**</sup>

Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

↑ denotes positive growth, ↓ denotes decline in scores.

<sup>1</sup>The analysis compared 2009 and 2013 since 2008 were not available.

## Progress on Achievement Gap.

For the purposes of this study, ‘achievement gap’ was defined as the difference between one school group and another along the same background attribute being evaluated. The gaps in NAPLAN scale scores and in the percentage of students reaching minimum standards are presented in Tables 7.5 and 7.6. On average, the magnitude of the achievement gaps is not very different between year levels and domains. The largest achievement gaps were found between primary schools and special schools, and NT and Victoria as evidenced in Figures 7.5 and 7.6. Each gap exceeded 100 scale score points in reading and was close to 100 points in numeracy. The gaps in scale score between the primary and special schools in Year 3 and Year 5 are equivalent to a near 40% and over 50% difference, respectively, in students achieving national minimum standards across both domains (Table 7.6).

Table 7.5

### *Average NAPLAN Score Gaps and Growth Trends (2008–2013)*

	Year 3		Year 5	
	Reading	Numeracy	Reading	Numeracy
NT VIC	125.4 (-0.10)	96.6 (-0.11)	109.8 (-0.08)	86.9 (-0.13)
Independent Government	34.0 (-0.14) *** ↑	27.2 (-0.11) *** ↑	31.4 (-0.16) *** ↑	27.6 (-0.15) *** ↑
Primary Special schools	104.5 (0.70)	87.5 (0.68)	112.9 (0.62)	98.8 (0.64)
25 <sup>th</sup>  75 <sup>th</sup> ICSEA quartile	87.3 (0.20) *** ↓	71.7 (0.20) *** ↓	78.3 (0.15) *** ↓	71.2 (0.13) *** ↓
25 <sup>th</sup>  75 <sup>th</sup> LBOTE quartile	7.3 (0.09) *** ↓	4.6 (0.10) *** ↓	1.4 (0.08) *** ↓	7.4 (0.07) *** ↓
25 <sup>th</sup>  75 <sup>th</sup> Attendance quartile	62.2 (0.16) *** ↓	52.2 (0.16) *** ↓	56.1 (0.08) * ↓	53.5 (0.07)

Note. \*\*\*  $p < .001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

The first number in each column represents the difference in average NAPLAN mean scale score over the six-year period between the two subgroups. The second number in brackets represents the beta coefficient from the trend analysis.

↑ Denotes a widening gap, ↓ denotes a narrowing gap, no arrow denotes statistically insignificant change.

Large proficiency gaps also exist between NT and Victoria. Compared to Victoria, 30–38% fewer NT students met proficiency standard across year levels or

domains. The proficiency gaps between high and low socioeconomic subgroups and high- and low-attendance subgroups were smaller for both years compared to jurisdictional and school type differences. In Year 3, they ranged between 11–13%, and in Year 5, 14–19%. Proficiency differences between government and independent schools were all below 10%. No gaps surfaced under the LBOTE measure.

Table 7.6

*Average NAPLAN Proficiency Gaps and Growth Trends (2008–2013)*

	Year 3		Year 5	
	Reading	Numeracy	Reading	Numeracy
NT VIC	31.8 (-1.69)*** ↑	29.5 (-1.76)*** ↑	38.1 (0.55)** ↓	33.5 (-3.60)*** ↑
Independent Government	4.7 (-0.32)*** ↑	4.6 (-0.23)*** ↑	7.0 (0.03)	5.5 (-0.66)*** ↑
Primary Special schools	36.9 (-8.88)*** ↑	39.1 (-7.36)*** ↑	55.9 (-43.74)*** ↑	51.3 (-35.95)*** ↑
25 <sup>th</sup>  75 <sup>th</sup> ICSEA quartile	12.8 (0.82)*** ↓	12.3 (0.56)*** ↓	18.5 (0.25)*** ↓	15.1(1.43)*** ↓
25 <sup>th</sup>  75 <sup>th</sup> LBOTE quartile	-0.4 (0.15)*** ↓	0.5 (0.16)*** ↓	0.4 (0.06)	1.5 (0.15)*** ↓
25 <sup>th</sup>  75 <sup>th</sup> Attendance quartile	11.7 (0.61)*** ↓	11.3 (0.40)*** ↓	16.4 (0.05)	13.5 (0.89)*** ↓

Note. \*\*\*  $p < .001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

The first number in each column represents the difference in average NAPLAN proficiency gap over the six-year period between the two subgroups. The second number in brackets represents the beta coefficient from the trend analysis.

↑ Denotes a widening gap, ↓ denotes a narrowing gap, no arrow denotes statistically insignificant change

Multiple regression results (in brackets in Tables 7.5 and 7.6) showed mixed movement in achievement gaps between advantaged and disadvantaged groups. Measuring by scale score, a large number of the gaps have narrowed and they were all statistically significant while others remained unchanged; by proficiency rate, the trends are mixed with gaps both narrowing and widening significantly. The scale score gaps (Table 7.5) between independent and government schools across year levels and domains have increased significantly, while other gaps have either decreased or remained unchanged. Among those that narrowed significantly were



the gaps between high and low ICSEA status, LBOTE representation and attendance level. No changes were observed between NT and Victoria, and between primary schools and special schools across year levels or domains. Under the proficiency measure (Table 7.6), special schools have lost ground relative to primary schools, and NT schools to Victoria schools across domains and year levels. The exception was Year 5 reading, where the proficiency gap between NT and Victoria narrowed significantly. Similar to the trends in scale score, the proficiency trends for other disadvantaged and advantaged pairs have generally narrowed.

### **Trends on NAPLAN Absence and Withdrawal**

Results from the withdrawal/absentee analysis indicate an increasing trend in these practices since the implementation of NAPLAN. As discussed in Chapter 5, withdrawal and absentee rates were collapsed into a new variable called 'exclusion rate'. Between 2008 and 2013, exclusion rates rose significantly across nearly all background categories. Special schools, schools in Queensland and SA had a larger than average increase of 3–4%. TAS, NSW and WA schools were the only groups without any significant changes. NT schools, the only school group that had avoided the upward trend in exclusion rate, recorded a 3% significant decline in Year 3 (Table 7.7). However, it must also be noted that NT schools began with the highest, and still had the highest, absentee and withdrawal rate among all jurisdictions in 2013 as seen in Figure 7.5. The average exclusion rate over six years was 12% in NT. Special schools had the second highest six-year average at 8% and it grew to 13% in 2013 (Figure 7.6). Schools at the lowest attendance quartile followed closely at 7%,

schools at the lowest SES quartile at 6%. ACT, one of the two top performing jurisdictions, also had a 6% average. The remaining groups averaged 5% or less.

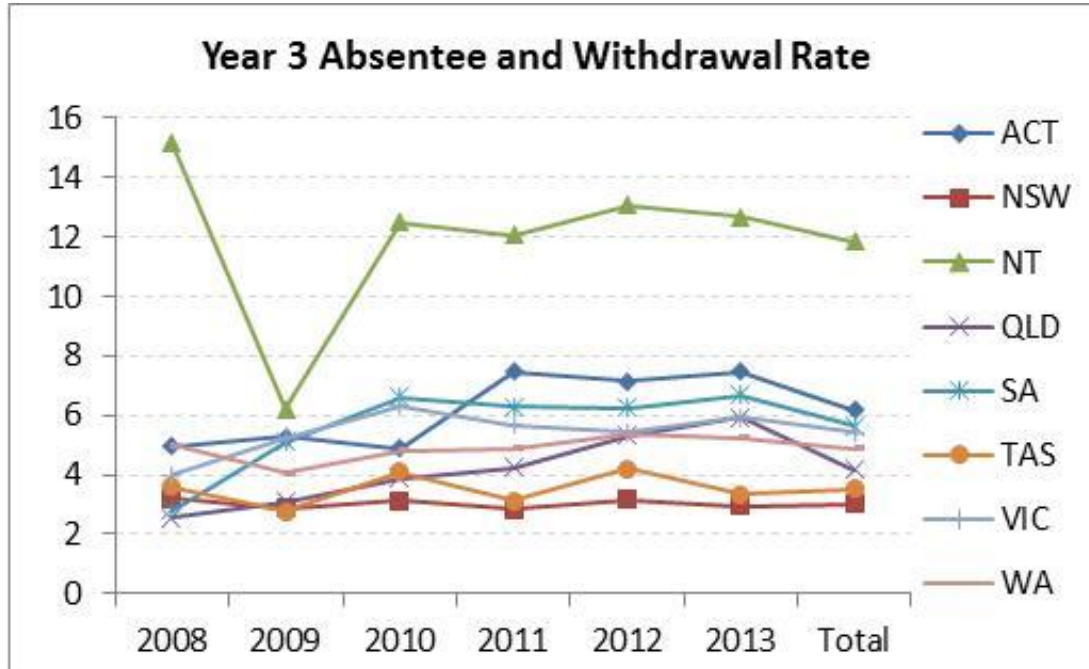


Figure 7.5. Year 3 Absentee/Withdrawal Percentage by Jurisdiction

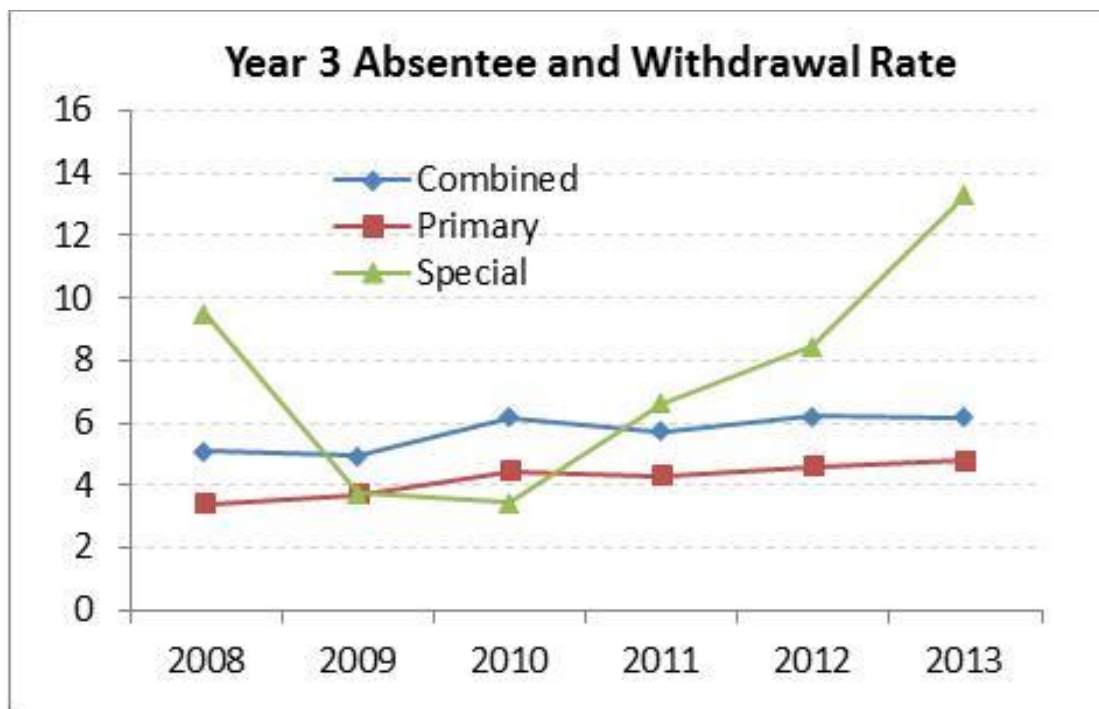


Figure 7.6. Year 3 Absentee/Withdrawal Percentage by School Type

Table 7.7

*NAPLAN Absentee/Withdrawal Percentage Change Over Time (2013 less 2008)*

<i>School subgroups</i>	<i>Year 3</i>		<i>Year 5</i>	
	<i>Reading</i>	<i>Numeracy</i>	<i>Reading</i>	<i>Numeracy</i>
Jurisdiction	↑ ACT (2.8)** QLD (3.3)*** SA (4.0)*** VIC (1.9)*** WA (0.1) ↓ NSW (-0.2) NT (-3.7)*** TAS (-0.1)	↑ ACT (2.3)** QLD (3.4)*** SA (3.7)*** VIC (2.1)*** WA (0.4) ↓ NSW (-0.2) NT (-1.2) TAS (-0.4)	↑ ACT (2.8)*** QLD (2.8)*** SA (3.0)*** VIC (1.2)*** WA (0.8) ↓ NSW (-0.1) NT (-3.1)*** TAS (-0.2)	↑ ACT (2.3)** QLD (2.9)*** SA (3.2)* TAS (0.1) VIC (1.6)*** WA (0.7)* ↓ NSW (-0.1) NT (-1.6)
ICSEA score	↑ Bottom 25 <sup>th</sup> % (0.9)*** 25 <sup>th</sup> % (1.9)*** 50 <sup>th</sup> % (1.6)*** 75 <sup>th</sup> % (0.9)***	↑ Lowest 25 <sup>th</sup> % (1.2)*** 25 <sup>th</sup> % (1.9)*** 50 <sup>th</sup> % (1.5)*** 75 <sup>th</sup> % (1.0)***	↑ Lowest 25 <sup>th</sup> % (1.2)*** 25 <sup>th</sup> % (1.2)*** 50 <sup>th</sup> % (1.3)*** 75 <sup>th</sup> % (0.6)**	↑ Lowest 25 <sup>th</sup> % (1.5)*** 25 <sup>th</sup> % (1.5)*** 50 <sup>th</sup> % (1.5)*** 75 <sup>th</sup> % (0.6)**
Attendance	↑ Bottom 25 <sup>th</sup> % (2.0)*** 25 <sup>th</sup> % (1.8)*** 50 <sup>th</sup> % (1.6)*** 75 <sup>th</sup> % (1.0)***	↑ Bottom 25 <sup>th</sup> % (2.3)*** 25 <sup>th</sup> % (1.7)*** 50 <sup>th</sup> % (1.8)*** 75 <sup>th</sup> % (1.0)***	↑ Bottom 25 <sup>th</sup> % (1.6)*** 25 <sup>th</sup> % (1.1)*** 50 <sup>th</sup> % (1.5)*** 75 <sup>th</sup> % (0.7)***	↑ Bottom 25 <sup>th</sup> % (2.0)*** 25 <sup>th</sup> % (1.5)*** 50 <sup>th</sup> % (1.8)*** 75 <sup>th</sup> % (0.8)***
LBOTE concentration	↑ Bottom 25 <sup>th</sup> % (1.8)*** 25 <sup>th</sup> % (1.1)*** 50 <sup>th</sup> % (1.3)*** 75 <sup>th</sup> % (0.6)*	↑ Bottom 25 <sup>th</sup> % (1.6)*** 25 <sup>th</sup> % (1.0)*** 50 <sup>th</sup> % (1.1)*** 75 <sup>th</sup> % (0.6)*	↑ Bottom 25 <sup>th</sup> % (1.5)*** 25 <sup>th</sup> % (0.8)*** 50 <sup>th</sup> % (0.8)*** 75 <sup>th</sup> % (0.5)*	↓ Bottom 25 <sup>th</sup> % (1.7)*** 25 <sup>th</sup> % (0.9)*** 50 <sup>th</sup> % (1.0)*** 75 <sup>th</sup> % (0.6)**
School sector	↑ Government (1.3)*** Independent (1.5)*** Catholic (1.2)***	↑ Government (1.4)*** Independent (1.5)*** Catholic (1.3)***	↑ Government (1.1)*** Independent (1.2)*** Catholic (0.9)***	↑ Government (1.4)*** Independent (1.0)*** Catholic (1.1)**
School type	↑ Primary (1.3)*** Combined (1.0)*** Special (3.4)***	↑ Primary (1.1)*** Combined (1.2)*** Special (4.4)***	↑ Primary (1.1)*** Combined (0.8)** ↓ Special (-0.5)***	↑ Primary (1.3)*** Combined (8.1)*** ↓ Special (-1.8)***

Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

## Discussion

The overall achievement trends observed in this study corroborate the limited number of previous studies (Ainley & Gebhardt, 2103; COAG Reform Council, 2013) evaluating primary student progress over time. On the whole, there was steady growth in reading across both year levels and modest to mixed results in numeracy with Year 3 showing more decline. As expected, the NT schools and schools serving special needs students particularly felt the impact of their disadvantaged backgrounds. Schools in the NT and special schools trailed their advantaged peers by a wide margin of about 30% and 50%, respectively. This finding is not surprising given the disproportionately large percentage of Indigenous students (41%) represented in the state of NT compared to 1% in Victoria or 2% in ACT, (Australian Government Productivity Commission, 2013). Furthermore, the same report shows that a quarter of the students in NT are in remote and very remote areas. Geographical remoteness has been identified repeatedly as one of the most important determinants of educational achievement in Australia (Bradley et al., 2007; Ford, 2013). Therefore, it is no accident that Indigenous students are considered the most disadvantaged in Australia by government and scholars. While it is encouraging that NT schools made significantly large gains compared to other school groups, the proficiency gaps between NT schools and top-performing jurisdictions are over 30% in both years and subject domains. These gaps have continued to persist, or have widened, since the launch of NAPLAN, despite impressive growth. Therefore, significantly more support under the Aboriginal and Torres Strait Islander Education Action Plan or other support programs are required to have an impact on the achievement gaps.

The same can be concluded for the achievement gaps between special schools and primary schools. However, given the limitations in the dataset discussed previously, it is not possible at this time to draw any conclusions on how all students with disabilities are making progress. Since students in special schools have very specific needs, it is likely that their results differ from those students with disability attending regular schools. While achievement gaps exist between disadvantaged and advantaged groups marked by socioeconomic status and school attendance rate, the gaps halved those observed between NT and Victoria. This further highlights the extreme inequity experienced by the Indigenous population. Furthermore, gaps between socioeconomic groups and gaps based on attendance level have been narrowing, albeit modestly. The smallest of all gaps, however, are those observed between government and independent schools. Notwithstanding the latter being better performing schools, the proficiency difference between the two groups was no greater than 7%. This small gap speaks to the overall equitable environment in Australia's education system as highlighted by the OECD (2010b).

One background characteristic that deviated from the general achievement trends is the LBOTE identification. The very existence of this category suggests that, in designing this category, ACARA had determined that the language background of a student, or of his or her family, predicts academic achievement. Various analyses in the current study have demonstrated the opposite. The results of this background factor contradicted the patterns observed in other background factors in three ways. First, this factor was not predictive of outcomes. Secondly, academic performance of schools in the top (highest LBOTE percentage) and bottom (lowest LBOTE percentage) quartile moved in opposite directions while the two higher performing

subcategories tended to be the middle quartiles. Thirdly, barely any achievement gap could be observed among the four different quartiles.

From these perspectives, the results seem to suggest that, as a group, students whose own background or whose parents' background is not English are not facing any challenge academically. In reality, however, the results might be validating the problematic definition issue raised by Creagh (2014) and Lingard et al. (2012). The LBOTE category's failure to capture the nuances of this student population suggests that this broad generalisation of language background might not capture the literacy or numeracy challenges that some segments of LBOTE students experience, or might overstate the challenges of other segments of LBOTE students. For example, the inclusion of students whose parents speak another language at home, even when those students might have been born and raised in Australian would clearly overstate their language challenge.

Lastly, the steady upward movement in absentee/withdrawal trends across each of school groups evaluated are indisputable and alarming. This trend substantiated qualitative findings in the literature of assessment exclusion (Athanasou, 2010; Thompson, 2013). Given the relatively small absolute changes, it may be too early to gauge measurable impact on achievement outcomes. However, given its upward direction and that higher absentee/withdrawal rates are generally associated with the disadvantaged school groups, this trend requires close monitoring. Furthermore, it is also important that research evaluates more than just the assessed rate – the measure used by ACARA, which only includes students who were present and exempt at the time of NAPLAN testing. Since the 'exempt' rate has

remained steady over time, the 'assessed' rate does not provide a full picture of test exclusion.

## Summary

Across jurisdictions, school types, school sectors, and school background characteristics, statistically significant progress has been made in NAPLAN reading among Year 3 and Year 5 students between 2008 and 2013. This progress is also evident in the Year 5 numeracy results. In Year 3, however, NAPLAN scores in numeracy declined across the majority of the school groups. Despite some encouraging gains, categories indicative of a less-advantaged school were consistently associated with lower achievement in reading and numeracy across both year levels. For example, large achievement gaps persisted between NT schools that serve over 40% Indigenous students and Victoria serving 1% of similar students, or between special schools serving only students with special needs and regular primary schools. Across most school characteristics, achievement gap movements between 2008 and 2013 were inconclusive. Unfortunately, the proficiency gaps between the two most disadvantaged groups (NT and special schools) and their advantaged counterparts have significantly widened or remained unchanged. Between high- and low-ICSEA groups, and high- and low-attendance groups, the gaps have narrowed. In sum, the trends are both encouraging and disappointing, and this speaks to the need for continued investment and support for the less advantaged schools and students. Lastly, the rise in absentee and withdrawal rates across school groups is alarming. The lack of participation cannot be good for the assessment, or for students who are being excluded because of the "potential for bias in estimates

of achievement” that could result from differential participation (ACARA, 2014, p. 324). Furthermore, it has the potential to inflate the NAPLAN progress in the long term if the practice of withdrawing the lowest-performing students from the assessment cited in the media (Cobbold, 2010) and identified in research (Thompson, 2013) continue.



## Chapter 8 Becoming Data-Driven Schools

Complementing the quantitative study presented in the two preceding chapters, the qualitative design of the present study sought to find connections between data-driven practice and trends in standardised assessment outcomes. Presuming linkage between data-driven practice and external assessment outcomes does not in any way suggest that other factors, such as curricular or instructional quality, play no part in student outcomes. However, just as the quantitative modelling in Chapters 6 and 7 concentrated on factors whose predictive powers for achievement outcomes have been demonstrated by a large body of evidence, the qualitative study focused on a key non-funding-related strategy of current education reform – data. This chapter reports the findings of six qualitative cases conducted in New South Wales, California and Hawaii. The findings drew upon interviews, observations and artefacts collected on site, and on publically available school annual reports and newsletters.

The presentation of the findings is structured as follows. The first section introduces the six participating schools. It aims to highlight the unique context in which each school operates and how this influences their decisions to implement data-driven practice. This section also clarifies the data parameter on which the semi-structured interviews were based and explains the support structure employed by the participating schools to assist students who fall behind. The second section maps the evolution of data-engagement at the participating schools. It reviews how schools and teachers perceived and used data, and the challenges they encountered in the process. Results from all three locations, and including both teachers and

administrators, were combined in the analysis and are reported as one set of findings. However, where a significant difference exists between countries, across schools, or between teachers and administrators, the difference is highlighted in a final section devoted to similarities and differences. Furthermore, unless a distinction is noted, the term ‘educator’ refers to both administrators and teachers collectively. Following this chapter, Chapter 9 applies the theoretical constructs to interpret the belief systems of educators that led them to the behavioural change necessary to adopt and to constructively engage with data as a valued end.

### **Case Context, Data Definition, and Student Support Structure**

According to Guba and Lincoln (1994), context and realism are necessary components of research. This is because causes and effects in any event are neither context- nor value-free; instead, they are interdependent. Therefore, this chapter begins by providing the contextual background of the participating schools by directing attention to the “complex and conflict-ridden social reality” (Byrne et al., 2009, p. 520) present at these schools. It will demonstrate how these social realities became an integral part of schools’ decisions to engage with data as part of school operation and teaching strategy.

#### **Case context.**

The six selected cases took place in three locations: NSW in Australia, and California and Hawaii in the US. For anonymity, the participating schools in NSW are referred to as Bilby and Koala Schools; in California, Almond and Walnut Schools; and in Hawaii, Kukui and Hibiscus Schools. Field work consisted of semi-structured one-on-one and small-group interviews, school- and grade-level meetings and

professional training observations, and collection of artefacts. It should be noted that the opportunity to observe in meetings only arose at two schools. Altogether, 43 participants were directly interviewed and 17 participants were observed by the researcher in their respective grade-level meetings. These numbers overlapped because only a fraction of the teachers observed in meetings also participated in the interviews. Furthermore, the total number of participants observed presented here excludes participants in the full-school data meeting and cross-district literacy professional training, which the researcher was invited to observe. There are two reasons for their exclusion. First, participant counts are large (over 30 each) and the researcher did not directly speak to any participant (unlike the grade-level data meetings). Secondly, the significance of these meetings, which will be discussed later in this chapter, pertains more to the very existence and purpose of these meetings than to the dialogues exchanged. Among the 50 participants interviewed, 17 were administrators and 33 were teachers. Of the teachers, 24 were classroom teachers or single-subject teachers and nine were special resource teachers. Among the administrators, five were instructional or data coaches and leaders, six assistant or deputy principals, four principals, one district accountability officer and one external data consultant.

As noted in Chapter 5, the six case studies varied significantly in school background, student population served and student performance. Coloured by their contextual differences, there was clear evidence that each school, each grade level and each teaching team interpreted policy through a different lens, as Fulcher (1989) conceptualised. The following vignettes highlight qualitative differences among the participating schools to set the context for the findings.

***Koala School, NSW.***

At the time of the field work, Koala School was re-emerging from a transformation process. Enrolment had dropped at Koala, the principal expect to be reassigned by the Department of Education to a larger school after this, his fourth, year. He would have liked to stay longer to finish the change he had implemented, believing that “in terms of culture change, that is [four years] probably a little bit short I think”. Based on his own account, as well as that of his executive members and teaching staff, Koala had undergone a massive cultural change in every aspect of its school operation. Prior to the principal’s arrival, student behavioural challenges had dominated day-to-day life at Koala and they continued to do so for the first year of his tenure, when he “suspended students 94 times”. He further stated that students were being “abusive to adults and teachers, telling teachers to f\*\*\* off and running out of school”, they “thr[ew] chairs” and “bags off balcony”.

Student behavioural issues paralysed all academic efforts and school morale.

One of the assistant principals summed up the challenge this way,

We [the administrators] were not setting a great example because the leaders were so caught up with all the flustering and the problems of dealing with negative behaviours, and putting out spot fires everywhere. We could not model good teaching practice...

Teacher turn-over was exceptionally high, not merely due to student behavioural problems, but because of a young teacher population of child-bearing age. According to the principal, ‘casual teachers’ (non-permanent or short-term staff) were the norm, rather than the exception.

This principal, a strong proponent of data, began to change school culture using data to inform and drive his decisions. He contended that the school had

always collected a large amount of data yet never used it. Merging disparate data points, the principal identified trends in the challenges the staff faced (e.g., at what time of day or during what type of activities did most suspensions happen?) and systematically addressed them. Two overriding trends surfaced that he believed had an impact on many of their difficulties: (1) the school was involved in too many programs and activities; and (2) students were not engaged. With this conclusion, as one of the assistant principals recalled, the school “cleared the deck” of everything and started from the ground up”. The leadership team cancelled all non-academic-related activities, including (as reported by the same assistant principal) “getting out of our sport programs” and getting the school “to be just about academic school development”. The principal instructed his faculty to use data to “program on a three-week cycle. So they are able to set their objectives and achieve them with the kids”. The school went back to basics, which another assistant principal described as akin to “the McDonald’s approach a little bit to get that simplicity”. With more clarity, both executive and teaching staff discovered that,

The curriculum and the learning part of school can be a really good tool to control negative behaviours. We ha[d] a kind of a view here for a long time that we had to fix the behaviour first then the curriculum. (Assistant Principal, Koala)

Furthermore, the principal stated that “there has been a cultural shift in the school about when we want to engage with parents, what is the conversation to be about”. The conversations moved from “deal[ing] with complaints six hours a day” from parents, to discussing the “business [of] teach[ing] kids” reading and mathematics.

These changes were drastic, but the results were equally dramatic. According to the principal, student suspension dropped to 20 in the year when the present

research took place. One of the two assistant principals, who had been at the school for a long time, noted:

There is a huge cultural change. We have gone from almost cloak and dagger existence and a fear in the school to it being a very open culture of being able to say 'I cannot do this, I do not know how to do this, can you help me?' whereas in the past we have not been able to do that.

NAPLAN results were also on the rise, particularly in Year 3; Year 5's results were slightly more mixed. The higher level education authority also had recognised the progress at Koala. The principal remarked,

When I had my meeting with my boss, this is what she talked about, that is the strength of the school at the moment that we are able to get those kids up here from that bottom.

Parent satisfaction had also gone up as reported by the survey results in the annual school report. This transformation, according to the principal, is attributable to their intentional use of data.

Without the data and the change in the curriculum to get the kids hooked in properly, we would still be in chaos. That is the long-term change in the school.

### ***Bilby School, NSW.***

Unsolicited accounts from participants painted Bilby as an extremely organised school where detailed guidelines for all teaching-related matters are outlined in a large 'Personal Portfolio' of which every teacher owns a copy. Having received a copy of this portfolio, the researcher can attest to the Year 5 teaching team's description of it as a manual containing detailed "guidelines ... for what they expected us to be doing with kids and where they were supposed to be", "it has everything in it that you are expected to have in your program...". It also provides

information on “what to track as well, what requirements we have in data, when we are collecting the data”. There were also guidelines for behavioural management, parent communication, report writing and more. As explained by an assistant principal, these guidelines and structures “guaranteed that I am going to walk into all the classes, and we are teaching the same way and we are expecting the same things from the students”. The deputy principal believed that such structured support is not common across schools, as she noticed that some teachers new to Bilby had never been exposed to a similar structure: “some people coming from some schools and they look at all that data collection and oh my God, it’s so formalised”. There is a general perception that the structure is good. A Year 3 and 4 composite teacher described it as “very systematic here”. However, she quickly added, “I felt coming from my last school... having that [Bilby’s] structure... I think I have become a better teacher”.

Perhaps it is this well-defined program that has contributed to the school being selected as a recipient of the *Centre of Excellence award for Improving Teacher Quality* under the Smarter Schools National Partnership supported by the federal government. Bilby’s reputation, according to the lead teacher in Year 5, brought in over 80 applicants for two open teaching positions in the prior year. Part of the school’s focus on excellence is setting high expectations for students, something all Bilby executive members in the study articulated. Despite the fact that 90% of Bilby’s students were of LBOTE background, Bilby’s NAPLAN scores were above the average for similar schools across year levels and all domains except for Year 3 reading, where Bilby trailed by 11 points during year of the field work. Yet the executives

were not resting on their laurels; they set new goals to move more students to the upper bands on NAPLAN.

Yes it's great that we don't have many in the bottom but it's not good that they are sitting here [middle bands] when we can make them better. So that's the area that we focus and we want to get that middle slump moving because it's not just okay to say they are doing okay.

***Almond School, CA.***

According to news reports<sup>3</sup> and the principal's account, the original Almond School was shut down less than ten years earlier as a consequence of: poor student outcomes on the CST; overcrowding; and high staff turnover. A new and smaller Almond School opened in its place. The redesign included the principal interviewed for the present study, who was new to the school, a duo-language immersion program and a philosophy of learning anchored by the following three pillars: language and culture focus; expedition learning; and family, school, and community Integration. Under the new leadership and new program, Almond won one of the few annual National Title I Distinguished School Awards in California. The award recognised Almond for raising student proficiency in mathematics from under 30% to over 70%, and in reading from under 20% to over 60% within five years of its redesign. This achievement came in spite of the fact that 96% of Almond's students came from the low-SES strata, 44% were English learners and the suspension rate was 12%. It also achieved an API score well above the state-wide goal of 800. All except for the low-SES and the disability subgroups achieved an API score of 800 or

---

<sup>3</sup> Citation is omitted to protect the anonymity of the school.



higher at the time of this research. The low-SES subgroup, in fact, was only two points shy of the API target.

These impressive achievements, however, appear to accomplish under a rather structured programming and instructional environment, where data are central to daily life of the school. Students' academic performance was colour coded and charted on a regular basis. The principal explained,

We colour code students on their first assessment... then after the second assessment which is going to come up at the end of this month, we move the cards.

The cards referred to were student name cards colour coded by assessment scores. These cards were slotted into pockets on scrolls and plastered all over the principal's office wall. Appendix E contains an example of a colour-coded worksheet used to direct where the individual name cards should move from week to week for target intervention. The principal explained the process,

So if a child scores in that category in the red, at that point that triggers an additional kind of assessment or diagnostic and progress monitoring ... Then there are kids who are in that approaching range which will be the yellow range. And if they score between like 50 and 70% on the test, so [at] that point the child[ren] [are] reading the test, they are understanding it, but they are getting a lot wrong. And so those students need some sort of repeat teaching or intervention that can often be done by the classroom teacher. And then you have the kids who are on benchmark or above, scoring 70% or more... oh, here you see the nice coloured graph... The level of analysis can get very, very detailed...

Compared to other principals in the study, Almond's principal was the most intimately involved with individual student data. Her intimate knowledge of many students suggested that she was as hands-on with student data engagement as her teachers. Her personal involvement was likely a result of the fact that her school is

small; it was the only school among the six participant schools without instructional coaches or assistant principals.

***Walnut School, CA.***

By contrast to Almond School, the environment at Walnut and its school district appeared more flexible. Although Walnut followed the Professional Learning Community (PLC) process conceptualised by DuFour and Eaker (1998) in the way teachers meet, plan and work with each other using data, each grade level was given the flexibility to program, instruct and assess as they saw fit to support their students. The instructional coach explained,

We have to use some of the curriculum that we are given to some degree ... but teachers have some flexibility as to how – what their next steps are based on the data.

For some grade-level teams, that meant co-teaching. The third-Grade teaching team decided to engage in a co-teaching environment where they rotate the entire grade's students among the four teachers on the team. Rotation decisions were based on the subject, student and the teacher's mastery of the topic being taught, and prior grade-level taught. The rotation afforded teachers the opportunity to know every student in their grade, as well as students' individual strengths and weaknesses. According to the instructional coach at Walnut, a total of three grades practised the same model.

Walnut district had the same philosophy regarding professional learning communities and data use. The district data officer expressed the view about working with schools that:

They [the schools] do have a lot of flexibility because it's really a results-based focus, so this is where we are headed, you know, we

are on track 'great you know what to do, fabulous'... I want to honour teachers wherever they are.

In response to a question about where the rest of the schools in the district were regarding data practice, the instructional coach at Walnut School validated the district data officer's claim regarding flexible management by district officials:

Our school has probably done them the longest, and then a couple of schools maybe followed within a year after. Some schools are just getting started or are trying to get started on them and just have not caught on.

***Kukui School, HI.***

Kukui School is a low-performing school where just slightly over half of the student body met reading (56%) and mathematics (53%) proficiency, and only 18% met the science proficiency at the time of the research. Similar to Koala School in NSW, Kukui School was also a school characterised by significant behavioural challenges until three years ago. According to the principal, who had been at the school for many years,

We used to get 10, 20 referrals a day from the ELL class alone, because kids weren't listening, they weren't paying attention. On the way to the class they'd get into fights, on the way back from class they'd get into fights. Their attitude in that class was horrible.

With a third of the students being English language learners, the disruption from moving in and out of class for intervention is easily imaginable. After some reflection, the administrative staff and teachers believed that their pull-out teaching model created the behavioural problems. Kukui's principal reasoned,

It's a society in your classroom. As soon as you start taking students out and saying, 'you have to go to a special class,' it makes them feel like they're not part of the class and it lowers their self-image... I now believe that; just because I have friends who've been pulled

out into special education classrooms their whole entire school life, so they don't expect much of themselves.

Coincidentally, this principal's philosophical belief is supported by the PISA findings (OECD, 2012b), which revealed that streaming based on ability erodes student motivation and performance. Kukui's solution, according to the principal, was inclusion.

[It's been] three years we've changed, and did full inclusion for special education, and then two years ago we did full inclusion for ELL.

We need to do that for our kids, because they live in an area where so little is expected of them. So we want them to know that they can get themselves out of there [their low income environment] through education.

What the school hoped to do was moving away from "perpetuating the cycle of low economic situation", said the instructional coach. The principal was pleased with the results thus far.

Look at our ELL scores jumped by almost 20%... So can't argue with that... When you pull them out into a separate room where all of them are either at the same level, they don't see their student models, the student peers, so their growth is minimal.

Indeed, public accountability data (Hawaii Department of Education, n.d.-b) validated Kukui's overall steep growth curve in students meeting reading proficiency from 44% two years prior to 59%.

Kukui also stood out as being a school operating under a highly structured curriculum and instructional environment among the schools in the sample. According to the principal, every school leader and faculty member knows "the three things that we're going to do consistently and pervasively, every day, every grade level". Instructional coaches walked around in their walkthrough to monitor that "all

the grade classes were doing the same thing”. During the morning of the interview for the present study, the instructional coaches found that two classes were “a day behind” on the pacing guide (Appendix F), which is highly structured in and of itself. Just as in Almond, the office and hallway walls were full of posters with coloured charts of assessment results.

***Hibiscus School, HI.***

Among the six participating schools, Hibiscus was a late adopter of data-driven practice. This was so perhaps because the school’s performance had been above the state average hence it may have gone under the radar of the state education authority. Nonetheless, not every student subgroup had met proficiency requirement under NCLB, particularly students with disability. For this reason, the principal’s improvement plan included data as part of the strategy. However, not every teacher was yet on board and the instructional coaches still performed a significant portion of the data-driven activities. For example, they compiled data sheets with test results for analytical purposes at the grade-level data meeting. At other schools, teachers were responsible for data compilation and interpretation. In addition, these Hibiscus instructional coaches supported teachers in data meetings using largely scripted meeting protocols (Appendices G and H).

The principal had been at the school for only two years. While a big believer in data-driven practice, he opposed highly structured data use because he believed it was too “oppressive” an approach, and considered it an “abuse” to monitor teachers and students with “a chart” to see how close they are to goals on a regular basis. Instead, what he aimed to create at Hibiscus was “system congruence”, which he

explained as an alignment of goals, curriculum, and instructions. Being a strong proponent of inclusive education, he instituted inclusive classes where students with disability learned alongside mainstream students in one of the grade-level classes. In these classes, general teachers and special resource teachers taught alongside each other. He had been pleased with the results in the inclusive classes:

The inclusion classes except for one grade level, they are either exceeding their peers on the grade level or they're right on par.

This claim might have been somewhat inflated. Based on results found in the Hawaii accountability data centre (Hawaii Department of Education, n.d.-b), students with disability were still behind their peers in reading and mathematics in the year in which the research took place. However, students with disability did make progress particularly in mathematics proficiency, which moved from 24% two years prior to 30%.

### ***Patterns of Data Practice***

As demonstrated above, each school's unique context produced varying motivations for data engagement, as well as differing application of data-driven practice. Their narratives of data practice can be loosely categorised into three overriding data engagement models: (1) strategic; (2) tactical; and (3) day-to-day (Figure 8.1). While every participating school appeared to engage with all three models, and this is particularly true of the tactical model, the contextual differences and leadership visions aligned some schools closer to one model than another.

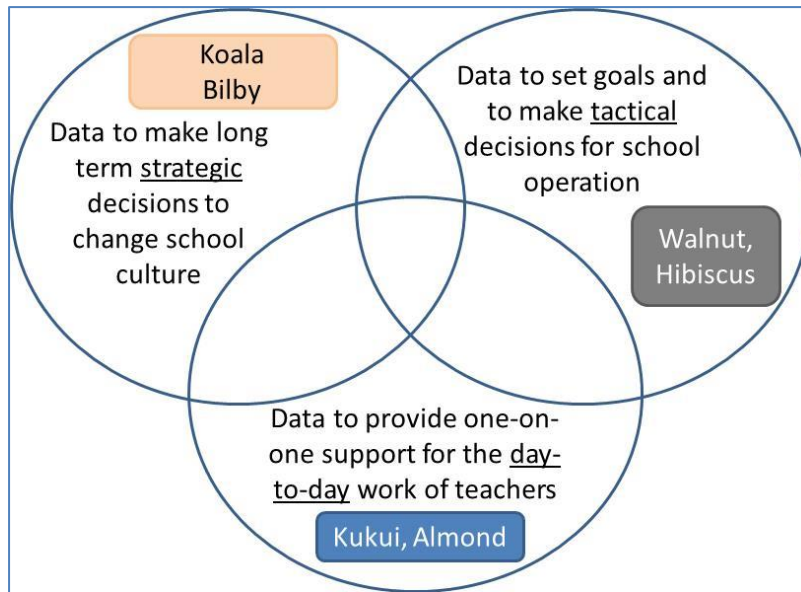


Figure 8.1. Data Engagement Models

All participating schools spoke of using data to obtain an overview of the school’s performance so that they could chart a course of action to make improvement. This is tactical use of the data, and is best expressed by the deputy principal at Bilby, whose view is representative of other schools:

So a lot of NAPLAN information helps us to identify key areas and need. Then using other school data, we can actually see if there’s correlation across the school and then put that into a management plan and certain area that we need to focus on.

Beyond tactical plans, however, two schools used data to inform strategic decisions. At Koala, the principal relied on data to solve larger problems and to make longer term investment decisions. He invested in playground equipment to reduce suspensions after noticing, through data, that most incidents happened in the playground. The principal also credited data and education research for his decision to implement a mathematics activity room “to get children hooked on the learning”, which resulted in the “decrease in abhorrent behaviours”. Similarly, the decision to invest in interactive white boards in every classroom was also informed by data.

We saw that three teachers who had interactive whiteboards, their students outperformed other students in the school, that is how we justified getting interactive whiteboards into the others for literacy. At Bilby, the executive members spoke about how they needed to liaise with

high schools to share their data practices, which helped them decipher the exact needs of the LBOTE population. They were not only concerned about their students' success in primary school, but wanted to ensure that these students maintained their growth trajectory as they moved to secondary schools.

In contrast, at Almond and Kukui, the schools appeared to be governed strictly by the large and small goals set at the beginning of the year, based on data. They engaged in an intense data-driven process where data was central to day-to-day teacher support. For these schools, the teacher data meetings ensured that nothing swerved off the path. Kukui's principal explained the purpose of data meetings, as follows:

Every week the teachers have an articulation day. So every seven days one grade level will meet... they have the whole day to sit, look at data, discuss with the coaches, and the special education teachers how the kids did, what they still need to work on, and then they'll change their lesson plans according to what the kids still need to work on.

At Almond, the principal was intimately involved with every teacher to monitor, as well as to guide them, on a one-on-one basis,

So after the teacher gets score report... [they] do item analysis and they can really see what question do they get wrong and how do I – oh, here you see the nice coloured graph. Then the teacher completes analysis of the test data. And when we started doing this all of this was done in one-on-one goals conferences, and I would meet with each teacher one-on-one and we would look first what whole class standards needed to be done for re-teaching.

The differences in policy interpretation and implementation illustrated here affirm the concept that policies are subject to interpretations and local adaptation



(Fulcher, 1989). What might appear to policy-makers to be a straight-forward policy directive on paper is anything but that because “actors adjust their activities based on situational dynamics”, as Wayman et al. (2012, p. 162) concluded from their work with school districts on data use.

**Data definition.**

In the world of school education, data can have numerous meanings. There are data associated with demographics, enrolment, absenteeism, matriculation, standardised assessments, internal assessments, survey results and financial data to name a few (Mandinach & Jackson, 2012). With the exception of Koala School where attendance and discipline referral data were also discussed, during the case interviews and in the reporting of these findings, reference to data or data use pertained strictly to student-performance data defined as external and internal assessment data. In fact, none of the participating schools included only external testing data in their data analytics. In the two Australian cases, external data refers to NAPLAN; in California, the California Standardised Test (CST); and in Hawaii, the Hawaii State Assessment (HSA). In each of the six cases, internal data included literacy program evaluations, assessments associated with specific intervention programs being implemented at the time, or teacher-created assessments as well as benchmark assessments. In the four US cases, there was an additional assessment – the benchmark assessment. Schools administered these three times a year to measure progress as students worked toward the targeted proficiency level on the annual external assessment. According to multiple participants, these benchmark assessments informed overall gaps in student learning and were good for student

grouping and lesson creation. However, according to a Grade 3 teacher at Walnut, they are not ideal:

It didn't really give us that specific detail as to each student what their needs were, it was just kind of like they get it, they don't get it, they kind of get it and so we were stuck as to what, how do we instruct them if we don't know what their needs are.

For more detailed diagnosis, the US schools in the sample used programs such as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), Developmental Reading Assessment (DRA) and other programs, as well as teacher-created assessments. In the two Australia cases, Best Start assessment was the program used to gauge entering kindergarteners' numeracy and literacy skills. Based on the results, intervention programs, such as Reading Recovery and Multi-Lit were provided to Year 3 students. Each of these programs comes with its own assessment schedules to monitor progress. From these assessments, teachers modified their pacing accordingly to support students. While the programs mentioned by the two Australian schools are an ongoing commitment made by the NSW Department of

Table 8.1

*The Meaning of 'Data' in the Interviews*

<b>School</b>	<b>External Data</b>	<b>Internal Data</b>
Koala	NAPLAN	Attendance; discipline referrals; Best Start assessments; Reading Recovery; Multi-Lit; teacher created formative assessments
Bilby	NAPLAN	Best Start assessments; Reading Recovery; Multi-Lit; teacher created formative assessments
Almond	CST	Triannual benchmark assessment; DIBELS; teacher created formative assessments
Walnut	CST	Triannual benchmark assessment; DRA; teacher created formative assessments
Kukui	HSA	Triannual benchmark assessment; DRA; teacher created formative assessments
Hibiscus	HSA	Triannual benchmark assessment; DIBELS; teacher created formative assessments

Education and Communities (n.d.-a) in all government schools to ensure students are on track with literacy and numeracy, the programs in the US samples were chosen by the school or district as there were no Department of Education mandated programs. Table 8.1 summarises the data sources referred by the individual participating schools during the interviews.

### **Support structure for low-performing students.**

From this study's quantitative results and existing literature, disadvantaged students tend to perform lower than their peers academically. Therefore, it is crucial to understand how the participating schools support their academic needs. The support structure described by participants can be summarised into four different models (Figure 8.2). For the most part, it was the level of the student's reading proficiency, more than their mathematics proficiency, that determined the type of support structure received. Across six participating schools, the push-in model was commonly used to support students whose reading proficiency was not significantly behind. This approach is consistent with what Idol (2006) described as the 'supportive resource program', one of the four forms of service delivery to support students with special needs. In this model, the special education teacher goes into the classroom to work with students who require additional support. These students could be students with disability, ELL or any low-performing students. Participants in the present study used this approach because it allowed these students to be exposed to their grade-level activities. The instructional coach at Walnut explained,

A lot of those kids, they are not in the regular classroom a whole lot, they were going out to this class for that, they are going out to that class for that... These kids were not at your classroom 5 hours, 6 hours a day, you are kind of like, 'well, what would I teach them

when they are not here?’ and so it’s been much more inclusive to have those kids and pretty much more [sic] I think beneficial for them because even if they are working at a different level, they are at least exposed to what all the rest of third grade is doing.

For students who were further behind and whose needs required targeted intervention, the pull-out model was common at all schools but Kukui. At Bilby and Koala in Australia, pull-out was the default model of the Reading Recovery and MultiLit programs which specify one-on-one interventions. Instructional Coach No. 1 at Hibiscus described her school’s pull-out model this way:

They pull out kids basically [for] the three of us as coaches. We also have counsellors and we also have off-ratio people like our Physical Education teacher and our Art teacher. We all pull out kids that are in this [the low] group and we give them special services in order to help them get over that mark for the benchmark for the HSA.

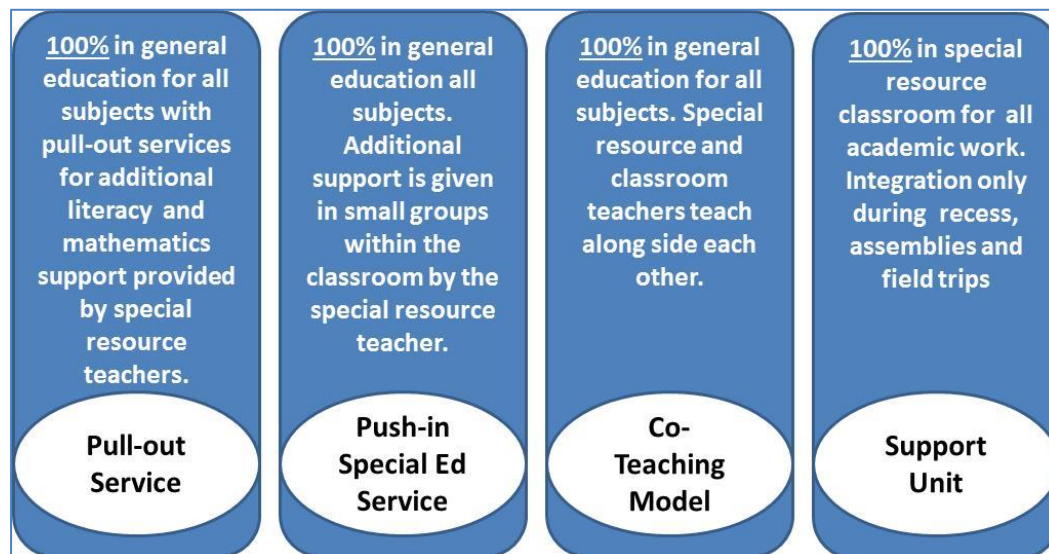


Figure 8.2. Models Supporting Low-Performing Students

The co-teaching model applied only at the two Hawaiian schools, although not uniformly. In this model, special resource teachers and classroom teachers were part of the same grade-level team, hence they taught and planned together.

We’re an inclusion school, so the kids are not pulled out... So there are special education teachers, and we have one ELL teacher. She

goes into the classroom too and works with the kids in the classroom. (Kukui Principal)

At Hibiscus, the fourth-grade team meeting that the researcher was invited to attend consisted of both special education teachers and classroom teachers who confirmed that they were a co-teaching team. The evidence of their planning and their description of co-teaching resembled the ‘cooperative teacher model’ described by Idol (2006), where special education and mainstream teachers work together to support all students. However, it is important to point out that students who were previously being supported in a separate environment were distributed to only one of the four classes in each grade level at Hibiscus. Furthermore, students who were significantly behind could still be pulled out for targeted intervention by the three instructional coaches. The principal had only just implemented what he considered the ‘least restrictive environment’ after noticing the lack of cooperation between special resource and general classroom teachers,

That gap between the two is so profound that the teachers started to develop a line there and they said that these kids can’t do...

In contrast, the support units were only available at the two Australian cases. Their special units were among the 1,500 plus support classes across NSW government schools set up by the Department of Education (New South Wales Department of Education, n.d.) to support students with special education needs. Families interested in placing their child into one of these units must request access from the NSW Department of Education and placement has “to be signed off by a district guidance officer”, as stated by Koala’s principal. According to the assistant principal overseeing these multi-category units at Koala, “probably 75% of our kids coming in on Special Transport Services are out of area”. However, once at Koala,

students, teachers and administrators in these units are incorporated into the school and become part of the school's responsibilities.

We plan with the rest of the school, we look at the targets for the rest of the school, and then we just make modifications and accommodations for our kids. (Assistant Principal of the Special Unit, Koala)

Integration with mainstream students only happened for social activities, for example, sport carnivals, recess, excursions and assemblies. Nonetheless, teachers and administrators both believed that inclusion in these activities constituted inclusive practice. Yet, when asked if mainstream students knew that students in these support classes had different needs than they did, a support unit teacher answered, "Yes, they do know and they accept them [special education students]".

As to how the special unit students felt about themselves being in a separate class, a Bilby support unit teacher offered this example,

There is only one student in the support class that she has been asking me, 'Am I in a support class, is this a support class?' like she sensed that something is different here.

Walnut's approach incorporated three of the four models. They had pull-out, push-in and self-contained service, and they had resource teachers and special education teachers. According to the instructional coach,

So resource teacher is brand new this year... So he does a support group in the morning for fourth and fifth graders outside, and then he pushes into classrooms and supports his kids with IEP's in the Special Day class... His aid does the same type of thing and we have one Special Day Class for K-2.

There were fewer than ten students in the Special Day class and each class had "one adult for every three or four kids". She continued,

The kids in Special Day class, they are in that contained classroom with the special education teacher all day... They are not really part

of the professional learning community program [grade-level data meetings] because their IEPs are so specific.

In sum, just as for data-engagement practices, support structures for students who were behind also differed significantly among the selected schools. More notably, what teachers and administrators deemed 'inclusive' also varied considerably. The educators interviewed spoke about the need to support all children in an inclusive manner, where diversity (be it social class, ethnicity, religion, ability or attainment) is welcomed, recognised and valued, a concept supported by inclusive scholars (Ainscow, Booth, & Dyson, 2006), the differing models of support provided suggest that principles and rhetoric, when translated to practice, can have very different meanings.

### **The Evolution of Data Engagement**

This section traces the path whereby participating schools began to embark on data practice. The interviews sought to understand the catalysts for data engagement and the eventual uses of and reliance on data. In all cases, participants reported that their schools, or they, had always kept internal assessment data. However, they previously did not regularly or systematically use data to determine which student needed help and to structure their day-to-day lessons to provide the needed support. For the majority of the participating schools, formalisation of data engagement had only begun to take place during the past two to six years. When asked what led to data engagement at their current level of intensity or formality, the deputy principal at Bilby replied,

Pressure to be accountable and also... one of the deputy principals here, that [data] was kind of her baby. Even though we had those

procedures in our school it probably wasn't as consistent or as evident throughout the school [until she came].

In the US, the need to meet strict accountability goals hastened the need to track progress through data; a process not every participating school found useful initially. The data officer at Walnut District recalled,

Where we are today is really far ahead of where we were five years ago, we were still using data but it's you know it's a shift... When I was principal at one of the sites, we had a data management system at a site level [which] we really did not like at all, and we didn't find it user-friendly, and we would do it because we were asked to do it.

In Hawaii, "the State wants all schools to start looking at data, building school-wide data teams", said Kukui's principal. However, she pointed out that Kukui began "grade-level articulation days for the last six or seven years" ahead of the mandate. For Kukui and Almond, accountability pressure was particularly acute due to the disproportionately large number of disadvantaged and low-performing students in those schools who did not meet the state benchmark. Both schools turned to data as a way to inform instructional practice and interventions to raise student performance. In fact, data-driven practice was one of the key strategies launched when Almond reopened its doors. Both schools had experienced growth through the data-driven process and so "every year we tweak it and we get better at looking at data," said Kukui's principal.

Accountability requirements of a different nature also sped up the adoption of data-driven practice at the two Australian Schools. Both were recipients of National Partnership Program funding. According to the principal at Koala, the school's foray into data was linked to this funding.



The school did not utilise data very much at all. They started to just [sic] before I came because of the need for National Partnership on Literacy and Numeracy.

The 'need' that the principal referred to was that, to be considered as a National Partnership candidate and continue receiving funding in the grant cycle, the school had to demonstrate data-supported self-assessment and annual academic growth. Regardless of requirements, participants pointed to societal demand to be more accountable and transparent as a catalyst for more formalised data practice.

An assistant principal at Bilby summed up the trend as follows:

I think nowadays and it is not just because of NAPLAN, I think it is because schools are the way that schools are becoming that we are collecting more data and we are being more accountable, so that we can tell parents where the child is sitting and how to make individual improvements and goals for certain students. So I just think because there is a need in our society to have more information about how students are going.

Accountability pressure and policy requirement aside, having a data champion in the leadership team also expedited data use at multiple schools, including Almond, Koala, Bilby and Hibiscus.

Over time, accountability requirements and leadership vision, as well as positive outcomes derived from data-driven practice, led to the adoption of grade-level meetings where systematic data review took centre stage and served as the anchor for goal setting, and curricular and instructional development. All participating schools except Almond held weekly or twice-a-week grade-level meetings ranging from two hours per week at Bilby to an entire day in a seven-day cycle at Kukui. At Walnut and Hibiscus, these meetings were called the 'professional learning community', at Kukui, a 'grade-level articulation day' and at Bilby and Koala, they were referred to as 'grade meeting' and 'stage meeting'. These grade-level data

meetings differed in name, but were unified in purpose. All were geared toward diagnosing learning trends, informing curricular and instructional practices, monitoring progress against goals, and evaluating the effectiveness of the curriculum and instructions to support student learning. These meetings also served as the forum where teachers come together to collaborate and share best practice.

**Perception of data utility.**

External data from standardised tests required under the current education reform lent marginal value to participating teachers. Principals in the studies found them slightly more useful. This is not unexpected as principals are more concerned with general school-wide trends and teachers are more focused on individual student needs. Consistent with findings in other reports and studies (The Senate Standing Committee on Education and Employment, 2013; Young & Kim, 2010), the long cycle of external assessment results lessened the value of external assessment results for most participating teachers. A Year 4 teacher at Walnut District complained,

We don't get that information till August or September<sup>4</sup> and we have already made the classes because they [students] have to know who their teacher is by August... So, to wait till September to get that official data and then put them in their class you've already wasted four weeks of instruction. So to that point, that's why it's not valuable.

Teachers also discounted the value of external assessment for its generality,

They [CST] are very useful for students who are on grade level and they are very useful for discrete skills. They are not useful for students who are far above grade level or below grade level. (Year 3 Teacher, Almond)

---

<sup>4</sup> The school year for most Californian and Hawaiian schools begins in mid to late August and ends in early to mid-June. In Australia, it begins in late January and ends in mid-December with more breaks during the school year.

For US teachers in the sample, benchmark assessments offer more value because of their regularity. Nonetheless, the issue of generality also applied to these assessments.

Furthermore, at Bilby and Koala, most teachers responded that NAPLAN was not within their purview. Instead, it is a data source used more widely by executives. This corroborates findings from a recent pilot study on data use (Pierce & Chick, 2011). Furthermore, because Bilby and Koala, like Australian schools in general, were not held to any specific benchmark on NAPLAN (unlike their US counterparts), they tended to rely on the annual external assessment results solely for trend analysis. According to one of Koala's executive members, in a sentiment echoed by other administrators and teachers in both Australian schools,

A prime use of NAPLAN data is to inform, it is good big-picture stuff, inform big professional learning that is needed within school rather than just individual, for an individual child... we can say there are problems here now, we will do these. But that is just once a year.

Despite being less informative at the individual student level, the same Koala assistant principal appreciated the opportunity to observe and compare trends,

The SMART<sup>5</sup> data is brilliant in the fact that you can sort of follow the trends, not just within the school, but with the other co-schools. Before that, it was very difficult to sort of gauge how you were doing against the rest of the schools around the area.

Another Koala executive noted that NAPLAN provided the school with a reality check on how its students were performing:

The NAPLAN showed that we were performing where we should not and where most kids in the state were performing, so we were underperformers.

---

<sup>5</sup> The electronic database where schools can access detailed NAPLAN results of their students for analytics.

This knowledge gave the school justification to refocus,

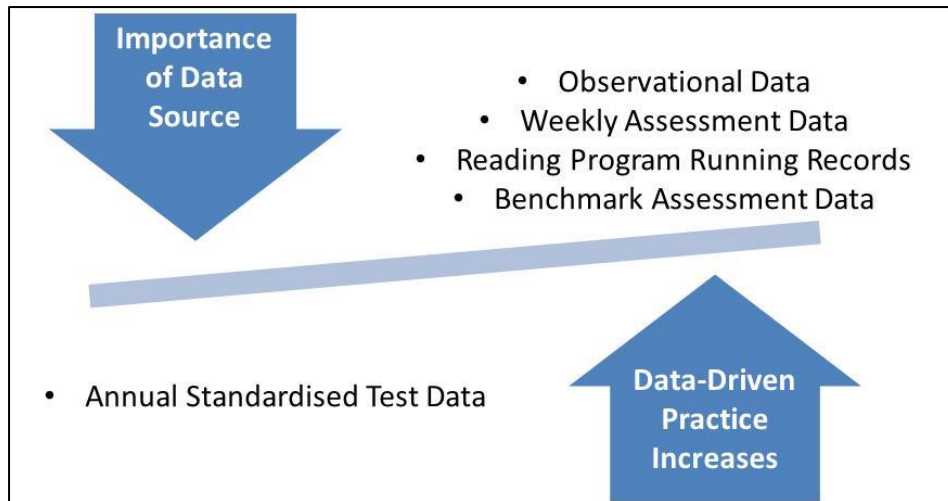
I think we took it and say well let us clear the decks, let us get rid of some things that are taking up our time so that we can concentrate on curriculum.

Regardless of value, Australian administrators and teachers generally agreed that the transparency offered through *My School* created unnecessary stress and thus reduced NAPLAN's perceived or real value. The assistant principal in Bilby's support unit expounded,

I do not think it is helpful in the way that, you know, it is advertised what certain schools get and all of that... I do not think is a positive at all, you know, our students are very low [performers] here, and it is good [for the school] to see that because then we try out new programs and we are constantly changing the way we are teaching.

Specifically, this assistant principal was concerned that the snapshot of student performance did not accurately provide context that the school had "a lot of learning needs", or reflected the school's efforts to support its students. A frustrated teacher in Year 5 at Bilby concurred, "That always upsets me a little bit when they [the public] do the comparisons". This finding is consistent with recent finding from a study (Thompson, 2013) surveying teacher perceptions on NAPLAN across two states and a pilot study on data use (Pierce & Chick, 2011).

For the most part, it appears that the more a school was immersed in data engagement as a way to guide programming and instructional decisions, the less it relied on external data, and the more it depended on internal assessment data. The result is an inverse relationship between the intensity of data engagement and the value of external assessment data as demonstrated in Figure 8.3.



*Figure 8.3.* The Relationship between Data Source and Depth of Data-Driven Practice

Based on these accounts, there is clear evidence that accountability pressure and accountability requirement were the catalysts for the initial data engagement in every case. However, as noted in prior research (Goldring & Berends, 2009; Young, 2006), it was the commitment of school leaders, and the provision of guidance and support in helping teachers see the connection between data use and student performance, that partly contributed to teachers' changing perceptions about data engagement.

I think that we've got to make, sometimes, that link for them [teachers], that okay yes we are collecting this from you but if we are not making clear enough that when we are programming together that we are using that information and they might be able to say 'that's a task'. (Deputy Principal at Bilby)

Once data practice is embraced, how did participating schools and educators use data to support students' academic growth? The next section details the numerous ways in which data inform strategic programming decisions and tactical curricular and instructional decisions.

**Nature of data use.**

This section begins by providing an overview of the uses articulated by the participants. It follows by sharing participants' descriptions of these uses and ends by discussing similarities and differences among schools and between countries. The nature of data use captured in Table 8.1 incorporates uses common among all the schools visited. At a glance, these uses do not differ significantly from those identified in previous research focusing on data use (Pierce & Chick, 2011; Wayman et al., 2007). These findings corroborate particularly closely the uses found at innovative schools (Supovitz & Klein, 2003). This is to be expected since the sample in the present study is non-random and biased towards schools that have delivered positive student outcomes as a result of data engagement.

Table 8.2

*Nature of Data Use*


---

Setting school, grade level and individual student goals
Informing student placement in regular classes or special services
Guiding curriculum and program development
Tailoring instruction to meet needs of individuals or small groups
Monitoring progress and evaluating curriculum and instructional effectiveness
Identifying successful and struggling teachers
Performing teacher evaluation
Reporting progress to the community at large
Allocating resources

---

**Set school, grade-level and individual student goals.**

All the executives in the sample mentioned using data to set school goals. These goals included raising student performance in spelling at Koala, moving more students to the upper band on NAPLAN at Bilby, and meeting the AYP targets at the

four US schools. As alluded to earlier, the executives at Bilby and Koala formed these goals based on the gaps that they observed in their own NAPLAN results, as well as comparisons with like-schools or the state average. The US sample's AYP targets were based mainly on their own trends from previous year's assessment results.

At the group leader level, which includes assistant principals and instructional coaches, data were used to set intermediary goals aimed at achieving the annual goals. However, because of the collaborative culture of these data meetings, teachers also took part in setting these goals. Their focus ranged from meeting a target in the benchmark assessment to specific skills at the end of each major unit or the end of an intervention program. Goals were adjusted on a three-week planning cycle at the two Australian schools and six-week cycle at the four US schools. A teacher in the Years 3 and 4 composite class at Bilby explained the dynamic nature of the goals,

And they readjust the goals for individual students. If the child has particular learning needs even though the benchmark might be set at level 20, we can adjust that goal.

At the individual teacher level, a third-grade teacher at Walnut said teachers created "mini goals" which they used as "checkpoints" to the quarterly goals. Teachers also developed individual student goals based on performance data and the needs of each student. The collective and individual goals were then communicated to each student and his or her parents.

**Inform student placement according to needs.**

All participants reported actively using data to inform student grouping decisions. For schools or teams that did not follow a co-teaching model, teachers

used data to decide how to group students so targeted instructions could be provided in the most efficient and effective manner. As a result of the availability of ongoing assessment data, student grouping was also a dynamic process, particularly at the US participating schools. A fourth-grade teacher at Walnut District articulated her grouping process as follows:

I go through [the tests], so '8' to '10' [assessment score of '8' or '10'] would be meeting standard, and then I say '7' and '8' [are] approaching. So I circle those because I want to meet with those students but not as often as the other ones. Then I highlight the ones that are '6' or below, and then I take that and I put them into my small group. So all the kids that are like '1s' and '2s', I would group together. And then all the kids that are '3s' and '4s' I would group together.

Teacher number 3 in Bilby's Year 3 and 4 composite team explained the significance of the dynamic process.

Just because you are in this maths group now, it does not mean that you will be next month because some children move faster than other children, so the groups are constantly changed according to the children's needs.

At schools or in teams that practice a co-teaching model, such as Hibiscus and selected grade levels at Walnut, teachers matched their personal expertise and knowledge to the particular needs of each student group as revealed through data. A Grade 3 Walnut teacher described in detail how they used data to match students with teachers' expertise,

At first we all kind of said 'okay I'll take that group, I'll take that group'... and we realized, wait a minute, we really should mind people's area of expertise. Like Nancy came from fifth grade, I came from fourth grade, so we really know what they need to know in order to prepare for the upper grades. So we have the two higher groups... and you know Jane came from first grade, so she really knows what those basic skills are, she's got a lot of strategies for working on those basic skills, so she took the lowest group. And



then Julie has been teaching third grade for a long time, so she is having the second grade combination class... she knows really well what they need to be able to do well in their third grade... so she's got that group that's the crucial group.

### **Guide curriculum, program and assessment development.**

Most of the teachers interviewed agreed that collecting data is “a lot of work”. Nonetheless, most also did not mind the work. As Teacher number 2 in Bilby's Year 3 and 4 composite class reasoned, “It is necessary in order to be able to do the next steps in our programming, to know where to go next”. For her, it was important to have the opportunity to incorporate “where [they] are falling down” based on what the leaders saw in data, into their programming.

At Koala, the teacher who created the special mathematics activity room used NAPLAN results to determine what activities to set up in the room. She focused on activities that could address challenges revealed by the data while supporting the needs of the largest number of students.

I usually look into the smart data like the NAPLAN and then I will say... ‘only two kids in our school got that question right and it is to do with measurement so that is the one [activity] I will use for that.’

The mathematics activity room was a strategic program informed by an overall lack of progress and engagement in mathematics that had been observed by the principal.

At Walnut, teachers created their own weekly assessments based on previous assessment results. The objective was to target areas that students still had not mastered, and to figure out “what's the breaking point for certain kids, like they can do an addition and subtraction problem but only when it's single digit”.

Understanding individual students' limits enabled the team to program their

interventions for the low performers, and to create extension programs for their high performers accordingly.

**Tailor support based on gaps in student performance.**

So this would be a weekly test that the kids would take and at the end of the test... I take [the results of ]my whole class and I look at it and say 'okay, four kids need reading comprehension', so I am going to pull them into small group and go over a main idea and details. 'Five of them didn't get the vocabulary,' so, I am going to pull them into my small group to work on vocabulary and... (Year 4 Teacher, Walnut District)

This type of process of item analysis to zero in on the exact areas that required intervention or reteaching was something that nearly all respondents mentioned as part of their data experience. This was also the exercise that the teaching teams engaged in during grade-level meetings. Almond's principal, the Australian assistant principals, and the US instructional coaches often also participated in helping teachers determine the topics and methods for reteaching. Indeed, it is at this level of data analysis where the value of data-driven approach materialised so that support for individual students' particular needs could be customised. The Walnut instructional coach shared an example of data helping a teaching team to make an accurate assessment of students' difficulties in mathematics:

What this data-driven process has really allowed the teachers to do is to see the root problem. Because throughout the meeting they had mentioned vocabulary and comprehension quite a few times as the challenge and this really helps to inform instruction [because] computation and arithmetic, they agreed, were not the problems.

In the past, when students missed a mathematic problem, the focus would have jumped straight to reteaching computation and arithmetic. Similarly,

behavioural problems had previously been thought to be child-related, but as Koala School and Kukui School had found, behavioural problems were linked to the curriculum, not the child. A Year 3 and 4 composite teacher at Bilby saw the same engagement benefit after adjusting their instructions to fix a vocabulary challenge observed on NAPLAN.

So then we can incorporate that into our programing by making sure that at the beginning [when] there is a text we are always pulling up the difficult words or vocab that they may need to know throughout the book or throughout their writings because it is all linked.

### **Monitor progress and measure curriculum and instructional effectiveness.**

Just as goal setting, data served as a tool for educators at all levels to monitor student progress. The executive used data to monitor school-wide program decisions. According to Bilby's deputy principal, "NAPLAN gives us sort of a ball park to look at, so that we can go 'okay comprehension is low'". Seeing a similar trend across the entire school, they could then determine the best course of action to ameliorate the problem. Once a solution was identified and implemented, staff returned "to monitor and track that data to see if [the students] are making improvements." During this process, data gives executives

The opportunity to have a chat with our staff saying 'okay, this works, this didn't work and kids are not getting this, what's happening in your classroom?' 'What are you doing to fix that?', and just having that bridge conversation about teaching and about, how the students are performing in their classes is really good.  
(Deputy Principal, Bilby)

Teachers used data to support their one-on-one conversations with students regarding performance. For the most part, when encouraging students to work

towards goals, teachers chose to focus on what students could do over what they could not do.

A big thing is that we make sure that we celebrate their successes, so that when they meet goals, you know that they know that it's a big deal, and so they strive to meet that goal and [to have] personal desire to do well. (Year 3 Teacher, Walnut)

### **Support teachers.**

Executives and instructional coaches who worked directly with teachers on a weekly or daily basis used data to provide instructional guidance. The deputy principal at Bilby believed the grade-level data meetings were an excellent forum in which to train teachers on the job:

Even if new teachers are there, and their input is not strong but they are sitting and listening, that's still an improvement in practice. They are still gaining some training and development from that and that's going to have an effect on student progress.

At Walnut, the instructional coach used the grade-level meeting as a place to collaborate with teachers and to help them drill deeper into the data to diagnose students' actual needs.

I will ask teachers to really look deep into what students are missing and what they mean to get in order to make that end goal and what kind of the target and instruction is best for them?

The coaches at Kukui did the same to help teachers identify trends and use them to formulate instructional strategies.

So it [grade-level meeting] is also a working hour as well. 'Do you see any trends? So what can we do, how can we address that?' Though they [the instructional coaches] have ideas, they wait for everybody to participate, 'what are we going to do then?'

**Report progress to the community at large.**

Participants who shared their thoughts on parental communication asserted that data offers a great way to start the dialogue. This is not merely because data-driven conversations are objective, but because they can be productive. Teachers used the dialogue as a way to engage parents to support their work. At Almond, the principal explained that assessment-outcome-related communication was a big deal and was done at multiple levels.

After each benchmark, I published all lists of the students who have hit proficiency and the students who have made significant growth so that growth being defined as 10%... Those are then sent home with monthly newsletter, and then... we celebrate that in our weekly community assembly. And then, we also do an individualised data letter to each family.

In one form or another, all schools communicated the goals that they committed themselves to with parents so that parents could be part of the process. However, as most of the schools in the sample served either a large English learner population or low-income families, language, cultural and educational barriers meant that the school's communication did not always succeed in bringing parents on board. Nonetheless, at Koala, the principal found that focusing parent discussions on data, as opposed to disciplinary complaints, not only reduced complaints but also freed the school from answering parenting complaints and enabled them to focus on schooling.

**Inform resource allocation decisions.**

All principals in the sample mentioned using data for resource allocation. Some decisions were related to programming, such as the mathematics activity room at Koala. At Hibiscus, data informed investment in a data system and in

classroom restructuring to a 'least restricted environment'. Data informed a number of professional development decisions such as training on spelling instruction at Koala after the school fared poorly across all grades in spelling in NAPLAN. At Walnut District, the researcher observed a training program for literacy based on the concept of 'universal access', which focused on strategies to raise the bar for every child and not only those falling behind. Lastly, staffing-related resources were another common area of data-driven decisions. The Almond principal described how this worked,

You see like this [referring to a data point she shared with the researcher] was actually a really low low low here for second graders. So I am giving more resources to second grade. We adjusted the instructional time for the second grade and all of those kind of things but we also expect to see a big movement this way and movement that way and that's generally is what happens then throughout the year.

At Kukui, after reviewing the school's data, the principal decided to shuffle teaching resources to help students build a strong reading foundation before they reached Year 3.

I wanted all my strongest teachers in lower elementary, so that my students are not behind in reading. For me it's important that everybody is reading at grade level or above. So I put my strongest ones in lower elementary.

From the narratives above, it is evident that data played a role in nearly every aspect of school operation at the participating schools. On the broad level, data helped schools set and communicate both accountability goals and working goals to all school constituents. On the operational level, data informed programing, curricular, staffing and professional development decisions. On the day-to-day level,

data enabled teachers to form efficient and effective groups to target students' individual needs as they worked towards school-wide goals.

## **Similarities and Differences**

In general, there were as many similarities as there were differences between the US and American samples. These extended beyond the nature of data use and also included attitudes about data driven-practice, and changes impacted by data and by accountability requirements.

### **Instructional leadership.**

At all participating schools, principals and assistant principals spoke at great length about being actively involved with teachers in making curricular decisions. Some also spent time in classrooms to help shape program development. They had intimate knowledge about students' needs and progress. These leaders exhibited many of the attributes associated with instructional leaders (Hallinger, 2005; Halverson et al., 2007) and transformational leaders (Leithwood & Jantzi, 2005). Leithwood and Jantzi's meta-analysis of 32 studies identified these attributes as: setting vision, group goals, and high performance expectations; modelling key values and practices; and building collaborative cultures and productive relations with parents and the community. These attributes were evident among leaders at the six participating schools. Leithwood and Jantzi also found a significant indirect effect of this form of leadership on student outcomes and engagement. Indeed, some researchers (Halverson et al., 2007; Mandinach & Jackson, 2012) had suggested that it was the advent of data that had ushered in this new generation of instructional school leaders.

In each of the cases in the present study, leaders were the driving force behind the removal of obstacles to ensure that data-driven practice had a better chance of sustaining in the long term. Almond Principal's active involvement in data diagnosis is an example of instructional leadership. In addition, she led by serving as a buffer for her teachers, taking the pressure from the higher education authority when their performance was poor, so that her teachers could focus on their work. Similarly, Koala's principal made the unpopular decision with parents and the community to cancel all non-academic activities so that teachers had time to focus on raising academic outcomes.

**Goal-oriented, accountable and transparent.**

In addition to having instructional leaders, schools across all three locations were highly goal-oriented. The schools' individual data, as well as comparative data from like-schools, helped them set believable and attainable goals for their students. More importantly, these goals were also publically communicated through their annual reports to parents and to the community at large. To that end, all schools have become much more accountable and transparent. Students, teachers, principals, and parents – everyone was aware of the school and the student's academic goals. In essence, everyone was accountable for delivering goals.

That affects the teaching and learning across the whole school because in the management plan those targets are delivered to the whole school and the strategies that are put in to then help us to gain those targets or achieve those targets go right through at K–6.  
(Deputy Principal at Bilby)

**Goal attainment – the currency for programming and instructional freedom.**

A large body of research (Chatterji, 2006; Duncan et al., 2007) demonstrates the predictive power of early literacy and numeracy on later school achievement. It



is by no accident that all schools in the sample focused relentlessly on literacy and mathematics, and on achieving their respective school goals for these domains. For three of these schools, an important incentive for reaching these goals was the opportunity to experiment with curricula and programs, and to try new instructional approaches. The Almond School principal described how success at reaching school goals had benefited her school.

At this point because we have strong test results, we are basically given full freedom [from higher education authority] compared to early years when test scores were little flatter.

With less monitoring from above, she said that her teachers had more time for the expeditionary learning program which linked literacy to social study projects to develop character growth, teamwork and reflection. The Walnut district data officer also confirmed that district officials generally allowed this type of instructional freedom when schools had met the benchmark.

In a slightly different circumstance, Koala also earned trust by demonstrating academic progress. The principal admitted that his conscious decision to get everything off the table and focused strictly on literacy and numeracy had alienated the community. However, just as in Almond and Walnut, that decision resulted in academic growth and behavioural improvement. These successes enabled the executive team to demonstrate to the community that, indeed, the school knew what it was doing and had strategies to succeed. In addition, academic progress gave the school legitimacy to set the terms of their engagement with the community. At the time of field work, the school had slowly begun to reintroduce non-academic-related programs, as both faculty and students had more capacity for other

activities. The principal also noticed an increase in parental engagement in the school.

**Not 100% on board.**

In all six cases, educators attested to the fact that not every teacher was on board with data-driven practice. In most cases, these teachers were identified by others as being older teachers and either comfortable with the way they had always taught, or not skilled in technology. Those teachers whom the Koala principal described as “crackerjacks” in technology, or who were technology savvy by their own admission, thrived in the new data environment. They tended to be younger and early-career teachers, as most executives in the participating schools noted.

For my younger teachers, they know this is the way they’re supposed to be working. They’re supposed to be focusing on the students and getting them to be better at meeting standards. For some of my older teachers are like, ‘Oh God, another thing [data analytics] I have to do’. (Principal at Kukui)

A new and young Year 5 teacher at Bilby validated that view:

For me where it [data driven practice] just becomes the standard, so it’s like I know only this.

The instructional coaches played a stronger supportive role in guiding teachers who were not yet fully on board through the process. Professional development was also part of the strategy. But, for the most part, administrators hinged their hope on the demonstrated benefits of data-driven practice to motivate them to come on board for the long haul. Demonstrating and modelling were the strategies engaged by many administrators to support resistant teachers. The data officer at Walnut District who began the process as a principal shared her experience supporting teachers.

It's been a journey and it's been a process. Everyone didn't just say 'wow, sure let's look at my data and see how my kids did.' It's been a lot of, you know, education and conversations and allaying of anxiety, and it's not about, you know, putting you down, but it's about, you know, what can we do to support the kids... so they feel that the time they spend in administering their assessment is not a loss of instructional time, but it's useful time because it gives them information to guide instruction.

Others chose not to provide an option, "We don't say 'do you want another data meeting?' We just say data meeting is next week", said an instructional coach at Hibiscus.

### **Strategic versus tactical orientation.**

Strategic over tactical use of data was probably the biggest difference observed across the two cultures in the sample. Perhaps because NAPLAN is truly a snapshot of school achievement (since data about the same student is available only every other year<sup>6</sup>), or because schools are not required to meet a particular state or national benchmark, administrators at the two Australian schools were just as interested in school trend and pattern identification as they were in grade-level and student-level performance. Trend identification led to many important school-wide and strategic decisions as discussed previously in the *Nature of data use* section. Grade-level meeting topics articulated by assistant principals across both schools focused as much on programming as they did on instructional effectiveness.

In comparison, in the US sample, meeting agendas focused almost entirely on individual student performance and strategies to help students meet goals.

Appendix I provides an example of a meeting agenda packed with topics related to meeting academic goals at one of the US schools. The entire year's agendas from the

---

<sup>6</sup> Students are tested in Year 3 and Year 5 so a student who takes NAPLAN in Year 3 will not take the test again until two years later.

same school did not deviate much from this narrow focus. Metrics and instructional or intervention tactics dominated at the US participating schools, whether in stories shared during the interviews, in grade-level meeting discussions, and at a school-level meeting with visiting district officials.

**Perceived versus actual accountability pressure.**

The derivatives of the word ‘compare’ were used often and exclusively among Australian participants and across all types of participant at both Australian schools. All mentioned the use of NAPLAN to compare their results against other schools, the state average or their peer group, to gauge where they fell short and what areas they needed to improve. This is how a Years 3 and 4 composite teacher at Bilby described the comparison exercise:

And often they [the executives] will show at the start of the year the NAPLAN result to say they might compare our school to the state average or something like that. So you can see the comparisons here we know that is the area that we need pull-up in so we know that we need to focus on that and include it [in our programming].

Ironically, this act of comparison alone created unprecedented pressure for the Australian participants, even though they were not technically under any state or federal pressure to deliver a certain level of performance (unlike their US counterparts). In the US, where API targets are mandated by law, participants used the term ‘compare’ only in the context of their school’s results from year to year despite the public availability and accessibility of comparison data. Just as *My School*, one can easily compare the results of School A and School B or C or D using the Hawaiian and Californian online tools.

Did the Australian media's use of data to create 'league tables' for schools influence the way educators perceived accountability at Bilby and Koala Schools? Has the fact that *My School* prominently presents results of one school's results alongside those of similar schools encouraged and legitimised the act of comparison? Conversely, at the four US schools, has the accountability requirement created inward-focused only schools? After all, whether a US school met its own target required under NCLB mattered more than how it performed relative to similar-schools or to the state average.

**Portfolio view versus strict assessment view.**

Between the two countries, Australian participants described a much more comprehensive formative assessment evaluation in their grade-level meetings than did participants in the four US schools. Observations and portfolio were described in the teacher handbook at Bilby and mentioned at both Australian schools as additional features to their assessment.

We encouraged the teachers as much as possible and as well to take any data records... I think you can learn a lot by observing a child just as much as you can learn by doing an actual pen and paper test. (Lead Teacher, Year 3 & 4 Composite Class at Bilby)

So we will do some formal testing, but more so it is ongoing anecdotal records, observations, student self-reflection, student goal-setting, looking at those learning continuums... We take video footage of children talking, so when they are talking about text ..."  
(Assistant Principal No. 2 at Koala)

Similar comments were absent in the US interviews, particularly not at Kukui or Almond, the two highly structured environments. Most decisions were informed by line item diagnosis of the ongoing exams. Although teachers at Walnut included qualitative data from the classroom to support their diagnoses of student needs

during the three grade-level meetings observed by the researcher, qualitative data was not formally collected in a student's portfolio, unlike at Bilby or Koala. Hibiscus teachers did not mention non-assessment-related data in their grade-level meeting where the researcher was present.

The experience observed and articulated by the US participants suggests that being tied to a stated benchmark might have hamstrung educators' ability to see the bigger picture. While the Australian participants clearly felt accountability pressure from comparative data, their funding and continued existence did not depend on meeting a particular proficiency target. This actual, versus perceived, pressure could be one reason why the two Australian schools could look beyond tactical challenges to strategic matters.

## **Conclusion**

This chapter has mapped the evolution of participating schools' decisions to systematically use student performance data to inform school operation and teachers' instructional practice for the purpose of raising student outcomes. While accountability policies were the impetus to focus on data in school operations, school dynamics played a larger role in each school's initial decision to incorporate data as an integral part of internal accountability. The respondents identified nine major examples of data-engagement activities to inform curricular, programming and instructional decisions. These examples are by no mean exhaustive but they were common across schools and geographical locations. They also corroborate uses identified in existing data use research (Pierce & Chick, 2011; Supovitz & Klein, 2003; Wayman et al., 2007).

Along with sharing their data activities, participants in the present study attributed performance growth in their schools to systematic data analysis and data use. To that end, they have come to embrace data not simply for accountability purposes but also for continuous improvement – the two overriding data use orientations documented by Jimerson and Wayman (2012). Participants' affirmative narratives about their experiences using data to support student learning is reassuring during a time when researchers generally agreed that effective data use was still a "vexing problem" (Wayman, Jimerson, et al., 2012), and many schools were still struggling with data-driven practice (Ronka, Geier, & Marciniak, 2010). While this research specifically sought out schools that were ahead in data engagement, given the negative publicity or framing in the media about educators' attitudes towards accountability and testing (Goldstein, 2011; Mockler, 2013; Shine, 2015), and the fact that not every teacher at the participating schools were on board, the rationale behind this particular sample's positive reception to data practice was puzzling. What influenced these educators' frame of mind and led them to commit to investing time and energy on a regular basis to use data as a primary tool to inform practice? Understanding these cognitive determinants could guide other schools hoping to implement data as part of the teaching inquiry process. It could also provide insights for policy makers as they contemplate how to effectively encourage schools to adopt data-driven practice. To answer these questions, the next chapter turns to the combined framework of efficacy belief and the theory of planned behaviour to interpret the findings.

## Chapter 9 Interpreting Behavioural Change

Literature exploring data-driven practice at schools in the current accountability era has applied organisational and leadership theories to explain patterns of practice (Booher-Jennings, 2005; Halverson et al., 2007; Jimerson & Wayman, 2012). The organisational view is important because current policy's focus on accountability has raised the bar in performance requirement and has demanded action from every member of the school. To meet accountability requirements, data-driven practice can no longer be a one-off activity, but must be a systemic and strategic practice at schools (Mandinach & Jackson, 2012). However, policy can be interpreted differently at every level, including the individual, grade and classroom level (Fulcher, 1989), and individuals have agentic capacity to decide the course of their own actions (Bandura, 1989). Thus it is as important to consider narratives and decisions about data use at the micro or individual level as it is at the organizational or policy level. Regardless of how data-driven practice was first introduced, ultimately, the exercise of control and agency to embrace and continue that practice depends on how educators, individually and collectively, interpret and process information to shape their decisions (Bandura, 1986b; Pajares, 1996). Unless teachers individually and collectively decide to embark on the behavioural change necessary to adopt data-driven practice, the prospect of sustaining such practice could be put in doubt even when the organisation itself has decided to embrace data.

From the perspectives of social behaviour (Bandura, 1986b) and social learning (Rotter, 1954), self-referent beliefs mediate between knowledge and action,



experience and action, and outcome expectancy and action. Sources of beliefs are both internal and external. Together, the theory of efficacy belief (Bandura, 1977, 1986a) and the theory of planned behaviour (Ajzen, 1991; Ajzen & Fishbein, 1980) provide a compelling framework to explain both internal and external antecedents influencing participating educators' data-driven practice-deliberating process. The combined construct suggests that the process of data-driven practice contributes positively to educators' development of personal and collective perceived efficacy beliefs, attitudes, subjective norms, and perceived behavioural control to commit to raising student outcome. In turn, these beliefs arm educators with resilience to stay committed to the practice in the face of internal and external challenges. The purpose of this chapter is to discuss how data-use contributes to the cognitive transformation influencing participating educators' willingness to commit to data engagement for continuous improvement.

### **The Theoretical Construct**

As will be recalled from Chapter 4, Bandura (2000) theorised that efficacy plays an important role in human functioning, both by having a direct impact on behavioural change and by indirectly affecting other determinants, such as attitude or subjective norm in relation to behavioural change (Ajzen, 1991). Efficacy judgment pertains to an individual's or a group's beliefs in their individual or collective capacity to affect a certain outcome (Bandura, 1977). As demonstrated in Figure 4.1, these judgments are moderated by the individual's or a group's mastery experience, vicarious experience, social persuasion and physical and emotional state (Bandura, 1986b). Over the past three decades, research has provided evidence connecting

student achievement to efficacy beliefs derived from students (Pajares, 1994, 1996), teachers (Hoy & Woolfolk, 1993; Tschannen-Moran et al., 1998) and the collective school perception (R. D. Goddard, 2002; R. D. Goddard et al., 2000). However, it is also important to recall that Bandura's theory does not concern itself with outcome expectations.

Given the present study investigates the impact of accountability pressure on academic performance outcomes, leaving outcome expectations and normative beliefs out of the equation in evaluating participants' belief systems would provide an incomplete analysis. Planned behaviour theory (Ajzen, 1991) offers a suitable framework to evaluate these additional important determinants, which are applicable in the respondents' narratives. TPB posits that *intention* and *perceived behavioural control* are immediate antecedents to behavioural change. TPB builds on the perception of ability. However, through the construct of perceived behavioural control, it also considers the perception of non-internal factors, such as the availability of requisite elements (e.g., financial resources, time, cooperation of others and so forth) (Ajzen, 1991). Intention is derived from an earlier construct called the theory of reasoned action (Ajzen & Fishbein, 1980), which suggests that behavioural decisions are influenced by both behavioural and normative beliefs. It accounts for a person's *attitude* about an action, based on his or her evaluation of potential outcomes, and *subjective norm*, the perceived social pressure to perform.

Together, these two constructs provide a useful framework to consider the cognitive beliefs underpinning respondents' commitment to using data for continuous improvement. The combined framework suggests that actions and outcomes associated with data-driven practice contribute to the development of

positive teacher efficacy and behavioural control judgment. They further moderate teachers' attitudes and subjective norm regarding data-driven practice. In turn, these beliefs affirm teachers' trust in the data-driven process. The ensuing discussion demonstrates how various data-use-related activities, and the resulting outcomes, played a role in affecting the antecedents of efficacy and behavioural control beliefs, as well as attitudes and subject norms, regarding the practice. The discussion considers evidence from the first-person accounts as well as from dialogue exchanges in meetings observed, through the lens of the antecedents underlying the two constructs; for example, mastery experience, vicarious experience, and others.

### **The Formation and Influence of Efficacy Belief, Intention and Behaviour Control**

Participating teachers alluded to the extra work involved and administrators to the resistance of some faculty members, in relation to data-driven practice. Yet, the practice had been embraced for the most part at the participating schools. How and why have teachers chosen to embrace a practice that did not begin from their own volition? Through the lens of efficacy belief and TPB, it can be inferred that many of the activities surrounding the data-driven process have enhanced teachers' perceptions of efficacy and control. Two important themes emerged from data-driven practice. The first involves educators' belief that they can influence students' achievement. The second relates to the belief that they have some control over the circumstances of their students' success even when an overwhelming majority of students come from a disadvantaged background. There was strong evidence throughout the interviews and dialogue exchanges in meetings that data-driven

practice provided a much-needed boost to educators' individual and collective perceived efficacy beliefs regarding their ability to support students who experience economical, physical, intellectual or linguistic challenges. The practice also appeared to have affected how far they were willing to stretch the goals for themselves and for their students, and the degree of effort to invest in achieving those goals. This boost seems to have been delivered through mastery experience, vicarious experience, subjective norm, positive persuasion and improved emotional states.

**Higher degree of perceived behavioural control and mastery experience.**

The perceived behavioural control and mastery experience constructs postulate that past performance results, as well as anticipated obstacles, predict future actions. Many aspects of data-driven practice have indeed afforded teachers the opportunity to raise their perceptions of mastery and of their control over personal struggles, as well as over those external to them. As a start, the joint goal-setting process based on data can raise teachers' perceptions of control and manageability by providing clarity. As a Grade 4 teacher in Walnut District suggested,

I definitely feel that I am better informed to know what I have to teach and it definitely guides what I do during the day and I know I need to focus on this otherwise – it holds me accountable because I actually have specifics that I need to teach and I think even from this year to last year, that I am better at knowing exactly what I need to teach, one from the data, because it tells me what I need to focus on.

This teacher considered it positive to know 'exactly' what needed to be taught. This is entirely understandable, because the concept of perceived behavioural control postulates that people have a more positive view when they know what to expect and can anticipate challenges regarding a goal.

Beyond clear goals, data further pinpointed precisely what students were missing and whether the teachers' own lessons and instructions were on target. The principal at Hibiscus felt that having that type of affirmation was, "the greatest thing that I know that is happening with the teachers". Furthermore, goals could also motivate teachers because, "people are pretty goal oriented and like to see results", as reported by the Walnut instructional coach.

The same rationale of positive perceived control development can also apply to students, who were also given clear goals to achieve. The instructional coach at Walnut contended that this process of the teacher communicating goals to students helped make the goals "seem more manageable" to the students, because teachers could break them into "manageable chunks". Furthermore, many participants also believed that approaching student learning with honesty and transparency, by sharing goals and progress, had raised students' desire to work harder.

Previously I wasn't as forthright maybe with the students in saying you know what you're reading at a 2.5 grade level and you're in fourth grade... I didn't want to affect their confidence... But I'm really honest with them [now] and they see their gains and they're willing to work. (Special Education Teacher at Hibiscus)

An instructional coach at Walnut summarised the importance of sharing goals with students in this way. First, it lets the student know that together "we got a lot of work to do". Secondly, it gives the opportunity for teachers to reassure the student that he or she will be supported, "let's see what we can do to – you know – make that goal". Thirdly, to allow the student to experience mastery of skills, "they get excited. They are like I may have not met the goal but I've made gains. And you're like yes, you did".

The relationship between goals and self-efficacy development has been documented (R. E. Wood & Locke, 1987), and the stronger a student judges his or her skills, the higher they set their goals (Zimmerman, Bandura, & Martinez-Pons, 1992). Together, self-efficacy and goal-setting had been found to predict 31–35% of academic outcomes (Zimmerman & Bandura, 1994; Zimmerman et al., 1992). The student experience described by respondents corroborates findings in a study (Friedel, Cortina, Turner, & Midgley, 2007) evaluating the impact of parental and teacher goals on students' disposition to adopt, and to come up with, strategies to overcome challenges to meet those goals.

Beyond having a more positive perception of their own efficacy, the transparency and the individual student goals were also reported to have raised students' self-esteem when they met those goals. Students could see that the process of learning and achieving is “a big cycle” and “they can tell you what it means, why it's important and how it's helping them in their education”, a special education teacher at Hibiscus explained. “Self-esteem” rises when students remember their experience of meeting their first goal, explained the data officer at Walnut District.

Additionally, educators' descriptions of the grade-level data meetings suggest that the structure and the process of these meetings can provide new teachers with timely opportunities to build their mastery. An assistant principal at Bilby described the impact of grade-level meetings on teachers new to the school,

People who are new to the grade or people who are new to teaching, you having [sic] that sort of environment where we plan together, when we talk about what we are doing together, is really supportive for members of staff from the grade.

Administrators in the sample believed that these data-driven grade-level meetings enabled new teachers to see in real time how data was being analysed and applied. According to the administrators, some of the benefits for teachers included having default mentors from day one and having a safe environment in which to learn and to ask “how do you do that?” In these meetings, new teachers learned how veteran teachers think and act in various situations. More importantly, new teachers did not feel exposed in sharing their challenges since sharing challenges is a feature of the grade-level data meeting process. In an empirical study on teacher efficacy, (Tschannen-Moran & Hoy, 2007), the importance of a support system was demonstrated to be of particular significance to novice teachers, who do not have mastery experiences from which to build their efficacy.

Even though grade-level data meetings were particularly beneficial to inexperienced teachers, seasoned teachers also “prefer the collaboration because the fact is that more heads are better than one”, according to Bilby’s deputy principal. Based on Goddard and Goddard’s (2001) efficacy research, one standard deviation increase in collective teacher efficacy can raise a teacher’s personal sense of efficacy by 0.25 standard deviation. This finding suggests that the grade-level data meetings, a part of the data-driven process, could have contributed to teachers’ personal sense of efficacy regarding their capacity to support student learning. Collective sense of efficacy and perceived behavioural control can be gleaned from a dialogue surrounding the weekly mathematics test results on a test designed and administered by the team of Year 3 teachers at Walnut.

Teacher No. 1: “The only two [students] that got a [score of] ‘2’ are Jane and Joe.”

Teacher No. 2: "Jane? That doesn't sound right."

Together, both teachers evaluated Jane's word problem and Teacher No. 2 noticed the following: "She makes all kinds of computation errors, she wrote '18-12', then she added them even though she set up the question correctly, that is weird."

Teacher No. 3 thought for a second and said, "Jane has been a bit distracted lately, I think her family life has been screwed, so [I] finally got them to bring her to school on time lately."

Teacher No. 1 continued to look through other problems done by Jane, and noticed two more careless mistakes which she read out to the group, then concluded, "She is not the kind of kid who makes computation errors, so if you see her (directing at Jane's homeroom teacher), talk to her about what's happening..."

Teacher No. 3, Jane's homeroom teacher also concurred that, "She's been doing fine on everything; she gets it all, that's the thing, why is this showing up like this?"

Then Teacher No. 4 checked the work, and noticed that Jane did the calculation right on one question but just wrote the answer wrong, so this teaching team concluded that they needed to follow up with Jane as their next step. They were concerned, but appeared neither nervous nor helpless. Their joint evaluation of Jane's work implied the existence of a collective belief that they all bore responsibility for Jane, regardless of which teacher Jane had been assigned to for mathematics or for homeroom. Furthermore, the level of their analyses of Jane's weekly test also attested to their collective in-depth knowledge of Jane, resulting from their decision to teach the entire third grade as a team by dividing up the subjects and changing grouping constantly.

The principal at Hibiscus aptly summed up the importance of efficacy as an antecedent to teachers' willingness to continue using data to influence student outcomes,



I don't have to go around and tell them [the teachers] they're doing a good job. And the more that I can get them to know it [through data] and develop that feeling of efficacy. Efficacy is everything. I mean that's why I come to work every day.

**Positive vicarious experience.**

According to efficacy theory (Bandura, 1997), vicarious experience can create and strengthen people's perceived efficacy. Vicarious experience can have a significant impact on teachers' perceived self-efficacy because "the impact of modelling on beliefs of personal efficacy is strongly influenced by perceived similarity to the models" (Bandura, 1995, p. 3). Furthermore, competent models can help to raise an observer's mastery belief by transmitting knowledge and teaching effective skills and strategies for managing challenges (Bandura, 1995). Observations from these collaborative grade-level data meetings suggest that these forums are conducive for the transfer and modelling of knowledge and skills, and in turn for teachers to vicariously imagine success through witnessing their peers succeed. An upper primary Bilby teacher described an experience she had when outstanding results were shared at a grade-level meeting by another teacher,

Oh you have got this kid to read at this level. What can we do?  
Can you show the rest of the school what you did to do that [to achieve the outstanding reading results].

In saying that, this teacher not only expressed her desire to achieve the same outcomes for her students, but wanted the teacher with good results to share her strategies.

Another instance where evidence of vicarious experience played a role in teachers' final buy-in to data-use can be inferred from the experience at Kukui when data-engagement was first introduced. The principal described the impact left on the

group when a seasoned teacher noticed her students' results using data during a meeting,

After the teacher kept saying, 'they can do it, they can do it, they can do it. We need to do it for our kids.' After she said that, and she was a well-respected teacher, then we just did it [committed to using data].

Students appear to be another source of vicarious experience that can raise teachers' belief in their capacity to support learning. Participants were asked whether they felt that they would have engaged with data if the push to engage with data were not required by policy. A Grade 4 teacher at Walnut District believed that motivation from her students' results would have eventually led them to the same place.

I think maybe eventually we would have fallen into this kind of thing, and definitely you know started gathering the data, you know because I think we are really seeing results at least in the kids and as far as their level of confidence, last year we know in our district schools, that there may not have a big change in our test scores, but the kids really felt good about their ability to problem solve.

Clearly, seeing her students' rising confidence boosted this teacher's belief in her own skills to support her students and motivated her to embrace data in her practice regardless of policy requirement.

In the same way that vicarious experience through the collaborative grade-level meetings could work in raising teacher efficacy, students could experience the same benefit through an inclusive learning environment. Kukui's principal described the benefits she saw when ELL students were put in mainstream classrooms,

As our students see what is expected of them and what other students who are at a higher level than they are, where they can see what their peers can do, then they raise their own expectations of themselves.

**An improved physiological and emotional state.**

Bandura (1986b) postulates that people's affective states are impacted by the level of stress, anxiety and excitement they face. The higher the stress level, the lower a person's judgment of efficacy (Bandura, 1995). Multiple participants recounted stories about stress and anxiety related to student behaviours and their impact on teaching. The principal at Koala recalled what he observed when he arrived, "teachers were a bit scared of the students and are [sic] disengaged with them". Based on Bandura's theory, this level of anxiety would negatively influenced teachers' perceived efficacy. Indeed, this has been proven in a study evaluating teacher instructional efficacy (Ho & Hau, 2004). The study found that student discipline and classroom management ability strongly correlated with personal teaching. In fact, this was the experience at Bilby. As highlighted in Chapter 8 *Case Context*, administrators at Bilby "were not setting a great example" for students and teachers for students because of the anxiety related to student behaviour.

Ironically, the drop in student behavioural challenge was one of the more significant outcomes of data-driven practice. Participants attributed the drop to their lessons being much more targeted towards each child's need. They asserted that their ability to customise instruction was directly related to their use of data to diagnose student needs. The drop in student disciplinary issues eased educators' anxiety and enabled them to focus their attention on schooling and teaching. The statement below from Special Resource Teacher No. 3 at Bilby highlights the impact on perceived efficacy when teachers experience less anxiety:

If we are teaching according to their needs, the children are more likely to be well behaved because we are catering for them, the

work is not too hard and not too easy, so it makes our jobs easier not hard.

Behavioural issues, however, were not the only source of stress for educators in the sample. With external accountability, and a relentless focus on school performance, particularly in the US, the literature is replete with educators experiencing stress. The structure and content of the grade-level data meetings, however, can offer teachers opportunities to temper their stress. For example, through data sharing in these meetings, teachers were able to put their students' poor performances in perspective. Teacher No. 2 at Hibiscus shared her feelings,

I felt so much better because, well everyone sort of you know didn't do so well on that. It wasn't just me... It [grade level data] kind of helped to validate that.

Another source of anxiety for most teachers in the sample revolved around the belief that their own students' outcomes were their responsibility alone; this was daunting for many teachers. This was true particularly for new teachers who were still developing mastery of their teaching skills, or for Years 3 and 5 Australian teachers, as their students are the ones tested by NAPLAN. However, it appears that administrators can leverage data to reduce teachers' anxiety by demonstrating how the result for every year level is in fact connected. The Year 4 and 5 composite class teaching team at Bilby described how knowledge of school-wide trends and school goals helped them to realise two facts. First, if students did not do well in the earlier years, the scores in Years 3 and 5 simply reflect the same trend. Secondly, they were not alone in the pursuit of respectable NAPLAN results. Teacher No. 3 in that team recalled what their executive shared,

It is not all about stressing out the Year 3 and Year 5 teachers because they have to do a huge amount of advising in that short period of time.

Instead, she assured the teachers that it was about everyone being responsible for both the “previous” and “future” goals of each child, so that by the time the child entered Year 3 or Year 5, he or she would have the necessary skills to do well on NAPLAN. Although accountability had raised the level of stress and tension at schools, data-sharing and data itself appear to have offered teachers a tool to moderate their stress, so as not to let it affect their efficacy judgment.

**Affirming social persuasion.**

Social persuasion is another source that can strengthen efficacy belief (Bandura, 1977). The shift from individual responsibility to grade-level responsibility in problem diagnosis, and collaborative lesson and instructional strategy development created unprecedented opportunities for phenomena such as social persuasions to surface. Through their work on data practice, Jimerson and Wayman (2012) “assert[ed] that data use is at the core a social venture” (p. 5). As such, during these ventures, teachers can confirm and support another teacher’s judgment, or reframe one another’s doubt to provide another perspective.

During field work, social persuasion was observed in the Grade 3 mathematics team meeting at Walnut. In one instance when a teacher was diagnosing the challenges of a child in her mathematics group, she faced the teacher who was the child's home room teacher and said, "your Johnny" or "your Mary", as a way of looking for affirmation or negation of her individual assessment of the child's strengths and weaknesses. In cases where more teachers have worked with the child being discussed, other teachers would chime in to affirm or to help with further

diagnoses. For example, when the team assessed whether to move a child who had been assigned to the second highest skilled group to the highest group because he had made good progress, they confirmed their assessment of his strengths with the teacher teaching the highest level to see whether he was ready for the move. The entire process resembled that of a group thinking out loud to figure out the appropriate level for this child.

Another powerful example of social persuasion was on display at the Walnut Kindergarten mathematics grade-level team meeting. When a teacher reported having had a few low-scoring students in the weekly assessment, she added,

I am so perplexed; I don't know if I just gave up, and didn't do what I normally do, I don't know what I've been doing.

A second teacher interrupted and said, "Well, you also have the 'stretch' group [the group that these teachers felt can handle harder mathematic problems]". A third teacher followed, "you have the hardest group". As a group, the first teacher's colleagues tried to persuade her that it was not her teaching ability, but perhaps the assessment that they collectively created was too "stretched" and hence confusing for the students. Through presenting a fact that their distressed colleague might have forgotten to consider, and by sharing responsibility for having created a possibly not-so-appropriate assessment, these teachers were effectively engaging in social persuasion, where someone is nudged into believing that they will be able to cope successfully despite a temporarily overwhelming situation (Bandura, 1977).

Social persuasion can be a particularly important source of efficacy development for special resource teachers who face many challenges. As the

principal at Hibiscus described, “I feel for special education teachers... The kids grow in such small increments; that is so hard for them to feel like they are doing a good job”. To help special resource teachers see their own value, he put them and special needs students into mainstream classrooms, so that these teachers could “see that they have a lot to offer in the classroom”. The principal shared this observation,

When they’re co-teaching together and the general education [teacher] is having a problem and ready to send the kid to the office, the special education teacher takes him [the child] aside and special education teacher can still deal with him. I found that the general education teachers have been saying ‘whoa, gosh, I was at my limit but you [special education teacher] taught me something’.

According to the principal, the special education teacher’s expertise was being recognised by the general education teacher. Such recognition could help to improve the special education teacher’s emotional state and in turn raise her perception of personal efficacy. Conversely, the fact that the general education teacher believed she had learned something from the special education teacher suggests that, through vicarious experience, she might have raised her perceived behavioural control regarding student disciplinary management.

In response to a question regarding how they, as teachers, felt when data revealed that their students were not doing so well, Teacher No. 1 in the Year 3 and 4 class at Bilby replied,

I do not feel like it is a reflection of, ‘oh I am doing a terrible job type thing.’ So I would never be embarrassed or ashamed to say ‘well that person in my class, and this is what they got type of thing. You want to keep trying to ask people, ‘well I have tried this, and that did not work, what can I do next?’

This teacher’s response illustrates the effect of social persuasion on her belief in her own capacity as well as her overall belief in the group’s efficacy. In seeking

support for the challenge she faced, she was engaging in what Bandura (2000) described as an exercise of proxy agency. What the collaborative environment had offered her and other teachers was a “socially mediated mode of agency, [where] people try to get other people who have expertise or wield influence and power to act on their behalf to get the outcomes they desire” (Bandura, 2000, p. 5).

### **Influences of attitude and subjective norms.**

The theory of reasoned action, before being expanded to the TPB, describes attitude and subjective norms as influences on behaviours mediated by intentions (Ajzen & Fishbein, 1980). The subjective value that an action could lead to the desired outcome under consideration serves to strengthen people’s intention to commit to that action. Normative beliefs refer to referent individuals’ or groups’ approval for the action under consideration. In the context of the present study, teachers’ positive response to data-driven practice as their *modus operandi* was clearly marked by their desire to continue to achieve the initial encouraging outcomes informed by data engagement. The initial outcomes both in student behavioural improvements, as well as academic growth, had a significant effect on teachers’ attitudes to the value of data-driven practice. The data officer at Walnut, who was once a principal, had anticipated attitude change in teachers regarding data practice, “they’ll [teachers] get there... they will take time [to work with data], when they see their [the data’s] value”.

The influence of subjective norms can be seen in the Kindergarten grade-level meeting at Walnut. Compared to the Grade 3 team at Walnut, this team had new members and was also new to data-driven practice. Their lack of practice with the process was evidenced through the way they strictly followed the meeting protocol



in their opening statement, in the way they took turns to talk, and in their general lack of ease during the meeting. The researcher noticed that the instructional coach provided much more feedback and reassurances than she did with the Grade 3 team, perhaps it was because of the members' tentativeness. The instructional coach also took on a different role in this meeting by participating more as one of the team members, probably to lighten the team's emotional state. At one point, the team members contemplated a strategy to use the special resource teacher and wondered whether this could work; the instructional coach affirmed that the strategy under contemplation had worked very well for another team. This affirmation, as well as her continued assurance, slowly changed the atmosphere of the meeting. Towards the second half, the team appeared slightly more relaxed and slightly more confident in recommending strategies. Other coaches and assistant principals spoke of similar strategies, which they use to guide, to support and to encourage teachers in data use with the goal of raising teachers' belief in their capacity to affect change in student outcomes.

## **Relevance and Implications of Strong Efficacy Belief and Planned Behaviour**

Considering the findings through the combined construct of efficacy and TPB helps to explain why this group of educators chose to embrace data-use beyond accountability compliance. According to leading teacher efficacy researchers (R. D. Goddard et al., 2004), the relationship between teacher collective efficacy and student outcomes depends on the reciprocation of a series of relationships including: teachers' personal efficacy, their professional practice, and their influence

over instructional decisions. The sections above have demonstrated that the structural design and the operational procedures of the data-driven process have provided appropriate settings and opportunities for shaping teachers' internal beliefs and for fostering these relationships. As these beliefs and relationships were evidenced in the selected cases in this research, it is possible to see that data-driven practice indirectly contributed to student outcomes by mediating the development of teachers' enhanced sense of individual and collective efficacy and behavioural control under the current accountability environment.

Beyond the obvious end value of student outcomes, how has the adoption of data-driven practice impacted teaching and learning? What evidence is there to suggest that data-driven practice is not a current policy fad but will become a sustained strategy at these schools? In the early years of NCLB, Hanushek and Raymond (2002) predicted that "focus on student outcomes will lead to behavioural changes by students, teachers, and schools to align with the performance goals of the system" (p. 81). Indeed, there were noticeable changes in teaching and learning at the participating schools. These changes emanated from the confluence of accountability, data availability, demand for transparency, and an enhanced internal belief system. Coming together, these events shifted the traditional model of schooling characterised by "loose coupling of administrative and teaching practice, teacher autonomy, individualized professional development and unmonitored instructional quality" (Halverson et al., 2005, p. 6) to one that is student-centric and collaborative at the selected schools. Data served as the hub in this change process. These changes have significant implications for disadvantaged students.

**Students are not the problem regardless of background.**

Across all schools, there is evidence that an increased sense of efficacy shifted participants' beliefs away from seeing performance problems as being inherent to students, and towards viewing performance problems as arising from curriculum and instructional misalignment. With comparative data, it was no longer an option to ignore glaring low-performance or to shift the blame to school and student background factors. This realisation was starkest at Koala, where executives were most reflective and realistic about what the data revealed.

We always blame, whether they [students] know [the test materials]... but the reality was there were other kids elsewhere at the same age doing much better and there is a reason for that and it is not only socioeconomic reasons that plays the part. (Assistant Principal, Koala)

Similarly, the principal recognised that the high rate of student suspension at Koala was not a result of their students being “out of control”. Instead, he believed it was inherent in the way they operated the school, because NAPLAN showed them that sister schools in the same neighbourhood, with similar populations produced very different results.

By shifting attitudes away from the idea that students cannot learn and towards a focus on curriculum and instructional practice, teachers became much more attuned to students' actual needs and the gaps in curriculum or in their own instruction. As Kukui's principal explained,

So the teacher has to change to fit the need of the student, and by looking at our data and following where our kids need the help, then that makes the teacher more responsive to the student's needs. You cannot just keep doing what you've been doing for last 20 years because that's how you've always done it, can't do that. You have to change.

This shift did not come naturally as there were still resistant teachers. However, once teachers moved past the initial feeling about change being a “requirement” by the school executives or education authority, and began to embrace change, a powerful effect on instructional practice and collaboration among teachers ensued.

An example is the interaction presented in the *Affirming Social Persuasion* section. In this example, a teacher felt discouraged about her students’ results, yet she saw them as her own shortcoming rather than the students’. The rest of the team members concurred that they had probably not created a suitable test, hence students were confused. More importantly, the team took collective responsibility for the failure, rather than allowing the individual teacher to do so. These teachers’ interaction resembles what Sachs describes as activist professionalism where “trust, obligation and solidarity work together in complementary ways” (Sachs, 2000, p. 81). Given recent findings (Nelson & Guerra, 2014) on the persistence of deficit beliefs among educators about diverse students at schools, the shift in attitude regarding students’ learning ability afforded by data at participating schools is reassuring, particularly in light of their work with disadvantaged students. As a group, minority students have typically expressed lower self-efficacy than their counterparts (Klassen, 2002, 2010). Educators’ changes in attitude should facilitate these students’ efficacy development.

#### **Student-centric practice.**

Throughout the interviews, participants with longer teaching careers recalled how they used to teach before the move towards data-driven practice. One such recollection was expressed by Kukui’s principal,

When I went through it [teacher training], assessment was what you did at the beginning before you planned your lesson, and then it was done at the end of the lesson... If they [the students] didn't get it, 'Oh, well. Next lesson.'

As participants at multiple schools suggested, there was a time when teaching was not about what students actually learned. Instead, it was about covering the required chapters in the text book in a given semester or a year.

When I first started teaching, it was sort of in the mentality that you had to do a program... and hand your program to your supervisor... And this is my program and I am going to get it taught and I am going to register that I have taught it all regardless. (Assistant Principal No. 2, Bilby)

This attitude continued when data was first collected for accountability purposes, because "there was no purpose for the data collection", reflected an instructional coach at Hibiscus. Teachers simply did not know what to do with the data, so they continued to do what they normally did regardless of whatever the data revealed.

The accountability era saw a slow migration from textbook-centric to student-centric teaching, and from whole-class lessons to targeted lessons. An instructional coach at Hibiscus noted, "Across the board, the school has become much more personalised per kid." This change was mentioned by all the schools in the sample.

There's a genuine belief, I think, everybody at the table believes that the kids come first in the model here. (Instructional Coach, Hibiscus)

I think teachers are having more conversations about where their students are and how they are progressing. (Instructional Coach, Walnut)

When the shift to student-centric practice spanned across faculty, staff, administrators and care-givers, a series of benefits for all constituents ensued. These benefits will be discussed in greater detail in the *Subsequent Direct and Indirect Benefits* section below.

**Collaboration over isolation.**

I was a teacher... It was very isolating and unless you go to talk to someone, all you know is what you do. (Instructional Coach, Walnut)

I mean I think back to the days where I start[ed] a program on my own and it [was] time consuming and it [was] painful... Everyone they [were] doing their little pie. (Deputy Principal, Bilby)

This isolation phenomenon has been well documented in the literature (Cuban, 1990; Hargreaves, 2000; S. M. Johnson, 1990). Along with the shift to student-centric practice, data has also transformed teaching from what this Walnut instructional coach considered “an isolated profession”, to a collaborative one as teachers collectively make sense of and make use of data to support students.

Collaboration and collegiality were visible in the grade-level meetings observed. In diagnosing and discussing students who were border-line between two skill groups, the Grade 3 Walnut team’s analyses expanded beyond a struggling student’s performance on the latest assessment to other factors that could have affected that student’s performance. For example, the environment in which the student thrived, his physical needs (the need for glasses) that might have contributed to slow test completion, and other conditions. Every member of the team shared knowledge about this particular student.

This team went a step further in their collaboration. They leveraged their strengths to teach students in areas in which they had expertise. The teacher with

lower grade experience took all the students who were behind, and the teacher with Year 5 experience took all the high performers, since they each knew what came before and after to provide the exact support needed by these two groups of students. In addition, they decided to divide up other subjects, such as science, art and social studies, to lessen the time needed to plan for every subject and to free up time to focus on students. These collaborative decisions are representative of what a Bilby teacher referred to as “catering” and “teaching according to their [students] needs.” In her opinion, when teachers work this way, children are engaged and behavioural challenges are reduced. The data officer at Walnut confirmed that many principals in her district had reported fewer disciplinary issues because of student engagement. It is therefore not impossible to imagine this Grade 3 team succeeding in meeting their goals for students, as research (Y. L. Goddard et al., 2007; McLaughlin & Talbert, 1993) has linked this type of collaborative commitment and focus on continuous improvement to positive impact on student learning.

For some school leaders, including Bilby’s deputy principal, data made it very clear that collaboration was the only path forward. If they were to have any success at raising student results, everyone had to be responsible for outcomes.

I mean [data] forced us as a school to look at the fact that it’s not just up to the individual teacher to process data, or whatever it is, that it has to be a whole school thing... So as a supervisor I can’t just say, oh I don’t know anything about the students or the teachers that I supervise or the students, I really got to be accountable for the learning of those students.

This line of thinking reflects Bandura’s theory that “many of the outcomes they [people] seek are achievable only through interdependent efforts. Hence, they have

to work together to secure what they cannot accomplish on their own” (Bandura, 2000, p. 75).

The deputy principal’s message had definitely funnelled down to faculty level. In a separate interview, the team of teachers for the Year 3 and 4 composite classes spoke about how the holistic view of data at their school had enabled them to see, first and foremost, that their students were not the only ones that did poorly, and secondly, to realise for the first time that they have shared responsibility for all students.

I have to tell myself that the kid who has been at the school for three years and still not doing well this year in my class can’t just be me... That child has been with us since Kindergarten... I had to really stop and say... ‘it is not my responsibility to make sure the kids are performing well in the NAPLAN, it is all of our responsibilities.’  
(Teacher No. 3, Year 3 & 4 Composite Class, Bilby)

Similar feelings were echoed elsewhere,

It’s really broken the barriers of the classroom walls, so it is not my 25 kids, it’s our 130 kids in second grade that need to get to third-grade level. (Data Officer, Walnut School District)

As schools in the sample shifted from isolating to collaborative data engagement, they noticed positive outcomes. Their success is not surprising as collaborative behaviours and collaborative processes have been linked to positive outcomes for disadvantaged students in a meta-analysis (B. M. Taylor, Pressley, & Pearson, 2000). The schools participating in the current study exhibited characteristics found in the meta-analysis: collaborative community where responsibilities for students are shared by both staff and faculty; progress monitoring being a core instructional practice; mutual learning and teaching support among faculty members to improve the art of teaching; and family outreach.



Collaborative models work because, in a socially constructed environment such as a school, outcomes “are achievable only through interdependent efforts” (Bandura, 2000, p. 75). These efforts only bear fruit when teachers and administrators have shared beliefs in their collective power to raise student outcomes – the main feature of collective efficacy (Bandura, 2000).

**Subsequent direct and indirect benefits.**

The fundamental changes described above in schooling and teaching, in turn, ushered in important direct and indirect benefits articulated by many constituents.

***Constructive and objective teacher support, development and evaluation.***

Several participants concluded that a key benefit of student-centric data-driven practice is the shift from social and administrative conversations at faculty meetings to constructive conversations about students. In the old model, the principal at Kukui said, “They just did whatever they felt like, discussing the field trips, what was happening at home.” “Nuts and bolts”, “issues” and “disciplinary problems” were widely mentioned as topics of conversations at every school. As Koala’s principal suggested, turning to student outcomes enabled schools to focus on the critical responsibility of schooling.

Student-centric practice via performance data also provided an effective and objective way for administrators to evaluate teacher’s work as articulated in the quote below. Teachers could no longer use anecdotal reasons to explain poor student results. Almond’s principal recalled a time when data had enabled her to reject casual explanations from a teacher and to provide constructive feedback after noticing gaps in his students’ performance.

Why didn't they [the students] learn this? This tells you a lot about the teacher... He says 'they didn't practise enough or they didn't get the question right, they didn't look at the meaning in the context of big picture.' At that point, it's really important to have the test with you, so you are not just looking at the test [results], but you are going back and looking at the actual questions... You can't tell here [the aggregate score] why didn't they get it right. You have to look at the question and the teacher needs to dig deeper to know why.

As the deputy principal at Bilby offered, having data to “bridge conversation” about teaching and student performance was not only valuable but less threatening. For some administrators such as Hibiscus' principal, student data also took away the “uncomfortable conversations” of having to counsel teachers without proof, and the risk of sounding “judgmental”. For him, student-centric data was not only a fact-based tool to approach teacher evaluation, but also a means to correct his own perception of teachers. He shared an evaluation experience where he was humbled by what the data revealed against his own judgment.

In at least two cases there were two teachers on my campus that I really did not have a high regard for professionally... in those two cases having data to see that the students are actually learning really made me kind of go, 'whoa, don't judge a book by its cover.'

Across the board, administrators were finding it easier to have conversations about students with teachers because the focus was on students, even though these conversations were indirectly about teachers' instruction. As the data officer at Walnut explained, it is not about “being evaluative” but about “how to make ourselves better” and “grow” from what is revealed in the data. At the same time, data also facilitated support and assurances for teachers who were less efficacious as in the following example:

When there's a teacher who's struggling like this teacher here... But the ability that the kids came in with, this is a much more needy group... And this whole year she's been struggling... It's affirming to

the teacher [to say to her] ‘you know what, you’re not crazy. This is a difficult cohort’. And maybe as a school we’re learning too that we could have done things differently from the beginning of the year.  
(Principal, Hibiscus)

Even teachers appreciated what student-centric data could reveal about their instruction because it helped them to improve as a teacher. A Grade 4 teacher at Walnut District reflected,

My teacher judgment isn’t always accurate... so it’s nice to have data... when I am, you know totally wrong, and it’s nice to not have the students be stuck in a location just because we thought ‘oh these students, this is what they are struggling’ and then it turns out that they are not.

***Rising student morale and self-esteem.***

The similar type of objectivity in supporting teaching also funnelled down to supporting student learning. The data officer at Walnut District suggested that student-centric data practice took away the tendency to generalise student abilities based on their background. As teachers became more aware of students’ needs, she noticed that they no longer said “this is what Hispanics need, but this is what struggling readers need.” Just as the administrators supported them, many teachers also conducted honest and open conversations with students about goals and progress.

With the advent of data, one of the big differences in my teaching practice, I think is the transparency with the students... I don't think I've ever done that before. This is my 16<sup>th</sup> year of teaching. So that for me was a big shift... I'm really honest with them [students] and they see their gains and they're willing to work. (Year 4 Teacher No. 1, Hibiscus)

The power of student-centric outlook lies with empowering students at “all parts of the spectrum and ability to shine and be successful”, including students in

special unit classes, said a special unit class teacher at Hibiscus. Enabling students to experience success can have a long-term positive impact on outcome, as prior success adds to mastery experience, a source of efficacy building (Bandura, 1977). In every investigation on student efficacy reviewed in a meta-analysis (Usher & Pajares, 2008), mastery experience significantly correlates with and consistently predicts self-efficacy. In goal-oriented and data-focused schools such as those in the sample, students can follow their progress and have more insights into their performance; therefore, a higher chance of experiencing the 'excitement' which participating teachers observed in their students.

In addition to the desire to work harder, participants also noted that students channelled their negative energy to positive activities. As discussed previously, across all schools, participants shared stories about the reduction in student behavioural issues and they attributed these reductions to the targeted lessons based on data. This is not surprising given the strong relationship among sources of student efficacy found in a meta-analysis (Usher & Pajares, 2008). The authors found that a student who achieves benchmark goals (mastery experience) receives recognition (subjective norm) and experiences positive feelings (affective state) is more likely to approach learning with a good attitude.

### ***Teacher professionalism.***

In the interviews among administrators, references to data-driven practice's impact on teacher professionalism were commonly heard. School administrators noted a move from congeniality to collegiality in teachers' day-to-day practice and

their relationships with one another. Throughout the interviews, 'collegiality' was referred to often among teachers and administrators.

Building the collegiality, I think that's one of the benefits too when looking at the data. (Teacher No. 2, Year 4, Hibiscus)

It [Data] forces you to have those conversations and learn from each other. (Instructional Coach, Walnut)

What did they mean when they used the term 'collegiality'? Reflecting on the words of the data officer at Walnut District, the 'collegiality' that teachers spoke of may in fact be 'teacher professionalism'.

I think the collegiality between teachers has been really strengthened, because even though in the past they may have been, you know very congenial and they like each other, they had things to talk about, their meetings were largely very often focused on nuts and bolts... so this has really taken it to a much more professional level of conversation where you are sharing best practices, but not just random 'here's what I did and the kids loved it', but 'here's what I did, and here are the results I got', so 'wow, they are great strategies', or 'let's try this, I've heard this really works', and then let's look at, you know what the data tells us.

It can be inferred from this explanation that 'collegiality' goes beyond being friendly with one another; rather, it refers to constructive cooperation, mutual support and intentionality. This fits well with the definition of professionalism as an occupational value discussed in the education literature (Evetts, 2008; Hargreaves, 2000), where the profession revolves around competence, trust, occupational identity, and cooperation.

Participants contended that discussions about student achievement were no longer about laying blame but about how to work together to help students. For this reason, teachers found the grade-level data meetings a "safe environment" to raise

their concerns, as well as a forum in which to share practice. The instructional coach at Walnut explained,

I think the benefit is that it allows a time for teachers to sit down and collaborate and learn from each other in a really safe environment where they can feel comfortable saying 'oh my gosh, my class did not do very well, what to do?'

These findings align with conclusions from research (Schnellert, Butler, & Higginson, 2008) linking teacher collaboration through data-inquiry cycles on professional development processes. The authors suggested that innovation in teaching afforded by data engagement is one of the reasons for a more professional environment. Hargreaves (2000) believes that this type of collaborative professionalism can bring many benefits including the opportunity to:

Develop common purpose, to cope with uncertainty and complexity, to respond effectively to rapid change and reform, to create a climate which values risk-taking and continuous improve, to develop stronger senses of teacher efficacy, and to create ongoing professional learning cultures for teachers that replace patterns of staff development, which are individualized, episodic and weakly connected to the priorities of the school (p. 165-166)

Based on participants' accounts in the current study, these benefits, while not yet universal, were evidenced at the participating schools.

It has been found in the literature that providing teachers with enough time to collaborate is an area in which many schools attempting to implement data practice have failed (Wayman & Stringfield, 2006b; Young & Kim, 2010). The schools in this study were cognisant of the critical element of time, and worked hard to prevent it from inhibiting the data process. For example, at Walnut, the school decided to use part of their funding to hire substitutes for the grad-level teams so they could meet to cover reading and numeracy data during school hours. This was a

strategic decision by the school to encourage grade-level meetings because “some teachers leave early right after school when the contract hours are done”, said the instructional coach. She felt that forcing these teachers to stay late for data-meetings would have turned these meetings into a chore and led to resistance. At Kukui, the principal scheduled specialist classes such as art, physical education and computing during the teacher articulation day also so that classroom teachers could meet during school hours.

***Culture of celebration and high expectations.***

Collaboration and student-centric practice also turned a culture of blame to one that focused on celebration of success.

A big thing is that we make sure that we celebrate their successes, so that when they meet goals... you know that they know that it's a big deal, and so they strive to meet that goal, and [increase] personal desire to do well. (Instructional Coach, Walnut)

Administrators across all the schools also explained that parents received congratulatory letters when their children met their goals. Student successes were also celebrated at assemblies. Around schools, posters and congratulatory flyers were seen in hallways particularly at the more structured schools, Kukui and Almond. At Koala, parents were invited to mathematics activity exploration evenings to learn about hands-on learning.

Bilby also stated that students in the support unit were put on the same growth model as opposed to being exempt from it. This meant that students in the special unit were expected to grow at the same pace as the rest of the school, albeit from their own baseline. At Walnut, the researcher was invited to sit in on an official district visit where grade-level teams presented their results. The meeting ended not

only with congratulations from district officials but with a champagne celebration which the Walnut principal initiated to acknowledge grade-level successes. In sum, the culture at most of the participating schools was positive and encouraging despite the many challenges they faced given the disadvantaged populations they served.

### **Costs of Data Engagement**

Benefits emanating from data-driven practice, however, also came at a cost.

The following subsections discuss the cost derived from data use.

#### **Constant testing.**

There is strong evidence that data needed for data-driven practice came at the expense of other programs and activities. To have data, weekly and monthly tests became a mainstay of the instructional process. This was more accentuated at the four US schools than at the Australian schools.

The State of California, we are getting very, very data driven, which you know is good in many ways, there are drawbacks because sometimes I feel that all we are doing is testing the kids, but I think that our particular lessons that we've chosen are meaningful, and do give us information so I think the data is really helpful. (Year 3 Teacher, Walnut)

It's just the reality that what gets assessed, gets paid attention to. So it's you know, it is definitely on the front burner of our conversations, and at least so at the three principal meetings that we have with cabinet and every data collection cycle so that's at least 7–8 times a year, that's a focused conversation. (Data Officer, Walnut District)

At these schools, teachers were busy preparing students for benchmark assessments and for the data-review cycles alluded to by the Walnut District Officer. It is hard to imagine excess time for non-literary and numeracy related activities.



**Seeing the forest for the trees.**

A side effect of data-driven practice appears to be the narrowing of school and instructional strategies. When schools such as Almond and Kukui put teachers within constricted curriculum pacing as a strategy to meet goals, they risk losing strategic vision for the school. From the amount of coloured charts and graphs of assessment results displayed throughout the principals' offices and in the school hallways, and the detailed evaluations of student performance as seen in Appendices J and K, one gets the impression that meeting assessment goals is the single most important work of these schools. Even an instructional coach whose job it is to guide teachers through data warned of the danger of solely focusing on data without looking at the big picture.

Data is strictly numbers, so sometimes I think you do have to take a step back from the numbers sometimes, and look at the whole child and see where this child started from. (Instructional Coach, Walnut)

At Hibiscus, the principal's incremental approach in providing "narrow achievement targets" to teachers, so they "know where to aim" risks delivering only skills and not conceptual understanding or knowledge to students. This narrow approach is what Entwistle and Smith (2002) described as target understanding where understanding is derived from the formal requirements of the syllabus. These authors argue that a more desirable form of understanding is personal understanding, where students are able to make sense of what they do. Evidence of the risk can be deduced from the annual school reports. At the four US Schools, these reports were devoid of any activities or programs not related to student performance. In contrast, the two Australian school annual reports were much more comprehensive on covering the holistic life of the school. The difference, however,

could be a result of Australia having a national curriculum, which covers a broad range of skills. In contrast, the 100% proficiency requirement in the US, while aspirational, might have constricted the breadth of knowledge and skills.

### **Compromising creativity.**

While systematic data-driven practice has contributed to improving teaching skills, it has the potential to run the risk of stifling creativity and creating a cohort of teachers who only rely on data for their lessons, and who forget about other creative resources and ideas. Indeed, education researchers (Schoen & Fusarelli, 2008) have warned that accountability and 21<sup>st</sup>-century school requirements, which include critical-thinking skills, adaptability, and creativity, are incompatible. This incompatibility is alluded to in the following comments among participants of the present study.

So I think that sometimes the teachers may feel a little, depending on the teacher, restricted to a narrow band, or they might want to work on character development because that's important. Or they might have a passion about a type of reading skill that they'd like everybody to have you know. (Principal, Hibiscus)

I think that you know you feel so like... 'oh my gosh we have to have this perfect test scores or else'... you know they put the fear of God in you as well these test scores, and like if you don't get those test scores, your school is all these bad things are going to happen [sic]... You can get so wrapped up and stressed out that you forget that, we need to have space to try these things out, you know that eventually yes this would affect their test scores... So sometimes we have to like remind ourselves that's it's, you know, it's okay to try those things out, and it could be a disastrous experiment, or it could be an experiment that has you know amazing benefits. (Grade 3 Teacher, Walnut)

From the students' perspective, while they might feel confident knowing that they could count and read, they might not be able to find innovative solutions for problems because they had little time for creative outlets. At Kukui, "well, we did

have a kid come and say they were being tutored too much”, said the principal. The student complained that there was “no time to play”, said Kukui’s instructional coach because “too many people pulling that student to do *response to intervention*<sup>7</sup>” continued the principal. This situation generally worsened as HSA testing approached, “so now that we’re in the last stretch before our last HSA, they’re getting triple and quadruple doses of response to intervention”, the principal confirmed. In addition to after-school tutoring, many low-performing students at Kukui were also required to attend a summer school aimed at helping them to get closer to their goals. Again, this risk appears more acute in the participating US schools. It is unclear whether Bilby teachers are exposed to the same risks, but at Koala, their experiment with the mathematics activity room leaned on the creativity spectrum.

**Limited visibility of students not at risk of failing.**

Ironically, the light shone on disadvantaged students has left shadows over exceptional or gifted learners. Two principals reflected on the imbalance in their curriculum and teaching focus.

I feel that, that’s a class of students [gifted] that we may not be serving as well as we could... And then our middle kids I still want to see the increased rigor... Although we’re making the state and benchmark test and we’re doing what is expected of us. I think our rigor could be increased. So that’s a long-range thing for me. (Principal, Hibiscus)

At Koala, it had dawned on the principal that they “were always focusing on the remediation” even though it is a fact that “there are gifted and talented kids in every school.” So to correct that bias, he instituted the following:

---

<sup>7</sup> *Response to Intervention* is a multi-tier early identification and support system for students with special needs in the US.

Extension groups for gifted and talented pupils or whatever, and we did gradually chipping away at that, you know, that belief that in low-SES school that you would not have kids who are gifted and talented.

At Almond, when asked directly whether the students who have met benchmarks were able to continue to grow, after some silence, the Principal offered the following:

There is an expectation that every child gets more reading instruction that will be either guided reading in the early grades, then moving towards more of book club or literature circle in the upper grades, and that's a place where it is really going, and what the student actually reading level is and they are now pushed towards higher levels.

Any educator would agree that a book club or literature circle is by no means a comprehensive extension strategy. Other administrators not yet mentioned here offered similarly unsatisfactory answers, which amounted to 'we are always meeting the students where they are'.

The exception seemed to be Bilby School. As discussed earlier, that school's goal aimed at moving more students to the above-average proficiency bands. Clearly, this school focused on stretching a large group of students. However, in this case, the school had the benefit of having fewer low-performers, which enabling them to devote attention elsewhere. In contrast, most other schools had a much larger population of students who had not yet met proficiency standards.

## **Conclusion**

In *Guiding School Improvement with Action Research*, Sagor (2000) outlines a seven-step process for improving teaching and learning in the classrooms. Data is featured prominently in this process. Drawing parallel to the way athletic teams use

data to improve their games and professionals on the Apollo Mission used data to solve problems, Sagor argues that, the “availability of data on performance and teacher authority to use the data to improve their instruction—are the prerequisites for building efficacy” (p.35). He further asserts that, data “is means to renew the efficacy that most teachers possessed when they left college, believing they could accomplish miracles” (p.35). In seeking to uncover the rationale behind teachers’ decisions to constructively engage with data to support student learning, this chapter found evidence to support data-driven practice’s influence on efficacy and demonstrated how efficacy development could be supported, through data and data-driven practice.

The combined theoretical framework of perceived efficacy and theory of planned behaviour suggests that the procedural operations of systematic data engagement offer opportunities where teachers’ internal belief systems could be influenced, including efficacy and control belief, attitude and intentions. Sources influencing the positive development of teacher beliefs, such as vicarious experience, social persuasion, mastery experience and affective state, could be identified in multiple aspects of data-driven practice. Whether sharing best practice, gaining a more holistic perspective of how different grades perform, having a clear direction and plan of action to reach goals, witnessing success achieved by peers, or experiencing success of their own, teachers had many opportunities through the data-driven process to alter their attitude and raise their perceived efficacy and control regarding their capacity to support the needs of the disadvantaged students they serve. Indeed, the educators’ narratives suggest that it was their elevated perceived efficacy, as opposed to the policy mandate, that led them to commit to

what data researchers (Mandinach & Jackson, 2012) considered a “philosophical and holistic transformation toward continuous improvement” (p. 19).

However, the participants’ accounts also highlighted some challenges of data-driven practice including putting creativity, flexibility, strategic vision and the joy of learning at risk. Despite that, they were fully committed to this change because the process had boosted their confidence in working with a large disadvantaged student population whose academic performance still trails advantaged peers by a wide margin. Although the current environment for practising systematic data-driven practice is not ideal, owing to accountability pressure, it has brought about significant shifts in the way in which these schools assumed responsibility for student learning and regard student abilities. It has also broken down walls between classrooms. These changes resulted in a much more professional faculty and more confident student body, and armed them with resilience in the face of challenges in today’s accountability environment.

## Chapter 10 Discussion and Conclusions

### Context

This study began with three overriding objectives, which it accomplished through mixed methods research. First, it aimed to evaluate the current trends in primary student outcomes on external assessments in Australia and two counties in California and to examine differences between advantaged and disadvantaged students (California) and schools (Australia). Secondly, it explored the relationship between data engagement and student outcomes by investigating the underlying beliefs that led teachers and administrators to embrace data engagement as a means to support student achievement at six schools in NSW, CA and Hawaii. Thirdly, it compared and contrasted the impact of the Australian and US accountability policies on learning and teaching. In the current era of external and internal accountability, where government policies and funding are based on data (Goldring & Berends, 2009; Hess & Finn Jr, 2007; Young & Kim, 2010), findings from this study build on and contribute additional knowledge to existing school reform literature in three important ways.

### Significance of the Study

First, the comparative nature of this study contributes a broader understanding of accountability reforms and their impacts in different educational systems “by providing a background of contrasts against which to examine [a particular system’s] problems” (Grant, 2000, p. 315). Secondly, this study adds to the nascent research in Australia evaluating the impact of accountability and external

testing on the actual outcomes of disadvantaged students and data-driven practice at schools. Thirdly, by focusing on educators' underlying belief systems to explain the rationale for their commitment to data engagement for internal accountability, this study contributes a different perspective to the growing body of research on data-driven practice. Whereas existing research focused mainly on the nature of data use and the obstacles and facilitators of data use, or explained the motivation for data-practice from an organisational perspective, the present study evaluated the antecedents to data-engagement decisions at the individual and collective teacher level. This is important, for policy can only go so far in making data-driven practice a requirement. In the tradition of cognitive behavioural theory, any account of behavioural change must consider self-generated influences as a factor because "people strive to exercise control over events that affect their lives. By exerting influence in spheres over which they can command some control, they are better able to realize desired futures and to forestall undesired ones" (Bandura, 1995, p. 1). Therefore, understanding the mechanisms that account for educators' willingness to change is vital to the proliferation of data use.

## **Key Findings**

### **Quantitative findings.**

The quantitative results from both countries provided evidence that most disadvantaged student groups have made real progress on external assessment in both scale scores and in proficiency rates between 2008 and 2013. The California results suggest that inclusion of student subgroups in external assessment is less of an issue today than it was when NCLB was launched. On the whole, the student



subgroups in California that achieved the largest gains in both subject domains were disadvantaged student subgroups, such as Hispanic students and students with disability. However, English language learners, who already had the lowest proficiency rate among all subgroups, were the only student group whose achievement declined between the two periods. While overall increase of up to 20% in proficiency rates for most other student groups is encouraging, it has not changed the achievement gap between advantaged and disadvantaged students. In 2013, the size of reading proficiency gaps between ELL and English-only students reached up to 50% in Walnut County; 40–47% between Black and White students in Almond County; and 31–40% between Hispanic and White students in both counties. In addition, reading progress did not keep pace with mathematics, and growth in reading and mathematics proficiency rates in Grade 3 lagged Grade 5.

Australian students also made progress on NAPLAN; they achieved larger growth in reading than in numeracy across both year levels. In fact, numeracy scores for the majority of school groups in Year 3 declined or made no significant growth in 2103 compared to 2008. Across both year levels and subject domains, the two lowest-performing groups were schools in the Northern Territory and special schools serving only students with special needs. However, on the whole, they also made one of the largest absolute score gains in reading albeit not in numeracy. Despite these growth rates, a proficiency gap of over 50% exists between special and primary schools across both domains in Year 5; between NT schools and schools in Victoria (one of the two highest performing jurisdictions), the gaps were 34% and 38% in numeracy and reading, respectively. The movements of achievement gaps between advantaged and disadvantaged Australian schools across different background

categories were also mixed, with some gaps widening while others narrowed.

According to the latest analysis conducted by the Grattan Institute (Goss, Sonnemann, Chisholm, & Nelson, 2016), an independent Australian think tank, the current gaps between advantaged and disadvantaged schools in Australia translate to a one–two year difference in schooling in Year 3. On the other hand, the results of students with a LBOTE background yielded no specific patterns. The LBOTE findings validated scholarly suggestions that the LBOTE category does not provide useful information for schools (Creagh, 2014). Lastly, with the majority of school groups showing a statistically significant increase in withdrawal and absentee rates between 2008 and 2013, the quantitative findings also confirmed the increasing practice of withdrawal from the NAPLAN assessment reported by educators in a recent study (Dulfer et al., 2012).

While direct comparison between the two sets of results cannot be made due to assessment differences, as well as student- (California) versus school-data sample (Australia), a few valuable insights can still be drawn based on the growth rates against their respective baseline scores. First, it can be concluded that Californian disadvantaged student outcomes have generally improved, while the results are still inconclusive in Australia. Secondly, each country's respective achievement gaps illuminate a stark difference in education equity. In Australia, NT schools, one of the lowest performing groups, trailed their advantaged counterparts by an average of 32% in Year 3 reading proficiency acquisition. However, all other gaps between disadvantaged and advantaged groups were in single-digits to low-teens contrast to the gaps in California. For example, in Almond and Walnut Counties, respectively, the reading proficiency gap was 43% between Grade 3 Black and White students,

and 32–37% between low-SES and high-SES students. In comparison, the SES-based reading proficiency gap was 12% in Year 3 in Australia. These noticeable differences in the two countries' achievement gaps confirm a conclusion in an OECD report (Schleicher, 2014) stating that Australia's education system is more equitable.

Thirdly, relative to each country's own proficiency measures, more Australian students achieved proficiency standards than did Californian students. While results from the Northern Territory clearly show the significant educational challenge facing Indigenous students, many more schools serving this student population (over 60%) met proficiency standards in both Year 3 domains than did every disadvantaged subgroup in California. On average, fewer than 20% US ELLs, and fewer than or just about a third of all other disadvantaged subgroups in Grade 3, met the Californian reading proficiency standards. Most disadvantaged groups fared better in meeting mathematics proficiency, but these rates are also far behind those of disadvantaged schools in Australia where a majority of the schools met their respective minimum proficiency standards. Many factors could have contributed to these differences. It may be that the Australian national partnership programs are better at targeting funding to support areas of greatest need than the broader Title I funding is at supporting struggling US schools; hence further contributing to the overall equity factors presented in an OCED report about school equity (OECD, 2013; Schleicher, 2014). Furthermore, albeit imperfect, the funding model in Australia discussed in Chapter 2 might result in less extreme achievement results than those in the US. Lastly, the national curriculum, or prior to that state curricula, in Australia might have provided clearer guidance to schools on how to achieve goals, rather than a mere directive to achieve a goal of 800 API. Knowing what without knowing how got

US schools less than half way to the national goal. In a more negative light, however, a recent publication by the Grattan Institute (Goss et al., 2016) concludes that the minimum proficiency standards in Australia “are set *very low*” (p. 2) and therefore must change to identify those students who are truly behind. This conclusion might be true relative to Australia’s education standard. However, Australian students’ consistently higher PISA results, compared to those of US students, suggest that its current standards provide Australian students with better foundational skills than those in the US.

Finally, the largest achievement gap difference between the Californian and Australian data samples is that between students with disability and those without disability. However, because the data samples are so different in their representation of disability status, where one included only students with very specific or severe needs and enrolled in special schools (Australia) and another contained students with disability and learning difficulty (California), conclusions about achievement cannot and should not be made. Instead, the difference highlights the urgency for the Australian local and federal education authorities to capture and release data pertaining to students with disability in mainstream schools, as it already does for the LBOTE and Indigenous populations.

#### **Qualitative findings.**

The absolute growth in external assessments at the six participating schools can be explained largely by their decisions to engage with data to inform strategic, curricular and instructional decisions. Data engagement was first driven by the need to meet accountability, and later by teachers’ individual and collective beliefs in data and in the process of data engagement to raise the outcomes of their disadvantaged

student segments. The case studies provide valuable insights that policy makers, regional and district officials, and school administrators would find useful to encourage data-driven practice.

First, transformative and instructional leaders are necessary to establish a culture of data-driven practice (Halverson et al., 2007; Wayman, Spring, et al., 2012). Transformative leaders in the present study used data strategically to establish a student-centric culture that operates from two principles: (1) poor student results are not indicative of student capability, but of a misalignment of curriculum and instruction; and (2) effective teaching is centred on the needs of students and not on delivering chapters in the text. Transformative leaders also disrupt the entrenched culture of isolation practice by bringing teachers to the table for joint data analysis and collaboration. These leaders further promote and model the process of data practice. Most importantly, they create an environment where sources of individual and collective teacher efficacy development can naturally emanate. Finally, effective instructional leaders help teachers to deflect accountability pressure, which can have a significant impact on teacher affective state.

Secondly, the type of data utilised and nature of data use varied little across the schools in this study, or from that identified in existing data-use research (Kerr et al., 2006; Young & Kim, 2010). Student external and internal performance data are mainly used in the following contexts: goal setting; identifying gaps in curriculum; program and instructions; monitoring student and teacher progress; informing student grouping; guiding resource allocation and professional development; and reporting to various constituents. The absence of new uses in this study might be because these uses are those for which “highly effective schools and classroom

teachers have been using data for years” (Mandinach & Jackson, 2012, p. 11).

However, because the practice has not been proliferated, as discussed in Chapter 1, schools not yet on this journey can still face challenges, as it is likely that there will be resistant teachers, as there were at the schools in this study. Therefore, beyond knowing what to do with data, it is imperative that interested school leaders and regional administrators also focus their attention on factors that will strengthen teacher internal motivations to engage with data.

Thirdly, accountability pressure, as evidenced in the case of most participating US schools and to some extent at Koala School, can stifle teacher creativity and provide students with only basic skills, including those students who are currently advanced. As one principal mentioned, as much as his teachers would like to focus on character development, they would not do it at the expense of redirecting attention from the core subjects. Another admitted to not having paid much attention to extension related-programs. This goes against the concept of ‘community education’, that Hargreaves (2002) advocates in his high-profile book, *Teaching in the Knowledge Society: Education in the Age of Insecurity*. He argues that the narrow focus of today’s accountability-based education “marginalizes the attention to personal and social development that is the foundation of community, and eliminates interdisciplinary attention to global education that is at the heart of cosmopolitan identity” (p. 5).

Finally, neither an accountability requirement nor strong leadership alone is enough to bring about effective data-adoption and engagement if teachers’ behavioural beliefs are not considered. This is because “personal factors, and environmental events all operate as interacting determinants” to a behavioural

change and “any account of the determinants of human action must, therefore, include self-generated influence as a contributing factor” (Bandura, 1989, p. 1175). In this study, data proliferation and collective data engagement created a powerful environment for teachers to adjust their internal beliefs. In all cases, teachers came along on the journey when they saw the growth in student outcomes. Their narratives suggested that students’ positive outcomes boosted their perceived efficacy because positive outcomes accentuated their mastery experience. Across all schools, data enabled teachers to cater to students’ particular needs, which in turn raised student engagement and lowered student behavioural issues. As Sagor (2000) argues, data is knowledge, and knowledge increases efficacy. Many participants spoke with pride about the positive change in student engagement. Behavioural management and student engagement have been linked to educator affective state improvement, which in turn leads to teacher efficacy (Ho & Hau, 2004). It is entirely possible that successes in both academic and behavioural outcomes also contributed to raising teachers’ perceived control over the circumstances of their students’ disadvantaged backgrounds. The efficacy theory (Bandura, 1977) states that internally driven beliefs, while malleable, are stronger than external related factors; hence, it is possible to see why teachers have chosen to embrace data practice.

Because efficacy is most malleable in the first years of teaching (as Woolfolk found through her teacher efficacy research (Shaughnessy, 2004)), the joint data articulation, planning and best-practice sharing opportunities through the advent of data could be particularly important for schools to support the long-term development of teacher efficacy. A large body of work (Hargreaves, 2000; Sachs, 2016) advocates professional communities as a way forward and empirical studies

(Y. L. Goddard et al., 2007; McLaughlin & Talbert, 1993) have found a positive impact on student learning when collaborative groups focus on continuous improvement. As evidenced in the present study's findings, activities such as goal setting, and a student-centric focus and collegiality can all contribute to a long-term commitment to raise student outcomes. Without an opportunity and a compelling reason to come together to collaborate, the direct and indirect benefits of data-driven practice would have been elusive, particularly for teachers who were not technology savvy or who were more accustomed to the traditional way of teaching that is more suitable for what Hargreaves (2000) describes as the 'pre-professional' and 'autonomous profession' ages.

## **Implications of the Study**

### **Prospects for closing the achievement gaps.**

As found in the quantitative results of this study, absolute progress has been made across many student and school subgroups including disadvantaged students. However, disadvantaged student- and disadvantaged school-outcomes lag their advantaged counterparts significantly and achievement gaps persist. Nonetheless, there is a silver lining in these findings: achievement gaps persist because most groups made progress since 2008. This paper began with the notion that any incremental change in student achievement is worth evaluating. Therefore, any strategy that has the potential to raise student outcomes should be given consideration. While a significant amount of work remains to raise the performance of disadvantaged students, there is evidence from the qualitative narratives from the six participating schools that data-driven practice has contributed to the academic



growth observed among disadvantaged students. Data-driven practice worked mostly because data can pinpoint student needs and instructional gaps. Such insight enabled teachers to target their interventions. It is no accident that data-driven practice is the central focus of policy makers across both continents (The Honorable Julia Gillard Australian MP, 2010, May 5; US Secretary of Education Ann Duncan, 2010) and the only element of the education reform that even its opponents could support (Hess & Finn Jr, 2007). However, it is necessary to highlight that standardised or external assessment data is not the only data source contributing to the benefits of data-driven practice. In fact, teachers in the current study indicated that it is the least informative data source, because it does not provide ongoing insights about individual student needs. Teachers rely more on formative assessments in their weekly data articulation sessions. To this end, policy makers' justification for standardised testing as a diagnostic tool is questionable. Many respondents in the study had acknowledged that it was not the standardised test data but the policy pressure that provided the impetus for data-driven practice. For this reason, one can argue that the continuation of data-driven practice and its associated benefits do not require external testing.

Although the academic growth since 2008 has been slow and not yet universal, if it continues, the prospects of more disadvantaged students acquiring proficiency are entirely imaginable. While 'closing the achievement gaps' is an aspirational goal, it is not a practical metric for which to measure school, teacher or student performance. Prominent education scholars (Darling-Hammond, 2010; Ravitch, 2010) have argued that, as long as economic gaps and education inequality persist, academic achievement gaps will linger. Given the historical social and

educational inequity in both countries, it would be naïve to believe that achievement gaps can be closed simply through testing or making use of data. In an attempt to project the convergence of various social and economic gaps between Indigenous and non-Indigenous Australians, researchers (Altman, Biddle, & Hunter, 2008) at the Centre for Aboriginal Economic Policy Research estimated that it would take another 63 years (starting in 1966) for attendance of youth (aged 15–24 years) at educational institutions to converge. Rather than focusing on the grand goal of closing or even narrowing the gaps, policy makers need to focus on attainable growth targets to empower disadvantaged students to achieve minimum proficiency standards. At the same time, they need to address other social factors that contribute to the academic gaps. Based on the two participating Australian schools' accounts, the Smarter Schools National Partnerships appear to have contributed to their positive outcomes and quality teaching. Continuation of these programs in Australia and consideration of similar programs in the US can help to continue the growth trends observed thus far.

However, if governments are truly committed to raising the outcomes of all students, more visibility must be given to students with disability. In both locations, big gaps remain regarding the progress of this student subgroup due to the lack of original data. In Australia, it was possible to evaluate students in special schools but not in mainstream schools; in California, it was possible to evaluate students who took the general assessment but not those who took the alternative assessments meant to accommodate their needs. As needs vary so much among students with disabilities, it is necessary to have transparency for all groups to appropriately support this subgroup.

**The 'real' value of data-driven practice.**

Policy makers justified standardised assessment by promoting its diagnostic value to schools and educators. This study found that beyond providing school and grade level trends and comparative data useful to school administrators, standardised results provided no diagnostic benefits to teachers. Instead, it was the process or operationalisation of data-driven practice that generated a host of indirect benefits at school. A large review (Sammons, Hillman, & Mortimore, 1995) of school effectiveness concluded that no single teaching style brings about school effectiveness; instead, there are characteristics that relate to school and classroom processes that contribute to school effectiveness. Paradigm shifts towards student-centric, goal-oriented and collaborative practice at the participating schools fit the definition of school and classroom process-related changes. Student data were the impetus to, and centre-piece in, these process changes. Data provided teachers with a reason to gather and to collaborate. The collective practice of diagnosing student needs, sharing practices, commiserating and celebrating personal or student failures or successes, all had an impact on teachers' decisions to be willing participants in data engagement. At data meetings, social persuasion and subjective norm occurred organically as part of the social interaction process; and they influenced teachers' judgments of personal and collective efficacy and control in affecting student outcomes. As Goddard et al. (2004) suggest, perceived collective efficacy has a strong influence on the normative environment of a school because, more than other common measures of school context, it is strongly related to teacher personal efficacy, which, in turn, influences student performance as Bandura's (1993) study demonstrates.

Through the participating schools' experience, it is also possible to identify some elements of teacher professionalism found in teacher identity discourse or advocated by leading experts (Hargreaves, 2002; Sachs, 2000, 2001). They include active trust, where professionals come together to negotiate shared values, principals and strategies and to take responsibility for each other's success (Sachs, 2000); practitioner research, where educators engage in mutual exchange and shared inquiry to improve practice (Sachs, 2000); and community engagement, where they move beyond the curriculum to engage other stakeholders to support the improvement efforts (Hargreaves, 2002). Nonetheless, it is undeniable that elements of the more controversial 'managerial professionalism' continue to dominate and here changes need to happen, including relaxing the "long line of authority in terms of [teachers'] accountability for reaching measurable outcomes that stretches through the principal, to the district/regional office, to the central office" (Sachs, 2001, p. 152).

Students benefited from the data-driven process as well. The schools' assessment results provided evidence of academic improvement and participants reported significant behavioural improvement. The shift to student-centric practice (made possible by data) contributed to these improvements through increased student engagement, because data enabled teachers to address their individual needs. More importantly, data validated the notion that 'every child can learn' and it is the school's responsibility to set high expectations and provide supportive services to help students achieve expectations (Slavin, 1996). Detailed and ongoing analyses of student summative and formative data illuminated where the real challenges lay: curriculum and instruction. The collaborative and professional environment in which

these insights were revealed also made it easier for teachers to own up to the problems and find appropriate solutions to support their students. For many of the participants, data-driven practice, adopted initially for accountability compliance, has evolved into a valued process for advancing student learning and professional development.

### **Lessons from policy-borrowing and policy-lending.**

The Australia and US reform policies share similar high-level features: testing, transparency, and public management. This is because Australia borrowed the policy concept from the US (Lingard, 2010). As discussed in Chapter 2, Australia's implementation of these features can be said to be more supportive and low-stakes, compared to the US's implementation of NCLB, which is considered high-stakes because it is sanction-oriented. Despite these differences, participants in both countries felt the stress of accountability. What this means is that perceived accountability pressure creates as much stress for educators as does real accountability pressure. The perceived pressure in Australia came, not from any actual mandate from the education authority, but from the media's misuse of league tables and possibly from ACARA's prominent publication of similar school results on *My School*. The actual pressure in America came from NCLB, which mandated schools to deliver 100% proficiency through incremental growth targets. Both forms of pressure can have a negative impact on schools and on teachers' affective states as they work towards meeting the school goals through data-driven practice.

Success in data-driven practice does not guarantee a reduction of pressure. A Year 5 Teacher at Bilby who believed in the data process said, "I feel the pressure, I definitely feel that". This was a sentiment shared by other participants regardless of

age, experience or grade. As much as data-driven practice has helped them to improve as teachers, and helped their students to improve as learners, no one enjoyed the constant pressure of having their results evaluated on an ongoing basis, even though administrators claimed that the data-driven process was not meant to be 'evaluative'. Even teachers whose class or school tends to do well felt that the 'community pressures' as well as the 'high expectation' simply put too much pressure on them and their students to perform.

Sachs (2016) argues that this type of "performance culture... ha[s] created the conditions for a more conservative and risk-averse teaching profession" and has turned teacher professionalism into "controlled or compliant professionalism" (p. 423). Compliant schools and teachers are less likely to be able to help students to develop the problem-solving and critical-thinking skills or creativity required to produce new knowledge that is necessary to succeed in the 'global' or 'knowledge economy' (Hargreaves, 2002; Sachs, 2016). Potential solutions to lessen the pressure while ensuring that positive growth trends continue include: (1) shifting the focus from meeting a particular benchmark, such as an 800 API score, to growth or continuous improvement; and (2) aligning external accountability and internal accountability by expanding learning communities to include more stake-holders to negotiate targets and programs to empower teachers to properly support student learning.

In fact, Australia is focusing on growth and the difference between Australia and the US is that the former takes a more strategic view of school operation, curricula and program development. In comparison, the US participating schools seemed to be bogged down at all times by the three cycles of the benchmark

assessments and the annual assessment. While some US administrators reported that continuous improvement was what their district cared about (as suggested by the data officer at Walnut), teachers still felt under pressure. This suggests that the entire school system, starting with the Department of Education, must also shift its accountability focus, so that the community can recalibrate its expectations of schools. Only then would teachers have the flexibility to experiment with curricula, programming and instruction to differentiate and support students' academic growth. The reauthorisation of the US education act, ESSA, in December 2015 has indeed opted for continuous growth over an absolute target. While this shift might take some time to funnel down to schools, it is movement in a positive direction. Focusing on growth rather than an unrealistic benchmark could also mitigate some side effects of accountability, such as diminishing creativity, flexibility, or attention to high performers. The evidence from Bilby, which operates in a less restrictive accountability environment and which is recognised as a model of excellence in quality teaching, hinted of that possibility.

Since learning or professional communities have engendered trust, collaboration, and accountability at grade and school levels, perhaps they can expand to also include regional or state education officials so that achievement goals can be aligned. Scholars (Hargreaves, 2002; Ravitch, 2010) have argued that educators should be trusted to have accountability for their student outcomes because they "have the collective wisdom of the profession to self-regulate practice" (Sachs, 2016, p. 416). What they lack then is trust from policy makers and from the community; and 'managerialism' in the form of external accountability, as discussed in Chapter 2, conveys just the opposite. As Sachs (2000) suggests, *active trust*

demands that different parts of the educational enterprise work collaboratively, not oppositionally, and this requires not only that “each party inhabit each other’s castles... but rather, that each party at least looks inside the other’s castles” (p. 82). If policy makers participate in these school-level professional communities on an ongoing basis, they might gain more insights into the complexity of different school contexts and environments. This can lead to the development of shared goals that are both motivational and reachable.

### **Limitations and Further Research Direction**

The findings in this study provide a starting point for policy makers considering the implementation of an appropriate approach to demand accountability, while providing schools with flexibility in how to meet goals. They also provide a starting point for schools that are interested in adopting data-driven practice in a manner that has a high chance of being embraced by all constituents. There remains a need for significant future research into data-driven practice as a means to raise the outcomes of disadvantaged students.

A key weakness in this study involves the lack of visibility of the performance of students with disabilities, arising either from challenges in accessing data or from an under-represented data sample. The result is an incomplete picture of this student subgroup. While the qualitative results clarified how students with disabilities are currently being supported in various structures, it is unclear how these support systems have an impact on their outcomes, since their outcomes were unavailable to the researcher. Future research specific to this student subgroup would be very useful.



A second related weakness is the muddled view of the LBOTE students' progress in Australia. Given the results of English language learners in California, it is hard to believe that no student in this category requires support in Australia. Further research combining other data sources and surveys could be helpful to parse out those who require support and those whose needs are overstated.

Thirdly, this study focused on leaders who believed in data-driven practice and teachers who have had a chance to hone that practice. Participants in this study also alluded to varying attitudes of teachers who were not yet on board with data-driven practice. Investigating the belief system of those educators who are not yet convinced of the merits of the process would complement these findings and offer useful insights to policy makers as they strategise how to encourage schools to adopt the practice.

Finally, in education, there is a general belief that secondary school systems are more complex than primary schools, since they are typically organised by subject domain rather than home rooms. To what extent will collaborative data practice apply at the secondary-school level, and would similar opportunities arise to influence teacher belief systems? Exploring the similarities and differences within the secondary-school context would assist policy makers to focus on elements that could buttress data-driven practice at all levels.

## **Conclusion**

In Australia and in the US, accountability policy has generated more criticism than applause, whether from scholars, practitioners, parents, researchers or the media. According to the existing literature, as well as sentiments shared by

respondents, the pressure that arises from accountability policy is something that many would like to see disappear. However, until the next reiteration of these policies (and there is no guarantee that this pressure will be lifted), educators must find ways to help disadvantaged students to raise their performance. In spite of the unwelcome pressure, participating administrators and teachers in the current study appear to have leveraged data-driven practice using a combination of summative and formative data to exercise control over their operational and instructional process to bring desirable outcomes to disadvantaged students. According to Bandura (1995), “the ability to affect outcomes makes them [people] predictable. Predictability fosters adoptive preparedness” (p. 1). Indeed, it was participants’ beliefs that they could effect outcomes that motivated many of them to stay the course of data-driven practice. In an era of strong accountability pressure, efficacy beliefs among different constituents in schools are ever more important if schools are to succeed in getting more disadvantaged students to proficiency levels in literacy and numeracy. Collaborative data-driven practice has facilitated the development of student, teacher, administrator and the community’s sense of efficacy towards achieving better outcomes for their students. Changes in participants’ attitudes, intentional practice and collective efficacy beliefs have also brought cascading direct and indirect benefits across all areas of schools and among all school constituents.

Yet, it must not be forgotten that these benefits can come at the expense of the advantaged students, curriculum, programming and instructional flexibility, and creativity. One may argue that the size of the achievement gaps between advantaged and disadvantaged students justify a narrow focus on disadvantaged

students, and largely on literacy and numeracy. However, this narrow focus is likely to create new challenges for the long term if policy makers do not change current accountability requirements or lessen the pressure placed on educators, students and families.

## References

- Aarts, H., Verplanken, B., & Van Knippenberg, A. (1998). Predicting behavior from actions in the past: Repeated decision making or a matter of habit? *Journal of Applied Social Psychology, 28*(15), 1355-1374.
- Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001). NAEP math performance and test accommodations: Interactions with student language background. *CSE Technical Report*
- ACARA. (2014). *NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2014*. Sydney: Author.
- ACARA. (2015). Guide to understanding ICSEA (Index of community socio-educational advantage) values. Retrieved from [www.myschool.edu.au](http://www.myschool.edu.au)
- ACARA. (n.d.-a). MySchool. Retrieved from <http://www.myschool.edu.au/>
- ACARA. (n.d.-b). NAP National Assessment Program. Retrieved from <http://www.acara.edu.au/assessment/assessment.html>
- Ainley, J., & Gebhardt, E. (2103). *Measure for measure: A review of outcomes of school education in Australia*. Retrieved from [http://research.acer.edu.au/ar\\_misc/15](http://research.acer.edu.au/ar_misc/15).
- Ainscow, M., Booth, T., & Dyson, A. (2006). *Improving schools, developing inclusion*. Milton Park, Abingdon: Routledge.
- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11-39). Berlin, Heidelberg: Springer.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Process, 50*(2), 179-211.
- Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control and the theory of planned behavior. *Journal of Applied Social Psychology, 32*(4), 665-683.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology, 22*(5), 453-474.
- Altman, J. C. (2000). The economic status of indigenous Australians. Retrieved from Australian National University <http://hdl.handle.net/1885/40098>
- Altman, J. C., Biddle, N., & Hunter, B. H. (2008). How realistic are the prospects for 'closing the gaps' in socioeconomic outcomes for Indigenous Australians? *Discussion Paper No. 287*. Canberra, ACT: The Centre for Aboriginal Economic Policy Research.
- Amrein-Beardsley, A. L. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*(2), 65-75.
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *The British Psychological Society, 40*(4), 471-499.
- Athanasou, J. A. (2010). NAPLAN and MySchool survey. *Independent Education Union of Australia*.

- Schools Assistance (Learning Together—Achievement Through Choice and Opportunity) Act. Act No. 135 (2004) (n.d.).
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267.
- Australian Bureau of Statistics. (2014). Socioeconomic factors and student achievement in Queensland. (Report 4261.3) *Educational outcomes, experimental estimates, Queensland, 2011*.
- Australian Education Union. (2013). Submission to the Senate Inquiry into the effectiveness of the national assessment program: Literacy and numeracy.
- Australian Government Department of Education. (2014). Review of the Australian Curriculum: Final report. Retrieved from [https://docs.education.gov.au/system/files/doc/other/review\\_of\\_the\\_national\\_curriculum\\_final\\_report.pdf](https://docs.education.gov.au/system/files/doc/other/review_of_the_national_curriculum_final_report.pdf).
- Australian Government Productivity Commission. (2013). Chapter 4: School education. *Report on Government Services 2013*. Canberra: Author. Retrieved from <http://www.pc.gov.au/research/ongoing/report-on-government-services/2013/2013>.
- Australian Ministerial Council on Education, Employment, Training and Youth Affairs. (2008). *Melbourne Declaration on educational goals for young Australians*. Melbourne: Author. Retrieved from [http://www.curriculum.edu.au/verve/resources/National\\_Declaration\\_on\\_the\\_Educational\\_Goals\\_for\\_Young\\_Australians.pdf](http://www.curriculum.edu.au/verve/resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf).
- Australian Primary Principals Association. (2009). *Australian primary principals association position paper on the publication of nationally comparable school performance information*. ACT: Author. Retrieved from <http://www.appa.asn.au/wp-content/uploads/2015/08/School-performance-information.pdf>.
- Australian Primary Principals Association. (2010). *The reporting and use of NAPLAN*. ACT: Author. Retrieved from [www.appa.asn.au/papers/Reporting-and-use-of-NAPLAN.pdf](http://www.appa.asn.au/papers/Reporting-and-use-of-NAPLAN.pdf).
- Ball, S. J. (2005a). The teacher's soul and the terrors of performativity. In S. J. Ball (Ed.), *Education policy and social class: The selected works of Stephen J. Ball* (pp. 143-156). London: Routledge.
- Ball, S. J. (2005b). What is policy? Texts, trajectories and toolboxes. In S. J. Ball (Ed.), *Education policy and social class. The selected works of Stephen J. Ball* (pp. 43-53). London: Routledge.
- Ball, S. J. (2012). *Global education inc.: New policy networks and the neo-liberal imaginary*. London, U.K.: Routledge.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122-147.
- Bandura, A. (1986a). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, 4(3), 359-373.
- Bandura, A. (1986b). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

- Bandura, A. (1989). Human agency in social cognitive theory. *American psychologist*, 44(9), 1175-1184.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117-148.
- Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing society* (pp. 1-45). New York, New York: Cambridge University Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and Company.
- Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, 9, 75-78.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52(1), 1-26.
- Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. McKinsey & Company.
- Beadie, N. (2004). Moral errors and strategic mistakes: Lessons from the history of student accountability. In K. A. Sirotnick (Ed.), *Holding accountability accountable: What ought to matter in public education* (pp. 35-50). New York, NY: Teachers College Press.
- Beck, L., & Ajzen, I. (1991). Predicting dishonest actions using the theory of planned behavior. *Journal of research in personality*, 25(3), 285-301.
- Biddle, B. J., & Berliner, D. C. (2002). A research synthesis / unequal school funding in the United States. *Educational Leadership*, 59(8), 48-59.
- Blinder, A. (2015, April 1). Atlanta Educators Convicted in School Cheating Scandal, *The New York Times*. Retrieved from <http://www.nytimes.com/2015/04/02/us/verdict-reached-in-atlanta-school-testing-trial.html? r=0>
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1-39.
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42(2), 2321-2268.
- Booth, T. (2003). Viewing inclusion from a distance: Gaining perspective from comparative study. In M. Nind, J. Rix, K. Sheehy & K. Simmons (Eds.), *Inclusive education: Diverse perspectives*. London: David Fulton Publishers.
- Booth, T., & Ainscow, M. (1998). *From them to us: An international study of inclusion in education*. London: Routledge.
- Bousfield, K., & Ragusa, A. T. (2014). A sociological analysis of Australia's NAPLAN and MySchool Senate Inquiry submissions: The adultification of childhood? *Critical Studies in Education*, 55(2), 170-185.
- Bradley, S., Draca, M., Green, C., & Leeves, G. (2007). The magnitude of educational disadvantage of indigenous minority groups in Australia. *Journal of Population Economics*, 20(3), 547-569.
- Broadfoot, P. (2000). Comparative education for the 21st Century: Retrospect and prospect [Special number 23]. *Comparative Education*, 36(3), 357-371.
- Broadfoot, P. (2003). Globalisation in comparative perspective: Macro and micro. *Comparative Education*, 39(4), 411-413.

- Byrne, D., Olsen, W., & Duggan, S. (2009). Causality and interpretation in qualitative policy-related research. In D. Byrne & C. C. Ragin (Eds.), *The SAGE handbook of case-based methods* (pp. 511-521). Los Angeles: SAGE.
- Caldwell, B. J. (2010). Why the education revolution is not transforming our school. Retrieved from [https://www.researchgate.net/publication/242456276\\_WHY\\_THE\\_EDUCATION\\_REVOLUTION\\_IS\\_NOT\\_TRANSFORMING\\_OUR\\_SCHOOLS](https://www.researchgate.net/publication/242456276_WHY_THE_EDUCATION_REVOLUTION_IS_NOT_TRANSFORMING_OUR_SCHOOLS)
- California Department of Education. (2013a). *California English language development standards implementation plan*. Sacramento: Author. Retrieved from <http://www.cde.ca.gov/sp/el/er/documents/nov2013impplanfinal.pdf>.
- California Department of Education. (2013b). California standardized testing and reporting: Post-test guide technical information. Retrieved from [www.startest.org/pdfs/STAR.post-test\\_guide.2013.pdf](http://www.startest.org/pdfs/STAR.post-test_guide.2013.pdf).
- California Department of Education. (n.d.-a). California assessment of student performance and progress (CAASPP) system. Retrieved from <http://www.cde.ca.gov/ta/tg/ca/>
- California Department of Education. (n.d.-b). The Public Schools Accountability Act of 1999 - CalEdFacts. Retrieved from [www.cde.ca.gov/ta/ac/pa/cefpsaa.asp](http://www.cde.ca.gov/ta/ac/pa/cefpsaa.asp)
- California Department of Education. (n.d.-c). Testing and accountability. Retrieved from <http://www.cde.ca.gov/ta/>
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Education Evaluation and Policy*, 24(4), 305-332.
- Casserly, M. (2007). America's great city schools: Moving in the right direction. In F. M. Hess & C. E. Finn Jr (Eds.), *No remedy left behind: Lessons from a half-decade of NCLB* (pp. 43-65). Washington, D.C.: The AEI Press.
- Center on Education Policy. (2005). *From the capitol to the classroom: Year 3 of the No Child Left Behind Act*. Washington, D.C.: Author. Retrieved from <http://epsl.asu.edu/epru/articles/EPRU-0504-120-OWI.pdf>.
- Chang, M. K. (1998). Predicting unethical behavior: A comparison of the theory of reasoned action and the theory of planned behavior. *Journal of Business Ethics*, 17(16), 1825-1834.
- Chatterji, M. (2005). Evidence on "what works": An argument for extended-term mixed method (ETMM) evaluation designs. *Educational Researcher*, 34(5), 14-74.
- Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology*, 98(3), 489-507.
- Cheng, K. M. (2010). Education for all, but for what? In J. E. Cohen & M. B. Malin (Eds.), *International perspectives on the goals of universal basic and secondary education* (pp. 43-65). New York: Routledge, Taylor & Francis Group.
- Cimera, R. E. (2007). *Learning disabilities: What are they? Helping teachers and parents understand the characteristics*. Lanham, Maryland: Rowman & Littlefield Education.

- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). The academic achievement gap in Grades 3 to 8. *NBER Working Paper (No.12207)*.
- COAG Reform Council. (2013). Education in Australia 2012: Five years of performance: Report to the Council of Australian Governments. *COAG Reform Council*.
- Cobbold, T. (2010). Rorting and cheating of school results is the future under My School. *Save Our School Education Policy Brief*. Retrieved from [http://www.saveourschools.com.au/file\\_download/40](http://www.saveourschools.com.au/file_download/40).
- Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*(4), 469-495.
- Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education, 5th edition*. London: Routledge.
- Cole, C. (2006). Closing the achievement gap series: Part III. What is the impact of NCLB on the inclusion of students with disabilities? *Center for Evaluation & Education Policy, 4*(11), 1-12.
- Common Core State Standards Initiative. (n.d.-a). Preparing America's students for success. Retrieved from <http://www.corestandards.org/>
- Common Core State Standards Initiative. (n.d.-b). Standards in your state. Retrieved from <http://www.corestandards.org/standards-in-your-state/>
- Conley, D. T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology, 28*(15), 1429-1464.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles: SAGE.
- Council for Learning Disabilities. (n.d.). Quantitative and qualitative research: Working together to inform education for students with learning disabilities. Retrieved from <http://www.council-for-learning-disabilities.org/quantitative-and-qualitative-research-working-together-to-inform-education-for-students-with-learning-disabilities>
- Council of Australian Governments. (n.d.). Schools and education. Retrieved from [https://www.coag.gov.au/schools\\_and\\_education](https://www.coag.gov.au/schools_and_education)
- Cranston, N., Kimber, M., Mulford, B., Reid, A., & Keating, J. (2010). Politics and school education in Australia: A case of shifting purposes. *Journal of Educational Administration, 48*(2), 182-195.
- Crawley, F. E. (1990). Intentions of science teachers to use investigative teaching methods: A test of the theory of planned behavior. *Journal of Research in Science Teaching, 27*(7), 685-697.
- Creagh, S. (2014). A critical analysis of problems with the LBOTE category on the NAPLAN test. *The Australian Educational Researcher, 41*(1), 1-23.
- Creswell, J. W. (2003). *Research design: qualitative, quantitative and mixed methods approaches* (2nd ed.). London: SAGE.
- Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, New Jersey Pearson Education Inc.



- Creswell, J. W., & Garrett, A. L. (2008). The "movement" of mixed methods research and the role of educators. *South African Journal of Education, 28*(3). Retrieved from [http://www.scielo.org.za/scielo.php?pid=S0256-01002008000300003&script=sci\\_arttext&tlng=pt](http://www.scielo.org.za/scielo.php?pid=S0256-01002008000300003&script=sci_arttext&tlng=pt)
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks: SAGE Publications.
- Crossley, M. (2014). Global league tables, big data and the international transfer of educational research modalities. *Comparative Education, 50*(1), 15-26.
- Crossley, M., & Vulliamy, G. (1984). Case-study research methods and comparative education. *Comparative Education, 20*(2), 193-207.
- Crossley, M., & Watson, K. (2003). *Comparative and international research in education, globalisation, context and difference*. London, New York: Routledge.
- Cuban, L. (1988). A fundamental puzzle of school reform. *Phi Delta Kappan, 70*(5), 341-344.
- Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19*(1), 3-13.
- Cuban, L. (2004). Looking through the rearview mirror. In K. A. Sirotnick (Ed.), *Holding accountability accountable: What ought to matter in public education* (pp. 18-34). New York, NY: Teachers College Press.
- Cuban, L. (2014). School leaders as marathoners, not sprinters [Blog post]. Retrieved from <http://larrycuban.wordpress.com/>
- Darling-Hammond, L. (2004). From "separate but equal" to "No Child Left Behind": The collision of new standards and old inequalities. In D. Meier & G. Wood (Eds.), *Many children left behind* (pp. 3-32). Boston: Beacon Press.
- Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. *Race Ethnicity and Education, 10*(3), 245-260.
- Darling-Hammond, L. (2010). *The flat world and education*. New York, London: Teachers College Press.
- Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing school systems use data to improve instruction for elementary students* California: Center on Educational Governance, University of Southern California.
- Davies, M. (2012). Accessibility to NAPLAN assessments for students with disabilities: A 'fair go'. *Australasian Journal of Special Education, 36*(1), 62-78.
- Davis, L. E., Ajzen, I., Saunders, J., & Williams, T. (2002). The decision of African American students to complete high school: An application of the theory of planned behavior. *Journal of Educational Psychology, 94*(4), 810-819.
- Day, C., Sammons, P., & Gu, Q. (2008). Combining qualitative and quantitative methodologies in research on teachers' lives, work, and effectiveness: From integration to synergy. *Educational Researcher, 37*(6), 330-342.
- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.
- Dempsey, I., & Davies, M. (2013). National test performance of young Australian children with additional educational needs. *Australian Journal of Education, 57*(1), 5-18.

- Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, 62(2), 80-88.
- Dixon, S., & Angelo, D. (2014). Dodgy data, language invisibility and the implications for social inclusion: A critical analysis of indigenous student language data in Queensland schools. *Australian Journal of Applied Linguistics*, 37(3), 213-233.
- Donnelly, K., & Wiltshire, K. (2014). *Review of the Australian curriculum*. Australian Government, Canberra: Department of Education and Training.
- Dowling, A. (2007). Australia's school funding system. *Policy Analysis and Program Evaluation Unit, ACER*. Retrieved from [http://research.acer.edu.au/cgi/viewcontent.cgi?article=1000&context=policy\\_analysis\\_misc](http://research.acer.edu.au/cgi/viewcontent.cgi?article=1000&context=policy_analysis_misc).
- DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Dulfer, N., Polesel, J., & Rice, S. (2012). *The experience of education: The impacts of high stakes testing on school students and their families - an educator's perspective*. Sydney: Whitlam Institute, the University of Western Sydney.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Sexton, H. (2007). School readiness and later achievement *Developmental psychology*, 43(6), 1428-1446.
- Eckes, S., & Swando, J. (2009). Special education subgroups under NCLB: Issues to consider. *Teachers College Record*, 111(11), 2479–2504.
- Ed-Data. (2011). Fiscal, demographics, and performance data on California's K-12 schools. from Education Partnership of the CDE, EdSource, and FCMAT <https://www.ed-data.k12.ca.us/Pages/Home.aspx>
- Ed-Data. (2012). District report and school report. Retrieved from [http://www.ed-data.k12.ca.us/App\\_Resx/EdDataClassic/fsTwoPanel.aspx?#!bottom=/layouts/EdDataClassic/profile.asp?level=04](http://www.ed-data.k12.ca.us/App_Resx/EdDataClassic/fsTwoPanel.aspx?#!bottom=/layouts/EdDataClassic/profile.asp?level=04)
- ED.gov. (2010). What's possible: Turning around America's lowest-achieving schools [Blog post]. Retrieved from <http://www.ed.gov/blog/2010/03/whats-possible-truning-around-americas-lowest-achieving-schools>
- Elliott, S. N., Davies, M., & Kettler, R. J. (2012). Australian students with disabilities accessing NAPLAN: Lessons from a decade of inclusive assessment in the United States. *International Journal of Disability, Development and Education*, 59(1), 7-19.
- Elmore, R. F., & Fuhrman, S. H. (2001). Holding schools accountable: Is it working? *Phi Delta Kappan*, 83(1), 67-72.
- Entwistle, N., & Smith, C. (2002). Personal understanding and target understanding: Mapping influence on the outcomes of learning. *British Journal of Educational Psychology*, 72(3), 321-342.
- Evetts, J. (2008). The management of professionalism. In S. Gewirtz, P. Mahony, I. Hextall & A. Cribb (Eds.), *Changing teacher professionalism: International trends, challenges and ways forward* (pp. 19-30). Hoboken, NJ: Routledge.
- Farrell, M. (2004). *Inclusion at the crossroads: Special education, concepts and values*. London: Taylor & Francis Routledge.

- Feldman, J., & Tung, R. (2001). *Whole school reform: How schools use the data-based inquiry and decision-making process*. Paper presented at the Annual meeting of the American Educational Research Association, April 2001, Seattle.
- Ferrari, J. (2012, September 3). Labor's 'top five' goal for schools, *The Australian*. Retrieved from <http://www.theaustralian.com.au/national-affairs/education/labors-top-five-goal-for-schools/story-fn59nlz9-1226463502869>
- Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE.
- Figlio, D., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbooks in economics* (Vol. 3, pp. 383-421). North-Holland, the Netherlands: Elsevier.
- Figlio, D., Rouse, C. E., & Schlosser, A. (2009). Leaving no child behind: Two paths to school accountability. *The Urban Institute*.
- Ford, M. (2013). Achievement gaps in Australia: What NAPLAN reveals about education inequality in Australia. *Race Ethnicity and Education*, 16(1), 80-102.
- Forlin, C., Chambers, D., Loreman, T., Deppeler, J., & Sharma, U. (2013). *Inclusive education for students with disability: A review of the best evidence in relation to theory and practice*. Retrieved from the Australian Research Alliance for Children and Youth website: [www.aracy.org.au](http://www.aracy.org.au).
- Friedel, J. M., Cortina, K. S., Turner, J. C., & Midgley, C. (2007). Achievement goals, efficacy beliefs and coping strategies in mathematics: The roles of perceived parent and teacher goal emphases. *Contemporary Educational Psychology*, 32(3), 434-458.
- Fryer, R. G. J., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics*, 86(2), 447-464.
- Fulcher, G. (1989). *Disabling policies? A comparative approach to education policy and disability*. London: The Falmer Press.
- Fullan, M. (2009). Large-scale reform comes of age. *Journal of Educational Change*, 10(2-3), 101-113.
- Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? . *Educational Researcher*, 36(5), 268-278.
- Gabriel, T. (2010, December 8). Parents embrace documentary on pressures of school, *New York Times*. Retrieved from Retrieved from <http://www.nytimes.com/2010/12/09/education/09nowhere.html>
- Gillard, J. (2008, April 4). Equity in education revolution [Speech]. Retrieved from [http://www.deewr.gov.au/Ministers/Gillard/Media/Speeches/Pages/Article\\_081022\\_145840.aspx](http://www.deewr.gov.au/Ministers/Gillard/Media/Speeches/Pages/Article_081022_145840.aspx)
- Gillard, J. (2010, April 15). A future fair for all - school funding in Australia, address to Sydney Institute [Speech]. Retrieved from [http://www.deewr.gov.au/Ministers/Gillard/Media/Speeches/Pages/Article\\_100416\\_113412.aspx](http://www.deewr.gov.au/Ministers/Gillard/Media/Speeches/Pages/Article_100416_113412.aspx)
- Gillard, J. (2010, June 10). Australia's productivity challenge: A key role for education [Speech]. Retrieved from [http://www.deewr.gov.au/Ministers/Gillard/Media/Speeches/Pages/Article\\_100610\\_130719.aspx](http://www.deewr.gov.au/Ministers/Gillard/Media/Speeches/Pages/Article_100610_130719.aspx)

- Gillborn, D., & Youdell, D. (2000). *Rationing education: policy, practice, reform and equity*. Buckingham, Philadelphia: Open University Press.
- Goddard, R. D. (2001). Collective efficacy: A neglected construct in the study of schools and student achievement. *Journal of Educational Psychology, 93*(3), 467-476.
- Goddard, R. D. (2002). Collective efficacy and school organization: A multilevel analysis of teacher influence in schools. In W. K. Hoy & C. G. Miskel (Eds.), *Theory and Research in Educational Administration* (Vol. 1, pp. 169-184). Greenwich, CT: Information Age Publishing.
- Goddard, R. D., & Goddard, Y. L. (2001). A multilevel analysis of the relationship between teacher and collective efficacy in urban schools. *Teaching and Teacher Education, 17*(7), 807-818.
- Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*(2), 479-507.
- Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33*(3), 3-13.
- Goddard, R. D., & Skrla, L. (2006). The Influence of school social composition on teachers' collective efficacy beliefs. *Educational Administration Quarterly, 42*(2), 216-235.
- Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record, 109*(4), 877-896.
- Godin, G., & Kok, G. (1996). The theory of planned behavior: A review of its applications to health-related behaviors. *American Journal of Health Promotion, 11*(2), 87-98.
- Goertz, M. E. (2005). Implementing the No Child Left Behind Act: Challenges for the States. *Peabody Journal of Education, 80*(2), 73-89.
- Goldring, E., & Berends, M. (2009). *Leading with data: Pathways to improve your school*. Thousand Oaks, CA: Corwin Press.
- Goldstein, R. A. (2011). Imaging the frame: Media representations of teachers, their unions, NCLB, and education reform. *Educational Policy, 25*(4), 543-576.
- Gonski, D. (2011, December). Review of school funding: Final report: Retrieved from <https://docs.education.gov.au/system/files/doc/other/review-of-funding-for-schooling-final-report-dec-2011.pdf>.
- Goss, P., Sonnemann, J., Chisholm, C., & Nelson, L. (2016). *Widening gaps: What NAPLAN tells us about student progress*. Grattan Institute. Retrieved from <http://grattan.edu.au/wp-content/uploads/2016/03/937-Widening-gaps.pdf>.
- Gottfried, M. A. (2010). Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal, 47*(2), 434-465.
- Graham, J. (2010). The trouble with MySchool. *Professional Voice, 8*(1), 7-11.
- Graham, L., & Jahnukainen, M. (2011). Wherefore art thou, inclusion? Analysing the development of inclusive education in New South Wales, Alberta and Finland. *Journal of Education Policy, 26*(2), 263-288.

- Grant, N. (2000). Tasks for comparative education in the new millennium [Special number 23]. *Comparative Education*, 36(3), 309-317.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks: Sage Publications.
- Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools*, 4(3), 211-239.
- Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2005). *The new instructional leadership: Creating data-driven instructional systems in schools*. Paper presented at the Annual Meeting of the National Council of Professors of Educational Administration, July 2005, Washington, D.C.
- Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The new instructional leadership: Creating data-driven instructional systems in schools. *Journal of School Leadership*, 17(2), 159-167.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques, third edition*. Waltham, MA: Elsevier.
- Haney, J. J., Czerniak, C. M., & Lumpe, A. T. (1996). Teacher beliefs and intentions regarding the implementation of science education reform strands. *Journal of Research in Science Teaching*, 33(9), 971-993.
- Hanushek, E., & Raymond, M. E. (2002). Sorting out accountability systems. *School accountability*. Stanford, CA: Hoover Institution Press, Stanford University.
- Hanushek, E., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Hardman, M. L., & Dawson, S. (2008). The impact of federal public policy on curriculum and instruction for students with disabilities in the general classroom. *Preventing School Failure*, 52(2), 5-11.
- Hardy, I., & Boyle, C. (2011). MySchool? Critiquing the abstraction and quantification of education. *Asia-Pacific Journal of Teacher Education*, 39(3), 211-222.
- Hargreaves, A. (2000). Four ages of professionalism and professional learning. *Teachers and Teaching*, 6(2), 151-182.
- Hargreaves, A. (2002). *Teaching in the knowledge society: Education in the age of insecurity*. New York: Teachers College Press. Retrieved from <https://phamtrung.wikispaces.com/file/view/TeachingInTheKnowledgeSociety.pdf>.
- Harr-Robins, J., Song, M., Garet, M., & Danielson, L. (2015). *School practices and accountability for students with disabilities (NCEE 2015-4006)*. Washington D.C.: Department of Education.
- Harris, D. N. (2007). High-flying schools, student disadvantage, and the logic of NCLB. *American Journal of Education*, 113(3), 367-394.
- Hartley, D. (2003). Education as a global positioning device: Some theoretical considerations. *Comparative Education*, 39(4), 439-450.
- Hattie, J. (2003). Teachers make a difference: What is the research evidence? *Australian Council for Educational Research*. Retrieved from [https://www.det.nsw.edu.au/proflearn/docs/pdf/qt\\_hattie.pdf](https://www.det.nsw.edu.au/proflearn/docs/pdf/qt_hattie.pdf).

- Hawaii Department of Education. (2011a). *School status and improvement (Report 107)*. Honolulu, HI: Author. Retrieved from <http://arch.k12.hi.us/school/trends/trends.html>.
- Hawaii Department of Education. (2011b). *School status and improvement (Report 236)*. Honolulu, HI: Author. Retrieved from <http://arch.k12.hi.us/school/trends/trends.html>.
- Hawaii Department of Education. (2011c). *Trend report: Educational and fiscal accountability (Report 107)*. Honolulu, HI: Author. Retrieved from <http://arch.k12.hi.us/school/trends/trends.html>.
- Hawaii Department of Education. (2011d). *Trend Report: Educational and fiscal accountability (Report 236)*. Honolulu, HI: Author. Retrieved from <http://arch.k12.hi.us/school/trends/trends.html>.
- Hawaii Department of Education. (n.d.-a). Strive HI performance system. Retrieved from <http://www.hawaiipublicschools.org/VisionForSuccess/AdvancingEducation/StriveHIPerformanceSystem/Pages/home.aspx>
- Hawaii Department of Education. (n.d.-b). Systems accountability office, system evaluation and reporting. Retrieved from <http://adc.hidoe.us/#/>
- Hehir, T. (2005). *New directions in special education, eliminating ableism in policy and practice*. Cambridge, MA: Harvard Education Press.
- Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*(2), 75-110.
- Hess, M. H., & Finn Jr, C. E. (Eds.). (2007). *No remedy left behind: Lessons from a half-decade of NCLB*. Washington, D.C.: The AEI Press.
- Hesse-Biber, S. N. (2010). *Mixed methods research, merging theory with practice*. New York, London: The Guilford Press.
- Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high and low-income countries. *American Journal of Sociology, 88*(6), 1162-1194.
- Heyvaert, M., Hannes, K., Maes, B., & Onghena, P. (2013). Critical appraisal of mixed methods studies. *Journal of Mixed Methods Research, 7*(4), 302–327.
- Ho, I. T., & Hau, K. T. (2004). Australian and Chinese teacher efficacy: Similarities and differences in personal instruction, discipline, guidance efficacy and beliefs in external determinants. *Teaching and Teacher Education, 20*(3), 313-323.
- Holme, J. J., & Wells, A. S. (2008). School choice beyond district borders: Lessons for the reauthorization of NCLB from interdistrict desegregation and open enrollment plans. *Improving on No Child Left Behind* (pp. 139-215). Retrieved from <http://www.prrac.org/pdf/ASW-interdistrict.pdf>.
- Howe, K. R. (2012). Mixed methods, triangulation, and causal explanation. *Journal of Mixed Methods Research, 62*(2), 89–96.
- Hoy, W. K., & Woolfolk, A. E. (1993). Teachers' sense of efficacy and the organizational health of schools. *Elementary School Journal, 93*(4), 355-372.
- Idol, L. (2006). Toward inclusion of special education students in general education: A program evaluation of eight schools. *Remedial and Special Education, 27*(2), 77-94.

- Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision-making: Barriers to the use of data to improve practice. *Teacher' College Record*, 106(6), 1258-1287.
- International Association for the Evaluation of Educational Achievement. (n.d.). Mission statement. Retrieved from [http://www.iea.nl/mission\\_statement.html](http://www.iea.nl/mission_statement.html)
- Irons, E. J., & Harris, S. (2007). *The challenges of No Child Left Behind: Understanding the issues of excellence, accountability, and choice*. Lanham, New York: Rowman & Littlefield Education.
- Jennings, J., & Stark Rentner, D. (2006). Ten big effects of the No Child Left Behind Act on public schools. *Phi Delta Kappan*, 88(2), 110-113.
- Jimerson, J. B., & Wayman, J. C. (2012). *Branding educational data use through professional learning: Findings from a study in three school districts*. Paper presented at the Annual Meeting of American Educational Research Association, Vancouver, British Columbia.
- Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed methods* (Third ed.). Los Angeles: SAGE Publications.
- Johnson, B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Johnson, R. S. (2002). *Using data to close the achievement gap: How to measure equity in our schools*. Thousand Oaks, CA: Corwin Press, Inc.
- Johnson, S. M. (1990). *Teachers at work: Achieving success in our schools*. Scranton, PA: Harpers Collins Publishers.
- Kalafatis, S. P., Pollard, M., East, R., & Tsogas, M. H. (1999). Green marketing and Ajzen's theory of planned behaviour: A cross-market examination. *Journal of Consumer Marketing*, 16(5), 441-460.
- Karvelas, P. (2013, March 28). Noel Pearson's Cape York trial 'changing lives'. *The Australian*. Retrieved from <http://www.theaustralian.com.au/national-affairs/indigenous/noel-pearsons-cape-york-trial-changing-lives/story-fn9hm1pm-1226608059546>
- Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112(4), 496-520.
- Kirsch, I. (1986). Early research on self-efficacy: What we already know without knowing that we knew. *Journal of Social Clinical Psychology*, 4(3), 339-358.
- Kitmitto, S. (2011). Measuring status and change in NAEP inclusion rates of students with disabilities: Results 2007-09 (NCES 2011-457). *National Center of Education Statistics*.
- Kitmitto, S., & Bandeira de Mello, V. (2008). Measuring the status and change of NAEP state inclusion rates for students with disabilities (NCES 2009-453). *National Center for Education Statistics*.
- Klassen, R. M. (2002). A question of calibration: A review of the self-efficacy beliefs of students with learning disabilities. *Learning Disability Quarterly*, 25(2), 88-102.

- Klassen, R. M. (2010). Confidence to manage learning: The self-efficacy for self-regulated learning of early adolescents with learning disabilities. *Learning Disability Quarterly, 30*(4), 19-30.
- Klein, A. (2014, January 8). No Child Left Behind turns 12 today. Now what? *Education Week*. Retrieved from [http://blogs.edweek.org/edweek/campaign-k-12/2014/01/no\\_child\\_left\\_behind\\_turns\\_twe.html](http://blogs.edweek.org/edweek/campaign-k-12/2014/01/no_child_left_behind_turns_twe.html)
- Klein, J. (2011). Scenes from the class struggle. *The Atlantic, 307*(5), 66-77.
- Klenowski, V. (2009). Public education matters: Reclaiming public education for the common good in a global era. *The Australian Educational Researcher, 36*(1).
- Klenowski, V. (2011). Assessment for learning in the accountability era: Queensland, Australia. *Studies in Educational Evaluation, 37*(1), 78-83.
- Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: the Australian story. *Assessment in Education: Principles, Policy & Practice, 19*(1), 65-79.
- Kozol, J. (2005). *The shame of the nation*. New York: Three Rivers Press.
- Krieg, J. M. (2008). Are students left behind? The distributional effects of the No Child Left Behind Act. *Education Finance and Policy, 3*(2), 250–281.
- Krieg, J. M. (2011). Which students are left behind? The racial impacts of the No Child Left Behind Act. *Economics of Education Review 30*(4), 654–664.
- Ladwig, J. G. (2010). What NAPLAN doesn't address (but could, and should). *Professional Voice, 8*(1), 35-39.
- Lamb, S., Rumberger, R., Jesson, D., & Tesse, R. (2004). School performance in Australia: Results from analyses of school effectiveness. Report for the Victorian Department of Premier and Cabinet. Retrieved from <https://www.eduweb.vic.gov.au/edulibrary/public/govrel/reports/schoolperformance-rpt.pdf>.
- Lauen, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis, 34*(2), 185-208.
- Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Cambridge, MA: The Civil Rights Project at Harvard University. Retrieved from <http://www.civilrightsproject.harvard.edu>.
- Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research, 78*(3), 608–644.
- Lee, J. (2010). Trick or treat: New ecology of education accountability system in the USA. *Journal of Education Policy, 25*(1), 73-93.
- Lee, J., Cerreto, F. A., & Lee, J. (2010). Theory of planned behavior and teachers' decisions regarding use of educational technology. *Educational Technology & Society, 13*(1), 152–164.
- Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis, 34*(2), 209-231.
- Lee, J. C., Zhang, Z., & Yin, H. (2011). A multilevel analysis of the impact of a professional learning community, faculty trust in colleagues and collective



- efficacy on teacher commitment to students. *Teaching and Teacher Education*, 27(5), 820-830.
- Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996 - 2005. *Leadership and Policy in Schools*, 4(3), 177-199.
- Lessinger, L. M. (1970). *Every kid a winner: Accountability in education*. New York: Simon and Schuster.
- Lincoln, Y. S., & Guba, E. G. (2000). The only generalization is: There is no generalization. In R. Gomm, M. Hammersley & P. Foster (Eds.), *Case study method* (pp. 27-45). London: SAGE Publications.
- Lingard, B. (2010). Policy borrowing, policy learning: Testing times in Australian schooling. *Critical Studies in Education*, 51(2), 129-147.
- Lingard, B. (2013). Historicizing and contextualizing global policy discourses: Test and standards based accountabilities in education. *The International Education Journal: Comparative Perspectives*, 12(2), 122-132.
- Lingard, B., Creagh, S., & Vass, G. (2012). Education policy as numbers: Data categories and two Australian cases of misrecognition. *Journal of Education Policy*, 27(3), 315-333.
- Lingard, B., Mills, M., & Hayes, D. (2006). Enabling and aligning assessment for learning: Some research and policy lessons from Queensland. *International Studies in Sociology in Education*, 16(2), 83-103.
- Louis, K. S., Febey, K., & Schroeder, R. (2005). State-mandated accountability in high schools: Teachers' interpretations of a new era. *Educational Evaluation and Policy Analysis*, 27(2), 177-204.
- Love, A., & Kruger, A. C. (2005). Teacher beliefs and student achievement in urban schools serving African American students. *The Journal of Educational Research*, 99(2), 87-99.
- Ma Rhea, Z. (2012). Partnership for improving outcomes in Indigenous education: Relationship or business? *Journal of Education Policy*, 27(1), 45-66.
- MacDonald, G. M., & Robinson, C. (1985). Cautionary tales about arbitrary deletion of observations: Or, throwing the variance out with the bathwater. *Journal of Labor Economics*, 3(2), 124-152.
- Madden, T. J., Ellen, P. S., & Ajzen, I. (1992). A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin*, 18(1), 3-9.
- Maddux, J. E., & Stanley, M. A. (1986). Self-efficacy theory in contemporary psychology: An overview. *Journal of Social and Clinical Psychology*, 4(3), 249-255.
- Mandinach, E. B., & Jackson, S. S. (2012). *Transforming teaching and learning through data-driven decision making*. Thousand Oaks: Corwin.
- Manstead, A., & Van Eekelen, S. (1998). Distinguishing between perceived behavioral control and self-efficacy in the domain of academic achievement intentions and behaviors. *Journal of Applied Social Psychology*, 1998, 28(15), 1375-1371-1392.
- Marks, G. N., Cresswell, J., & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation*, 12(2), 105-112.

- Marks, G. N., McMillan, J., & Hillman, K. (2001). Tertiary entrance performance: The role of student background and school factors. *LSAY Research Reports, 24.*: Australian Council for Educational Research.
- McLaughlin, M. W., & Talbert, J. E. (1993). Contexts that matter for teaching and learning: Strategic opportunities for meeting the nation's educational goals.
- Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). Implementing data-informed decision making in schools: Teacher access, supports and use. *US Department of Education.*
- Means, B., Padilla, C., & Gallagher, L. (2010). *Use of education data at the local level from accountability to instructional improvement.* Washington, D.C.: US Department of Education.
- Merriam-Webster. (n.d.). Reform. Retrieved from [www.merriam-webster.com/dictionary/reform](http://www.merriam-webster.com/dictionary/reform)
- Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San Francisco: Jossey-Bass
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (Fourth ed.). San Francisco: Jossey-Bass.
- Mertens, D. M., & Hesse-Biber, S. (2012). Triangulation and mixed methods research: Provocative positions. *Journal of Mixed Methods Research, 62*(2), 75-79.
- Mertler, C. A., & Vannatta, R. A. (2002). *Advanced and multivariate statistical methods: Practical application and interpretation* (Second ed.). Los Angeles: Pyrczak Publishing.
- Miller, P. W., & Voon, D. (2011). Lessons from MySchool. *The Australian Economic Review, 44*(4), 366-386.
- Miller, P. W., & Voon, D. (2012). Government versus non-government schools: A nation-wide assessment using Australian NAPLAN data. *Australian Economic Papers, 51*(3), 147-166.
- Mockler, N. (2013). Reporting the 'education revolution': MySchool.edu.au in the print media. *Discourse: Studies in the cultural politics of education, 34*(1), 1-16.
- Monpas-Huber, J. B. (2010). Explaining teachers' instructional use of state assessment data: A multilevel study of high school teachers in Washington State. *Journal of School Leadership, 20*(2), 208-237.
- Moolenaar, N. M., Slegers, P. J., & Daly, A. J. (2012). Teaming up: Linking collaboration networks, collective efficacy, and student achievement. *Teaching and Teacher Education, 28*(2), 251-262.
- Multicultural Development Association (MDA). (2010). Submission for the Senate Inquiry into the administration and reporting of NAPLAN testing.
- National Center for Education Statistics. (2013a). A first look: 2013 mathematics and reading. National assessment of educational progress at grades 4 and 8, (NCES 2014-451). Retrieved from [http://nationsreportcard.gov/reading\\_math\\_2013](http://nationsreportcard.gov/reading_math_2013).
- National Center for Education Statistics. (2013b). Mega-States: An analysis of student performance in the five most heavily populated states in the nation, (NCES 2013-450). Retrieved from <http://nces.ed.gov/nationsreportcard/pdf/main2011/2013450.pdf>.

- National Center for Education Statistics. (n.d.). NAEP technical documentation. Retrieved from <http://nces.ed.gov/nationsreportcard/tdw/analysis/infer.aspx>
- National Council of Teachers of English. (2008). English language learners: A policy research brief. *NCTE Policy Research Brief*, 1-8. Retrieved from [www.ncte.org/library/NCTEFiles/Resources/.../ELLResearchBrief.pdf](http://www.ncte.org/library/NCTEFiles/Resources/.../ELLResearchBrief.pdf)
- Nelson, S. W., & Guerra, P. L. (2014). Educator beliefs and cultural knowledge: Implications for school improvement efforts. *Educational Administration Quarterly*, 50(1), 67-95.
- Neuman, W. L. (2006). *Social research methods, qualitative and quantitative approaches* (Sixth ed.). Boston: Pearson.
- Neuman, W. L. (2009). *Understanding research*. Boston: Pearson Education Inc.
- New America EdCentral. (n.d.). No Child Left Behind. Retrieved from <http://www.edcentral.org/edcyclopedia/no-child-left-behind-overview/>
- New South Wales Department of Education. (n.d.). Visit our schools. Retrieved from <http://www.dec.nsw.gov.au/about-us/careers-centre/school-careers/focus-areas/learning-and-support/visit-our-schools>
- Nichols, S. L., & Berliner, D. C. (2005). The inevitable corruption of indicators and educators through high-stakes testing. *Education Policy Research Unit*.
- Noguera, P. A. (2004). Standards for what? Accountability for whom? Rethinking standards-based reform in public education. In K. A. Sirotnick (Ed.), *Holding accountability accountable: What ought to matter in public education* (pp. 66-81). New York, NY: Teachers College.
- NSW Department of Education and Communities. (n.d.-a). Best start. Retrieved from <http://www.curriculumsupport.education.nsw.gov.au/beststart/index.htm>
- NSW Department of Education and Communities. (n.d.-b). Selective high school students with language backgrounds other than English. Retrieved from [https://www.det.nsw.edu.au/media/downloads/about-us/statistics-and-research/key-statistics-and-reports/sel\\_high.pdf](https://www.det.nsw.edu.au/media/downloads/about-us/statistics-and-research/key-statistics-and-reports/sel_high.pdf).
- O'Day, J., Elledge, A., Le Floch, K. C., Taylor, J., & Anderson, L. (2009). State and local implementation of the No Child Left Behind Act, Volume V-Implementation of the 1 percent rule and 2 percent interim policy options. *US Department of Education, Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Services*.
- Oakes, J., Blasi, G., & Rogers, J. (2004). Accountability for adequate and equitable opportunities to learn. In K. A. Sirotnick (Ed.), *Holding accountability accountable: What ought to matter in public education* (pp. 82-99). New York, NY: Teachers College.
- Obama. (2012, February 9). President Obama speaks on No Child Left Behind reform [Speech]. <https://www.whitehouse.gov/photos-and-video/video/2012/02/09/president-obama-speaks-no-child-left-behind-reform>
- Obama, US President. (2015, December 10). Remarks by the President at Every Student Succeeds Act signing ceremony [Speech]. Retrieved from <https://www.whitehouse.gov/the-press-office/2015/12/10/remarks-president-every-student-succeeds-act-signing-ceremony>

- Ochs, K., & Phillips, D. (2004). Processes of educational borrowing in historical context. In D. Phillips & K. Ochs (Eds.), *Educational policy borrowing: Historical perspectives* (pp. 7-23). Oxford, U.K.: Symposium Books.
- OECD. (2003). *Chapter 1: Diversity, inclusion and equity: Insights from special needs provision*. Paris: OECD Publishing. Retrieved from <http://www.oecd.org/edu/innovation-education/26527517.pdf>.
- OECD. (2005). *School factors related to quality and equity: Results from PISA 2000*. <http://doi.org/10.1787/9789264008199-en>.
- OECD. (2007). *Students with disabilities, learning difficulties and disadvantages: Policies, statistics and indicators - 2007 Edition*. Paris: OECD Publishing. Retrieved from [http://www.oecd.org/document/40/0,3746,en\\_2649\\_39263231\\_40299432\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/40/0,3746,en_2649_39263231_40299432_1_1_1_1,00.html).
- OECD. (2010a). *PISA 2009 results: Executive summary*. Paris: OECD Publishing. Retrieved from <http://www.oecd.org/pisa/pisaproducts/46619703.pdf>.
- OECD. (2010b). *PISA 2009 results: Overcoming social background - Equity in learning opportunities and outcomes (Volume II)*. <http://dx.doi.org/10.1787/9789264091504-en>.
- OECD. (2010c). Viewing education in the United States through the prism of PISA *Strong performers and successful reformers in education: Lessons from PISA for the United States* (pp. 9-35). Paris: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/9789264096660-en>.
- OECD. (2011). *Against the odds: Disadvantaged students who succeeded in school*. <http://dx.doi.org/10.1787/9789264090873-en>.
- OECD. (2012a). *Equity and quality in education: Supporting disadvantaged students and schools*. <http://dx.doi.org/10.1787/9789264130852-en>.
- OECD. (2012b). *Pisa 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Paris: OECD Publishing. Retrieved from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf>.
- OECD. (2012c). *United States country notes - Results from PISA 2012*. Paris: OECD Publishing. Retrieved from <http://www.oecd.org/unitedstates/PISA-2012-results-US.pdf>.
- OECD. (2013). *Education policy outlook: Australia*. Paris: OECD Publishing. Retrieved from [http://www.oecd.org/edu/EDUCATION%20POLICY%20OUTLOOK%20AUSTRALIA\\_EN.pdf](http://www.oecd.org/edu/EDUCATION%20POLICY%20OUTLOOK%20AUSTRALIA_EN.pdf).
- Opening All Options. (n.d.). Learning difficulty versus learning disability Retrieved from <http://www.adcet.edu.au/Oao/view.aspx?id=4511>
- Orbell, S., Hodgkins, S., & Sheeran, P. (1997). Implementation intentions and the theory of planned behavior. *Personality and Social Psychology Bulletin*, 23(9), 945-954.
- Ouwueghuzie, A. J., & Johnson, B. (2006). The validity issue in mixed research. *Research in the Schools*, 13(1), 48-63.
- Pajares, F. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. *Journal of Educational Psychology*, 86(2), 193-203.

- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578.
- Phillips, D. (2000). Learning from elsewhere in education: Some perennial problems revisited with reference to British interest in Germany [Special number 23]. *Comparative Education, 36*(3), 297-307.
- Phye. (2008). Data collection methods: Potential pitfalls when implementing early reading programs as educational interventions. In S. K. Thurman & C. A. Fiorello (Eds.), *Applied cognitive research in K-3 classrooms* (pp. 189-211). New York, London: Routledge Taylor & Francis Group.
- Pierce, R., & Chick, H. (2011). Teachers' intentions to use national literacy and numeracy assessment data: A pilot study. *Australian Educational Researcher, 38*(4), 433-447.
- Pligt, J., & Vries, N. K. (1998). Belief importance in expectancy-value models of attitudes. *Journal of Applied Social Psychology, 28*(15), 1339-1354.
- Polesel, J., Dulfer, N., & Turnbull, M. (2012). *The experience of education: The impacts of high stakes testing on school students and their families*. Sydney, NSW: Whitlam Institute, University of Western Australia. Retrieved from [http://www.whitlam.org/data/assets/pdf\\_file/0010/409735/High Stakes Testing An Educators Perspective.pdf](http://www.whitlam.org/data/assets/pdf_file/0010/409735/High_Stakes_Testing_An_Educators_Perspective.pdf).
- Polikoff, M. S. (2012). Instructional alignment under No Child Left Behind. *American Journal of Education, 118*(3), 341-368.
- Pratt, J. (2003). A Popperian approach to policy research. In J. Swann & J. Pratt (Eds.), *Educational research in practice: Making sense of methodology* (pp. 51-66). London: Continuum.
- Pullin, D. (2005). When one size does not fit all: The special challenges of accountability testing for students with disabilities. *Yearbook of the National Society for the Study of Education, 104*(2), 199-222.
- Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin, 23*(4), 3-13.
- Raudenbush, S. W. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher, 34*(5), 25-31.
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choices are undermining education*. New York: Basic Books.
- Ready, D. D. (2010). Socioeconomic disadvantage, school attendance, and early cognitive development: The differential effects of school exposure. *Sociology of Education, 83*(4), 271-286.
- Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics, 92*(5-6), 1394-1415.
- Reeves, D. B. (2004). *Accountability for learning: How teachers and school leaders can take charge*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Reid, A. (2009). *Is this a revolution? A critical analysis of the Rudd government's national education agenda*. Paper presented at the Biennial Conference of ACSA on 2 October, 2009. [www.acsa.edu.au/pages/images/ACSA%20Boomer%20address.pdf](http://www.acsa.edu.au/pages/images/ACSA%20Boomer%20address.pdf)

- Reimers, F. (2010). Educating for global competency. In J. E. Cohen & M. B. Malin (Eds.), *International perspectives on the goals of universal basic and secondary education* (pp. 183-202). New York: Routledge, Taylor & Francis Group.
- Rizvi, F., & Lingard, B. (2000). Globalisation and education: Complexities and contingencies. *Educational Theory*, 50(4), 419-426.
- Ronka, D., Geier, R., & Marciniak, M. (2010). *A practical framework for building a data-driven district or school: How a focus on data quality, capacity, and culture supports data-driven action to improve student outcomes*. PCG Education
- Rorris, A., Weldon, P., Beavis, A., McKenzie, P., Bramich, M., & Deery, A. (2011). Assessment of current process of targeting schools funding to disadvantaged students. A report prepared for the Review of Funding for Schooling Panel. Retrieved from [http://research.acer.edu.au/policy\\_analysis\\_misc/10/](http://research.acer.edu.au/policy_analysis_misc/10/).
- Rose, L. C., & Gallup, A. M. (2006). The 38th Annual Phi Delta Kappa/Gallup Poll: Of the public's attitudes toward the public schools. *Phi Delta Kappan*, 88(1), 41-56.
- Rotter, J. B. (1954). *Social learning and clinical psychology*. New York: Prentice-Hall.
- Rowe, K. (2003). The importance of teacher quality as a key determinant of students' experiences and outcomes of schooling. *2003-Building Teacher Quality: What does the research tell us?*, 3. Retrieved from [http://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=research\\_conference\\_2003](http://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=research_conference_2003).
- Sachs, J. (2000). The activist professional. *Journal of Educational Change*, 1(1), 77-95.
- Sachs, J. (2001). Teacher professional identify: Competing discourses, competing outcomes. *Journal of Education Policy*, 16(2), 149-161.
- Sachs, J. (2016). Teacher professionalism: why are we still talking about it? *Teachers and Teaching*, 22(4), 413-425.
- Sagor, R. (2000). *Guiding school improvement with action research*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Sale, J. E. M., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality & Quantity*, 36(1), 43-53.
- Sammons, P. (2010). The contribution of mixed methods to recent research on educational effectiveness. In A. Tashakkori & C. Teddlie (Eds.), *Mixed methods in social & behavioral research* (pp. 697-723). Los Angeles: SAGE.
- Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. 9 Headlands Business Park, Ringwood, Hants BH24 3PB, England, United Kingdom: B & MBC Distribution Services.
- Sawchuk, S. (2010, March 9). NAEP board curbs exclusion of test-takers: More special-needs students, ELLs to be given assessment, *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2010/03/09/25nagb.h29.html>

- Schleicher, A. (2014). *Equity, excellence and inclusiveness in education: Policy lessons from around the world*. Paris: OECD Publishing. doi:10.1787/9789264214033-en.
- Schnellert, L. M., Butler, D. L., & Higginson, S. K. (2008). Co-constructors of data, co-constructors of meaning: Teacher professional development in an age of accountability. *Teaching and Teacher Education, 24*, 725-750.
- Schnepf, S. V. (2007). Immigrants' educational disadvantage: An examination across ten countries and three surveys. *Journal of Population Economics, 20*(3), 527-545.
- Schoen, L., & Fusarelli, L. D. (2008). Innovation, NCLB, and the fear factor: The challenge of leading 21st-century schools in an era of accountability. *Educational Policy, 22*(1), 181-203.
- Schriewer, J., & Marinez, C. (2004). Constructions of internationality in education. In G. Steiner-Khamsi (Ed.), *Global politics of educational borrowing and lending* (pp. 29-53). New York and London: Teachers College Press.
- Searle, J. R. (1995). *The construction of social reality*. New York, N.Y.: The Free Press.
- Shaughnessy, M. F. (2004). An interview with Anita Woolfolk: The educational psychology of teacher efficacy. *Educational Psychology Review, 16*(2), 153-176.
- Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*(1), 91-100.
- Shine, K. (2015). Are Australian teachers making the grade? A study of news coverage of NAPLAN testing. *Media International Australia, 154*(1), 25-33.
- Siegel, H. (2004). What ought to matter in public schooling: Judgment, standards, and responsible accountability. In K. A. Sirotnick (Ed.), *Holding accountability accountable: What ought to matter in public education* (pp. 51-65). New York, NY: Teachers College Press.
- Sirotnick, K. A. (2004). Introduction: Critical concerns about accountability concepts and practices. In K. A. Sirotnick (Ed.), *Holding accountability accountable: What ought to matter in public education* (pp. 1-17). New York, NY: Teachers College.
- Slavin, R. E. (1996). *Education for All*. Lisse: Swets & Zeitlinger Publishers.
- Slins, H. C., & Murray-Harvey, R. (2000). Students as a central concern: School students and outcome measures. *Journal of Education Administration, 38*(3), 230-246.
- Smith, E. (2005). Raising standards in American schools: The case of No Child Left Behind. *Journal of Education Policy, 20*(4), 507-524.
- Sparks, P., & Guthrie, C. A. (1998). Self-identity and the theory of planned behavior: A useful addition or an unhelpful artifice? *Journal of Applied Social Psychology, 28*(15), 1393-1410.
- Spelling, M. (2005). *No Child Left Behind: A road map for state implementation*. Retrieved from <https://www2.ed.gov/admins/lead/account/roadmap/roadmap.pdf>.
- Spring, J. (2009). *Globalisation of education*. New York: Routledge, Taylor & Francis Group.

- Sriprakash, A., & Proctor, H. (2013). Social class and inequality. In R. Connell, C. Campbell, M. Vickers, A. Welch, D. Foley & N. Bagnall (Eds.), *Education, change and society (3rd ed.)* (pp. 79-97). Melbourne: Oxford University Press.
- Stake, R. E. (2000). The case study method in social inquiry. In R. Gomm, M. Hammersley & P. Foster (Eds.), *Case Study Method* (pp. 18-27). London: SAGE Publications.
- Stake, R. E. (2010). *Qualitative research: Studying how things work*. New York: Guilford Press.
- State of Hawaii Department of Education. (2010). Overcoming challenges, 2010 superintendent's 21st annual report. Retrieved from [http://arch.k12.hi.us/PDFs/state/superintendent\\_report/2010/2010SuptRptFinal20110218.pdf](http://arch.k12.hi.us/PDFs/state/superintendent_report/2010/2010SuptRptFinal20110218.pdf).
- State of Hawaii Department of Education. (2011). Accountability Resource Center Hawaii. <http://arch.k12.hi.us/datacenter/adc.html>
- Steiner-Khamsi, G. (2004a). Globalisation in education: Real or imagined. In G. Steiner-Khamsi (Ed.), *Global politics of educational borrowing and lending* (pp. 1-11). New York and London: Teachers College Press.
- Steiner-Khamsi, G. (Ed.). (2004b). *The politics of education borrowing*. New York and London: Teachers College Press.
- Supovitz, J. A. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, 10, 211–227.
- Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. University of Pennsylvania: Consortium for Policy Research in Education.
- Sweller, N., Graham, L., & Van Bergen, P. (2012). The minority report: Disproportionate representation in Australia's largest education system. *Exceptional Children*, 79(1), 107-125.
- Systems Accountability Office. (n.d.). Strive HI performance system. Retrieved from <http://arch.k12.hi.us/school/strivehi/strivehi.html>
- Taylor, B. M., Pressley, M., & Pearson, D. (2000). *Effective teachers and schools: Trends across recent studies*. Retrieved from <http://files.eric.ed.gov/fulltext/ED450353.pdf>.
- Taylor, S., & Todd, P. (1995). Decomposition and crossover effects in the theory of planned behavior: A study of consumer adoption intentions. *International Journal of Research in Marketing*, 12(2), 137-155.
- Teddlie, C., & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Mixed methods in social & behavioral research* (pp. 1-41). Los Angeles: SAGE.
- The Honorable Julia Gillard Australian MP. (2010, May 5). Transcript of interview on the ABC's Lateline. Retrieved from <http://www.abc.net.au/lateline/content/2010/s2891489.htm>
- The Senate Standing Committee on Education and Employment. (2010). Inquiry into the administration and reporting of NAPLAN testing. *Parliament of Australia*. Retrieved from [http://www.aph.gov.au/Parliamentary\\_Business/Committees/Senate/Educate](http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Educate)



- [ion Employment and Workplace Relations/Completed inquiries/2010-13/naplan/index.](#)
- The Senate Standing Committee on Education and Employment. (2013). *Effectiveness of the national assessment program - literacy and numeracy*. Canberra: Senate Printing Unit, Parliament House.
- The Senate Standing Committee on Education and Employment. (2014). *Effectiveness of the national assessment program: Literacy and numeracy. Final report*. Canberra: Senate Printing Unit, Parliament House.
- The U.S. Office of Under Secretary. (2002). *No Child Left Behind: A desktop reference*. Washington, D.C.: US Department of Education.
- The United Nations. (2009). *Millennium goals development report 2009 (DPI/2539B)*. New York: Author. Retrieved from [http://www.un.org/millenniumgoals/pdf/PR\\_Global\\_MDG09\\_EN.pdf](http://www.un.org/millenniumgoals/pdf/PR_Global_MDG09_EN.pdf).
- The White House. (n.d.). *Reforming No Child Left Behind*. Washington DC: Author. Retrieved from <https://www.whitehouse.gov/issues/education/k-12/reforming-no-child-left-behind>.
- Thompson, G. (2013). NAPLAN, MySchool and accountability: Teacher perceptions of the effects of testing. *The International Education Journal: Comparative Perspectives*, 12 (2), 62–84.
- Thompson, G., & Harbaugh, A. G. (2013). A preliminary analysis of teacher perceptions of the effects of NAPLAN on pedagogy and curriculum. *Australian Educational Researcher*, 40, 299–314.
- Thomson, S., & De Bortoli, L. (2008). *Exploring scientific literacy: How Australia measures up, the PISA 2006 survey of students' scientific, reading and mathematical literacy skills*. Camberwell, VIC: ACER Press.
- Thurlow, M. L., & Johnson, D. R. (2000). High-stakes testing of students with disabilities. *Journal of Teacher Education*, 51(4), 305-314.
- Tolofari, S. (2005). New public management and education. *Policy Futures in Education*, 3(1), 75-89.
- Tooley, M. (Dec 24, 2015). No Child Left Behind is gone, but will it be back?, *The Atlantic*. Retrieved from <http://www.theatlantic.com/education/archive/2015/12/no-child-left-behind-is-gone-but-will-it-be-back/421329/>
- Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, 23(6), 944-956.
- Tschannen-Moran, M., Woolfolk Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202-248.
- Uncommon Schools. (n.d.). Frequently asked questions about public, charter schools. Retrieved from <http://www.uncommonschools.org/faq-what-is-charter-school>
- United States Census Bureau. (2015, June 2). Per pupil spending varies heavily across the United States, (Release Number: CB15-98). Retrieved from <https://www.census.gov/newsroom/press-releases/2015/cb15-98.html>
- US Department of Education. (2003). Questions and answers on No Child Left Behind. Retrieved from <http://www2.ed.gov/nclb/accountability/schools/accountability.html#4>

- US Department of Education. (2013). No child left behind: Early lessons from state flexibility waivers. Testimony of Secretary of Education Arne Duncan [Press release]. Retrieved from: <http://www.ed.gov/news/speeches/no-child-left-behind-early-lessons-state-flexibility-waivers>
- US Department of Education. (2015). Every Student Succeeds Act (ESSA). Retrieved from <http://www.ed.gov/essa?src=rn>
- US Department of Education. (n.d.). Budget Office—US Department of Education. Retrieved from <http://www2.ed.gov/about/overview/budget/index.html?src=gu>
- US Secretary of Education Ann Duncan. (2010). Unleashing the power of data for school reform: Secretary Arne Duncan's remarks at the STATS DC 2010 data conference [Speech]. Retrieved from <http://www.ed.gov/news/speeches/unleashing-power-data-school-reform-secretary-arne-duncans-remarks-stats-dc-2010-data-conference>
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research, 78*(4), 751-796.
- Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk, 10*(3), 295-308.
- Wayman, J. C., Cho, V., & Johnston, M. T. (2007). *The data-informed district: A district-wide evaluation of data use in the Natrona County School District*. Austin: The University of Texas.
- Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the data-informed district. *School Effectiveness and School Improvement, 23* (2), 159-178.
- Wayman, J. C., Midgley, S., & Stringfield, S. (2006). *Leadership for data-based decision-making: Collaborative educator teams*. Paper presented at the 2006 Annual Meeting of the American Educational Research Association, San Francisco CA.
- Wayman, J. C., Spring, S. D., Lemke, M. A., & Lehr, M. D. (2012). *Using data to inform practice: Effective principal leadership strategies*. Paper presented at the Annual meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.
- Wayman, J. C., & Stringfield, S. (2006a). Data use for school improvement: School practices and research perspectives. *American Journal of Education, 112*(4), 563-568.
- Wayman, J. C., & Stringfield, S. (2006b). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education, 112*(4), 549-571.
- Welch, A. (2001). Post-modernity and the state: Comparative education facing the third millennium. *Comparative Education, 37*(4), 475-492.
- Welch, A. (2010). Making education policy. In R. Connell, C. Campbell, M. Vickers, A. Welch, D. Foley, N. Bagnall & D. Hayes (Eds.), *Education, change and society* (Second ed., pp. 235-271). Victoria, Australia: Oxford University Press.

- Welch, A. (2011). The challenge of comparative research: A Critical introduction. In L. Markauskaite, P. Freebody & J. Irwin (Eds.), *Methodological choice and design* (pp. 187-208). London, New York: Springer.
- Welch, A. (2013). Making education policy. In R. Connell, C. Campbell, M. Vickers, A. Welch, D. Foley & N. Bagnall (Eds.), *Education, change and society (3rd ed.)* (pp. 186-212). Melbourne: Oxford University Press.
- Westwood, P. (2008). *What teachers need to know about learning difficulties*. Victoria, Australia: ACER Press.
- Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and choice in education. The school, the state and the market*. Scarborough, U.K.: ACER Press.
- Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues, 25*(4), 41-78.
- Wiersma, W. (2000). *Research methods in education: An introduction* (Seventh ed.). Boston: Allyn and Bacon.
- Wigglesworth, G., Simpson, J., & Loakes, D. (2011). NAPLAN language assessments for indigenous children in remote communities: Issues and problems. *Australian Review of Applied Linguistics, 34*(3), 320-343.
- Wöbmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2007). School accountability, autonomy, choice and the level of student achievement: International evidence from PISA 2003. *OECD Education Working Papers, No. 13*. Paris: OECD Publishing. doi: <http://dx.doi.org/10.1787/246402531617>.
- Wong, M., Cook, T. D., & Steiner, P. M. (2009). No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series. *Institute for Policy Research (Documento de trabajo 09-11), 18*.
- Wood, G. (2004). A view from the field: NCLB's effects on classrooms and schools. In D. Meier & G. Wood (Eds.), *Many children left behind* (pp. 33-50). Boston: Beacon Press.
- Wood, R. E., & Locke, E. A. (1987). The relation of self-efficacy and grade goals to academic performance. *Educational and Psychological Measurement, 47*(4), 1013-1024.
- Wyn, J., Turnbull, M., & Grimshaw, L. (2014). *The experience of education: The impacts of high stakes testing on school students and their families: A qualitative study*. The Whitlam Institute
- Yin, R. K. (1994). *Case study research, design and methods* (Second ed.). Thousand Oaks: SAGE Publications.
- Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*(4), 521-548.
- Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives, 18*(19), doi: <http://dx.doi.org/10.14507/epaa.v14518n14519.12010>.
- Yunupingu, G. (2008, December). Tradition, truth and tomorrow. *The Monthly*. Retrieved from <https://www.themonthly.com.au/issue/2008/december/1268179150/galarr-wuy-yunupingu/tradition-truth-tomorrow>

- Zhou, F. (2010). *Revenues and expenditures for public elementary and secondary education: School year 2007–08 (fiscal year 2008). First look*. NCES 2010-326. Washington, D.C.: National Center for Education Statistics. Retrieved from <http://files.eric.ed.gov/fulltext/ED509800.pdf>.
- Zimmerman, B. J. (1996, April). *Measuring and mismeasuring academic self-efficacy: Dimensions, problems and misconceptions*. Paper presented at the Symposium presented at the Annual Meeting of the American Educational Research Association, New York.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91.
- Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, 31(4), 845–862.
- Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663-676.
- Zymek, B., & Zymek, R. (2004). Traditional - National - International: explaining the inconsistency of educational borrowers. In D. Phillips & K. Ochs (Eds.), *Educational Policy Borrowing: historical perspectives* (pp. 7-23). Oxford, U.K.: Symposium Books.

## Appendices

### Appendix A Recruitment Letter



ABN 15 211 513 464

**Dr. Ilektra Spandagou**  
Senior Lecturer in Inclusive Education

**Faculty of Education and  
Social Work**

Room 705  
Education Building A35  
The University of Sydney  
NSW 2006 AUSTRALIA  
Telephone: +61 2 9351 6379  
Facsimile: +61 2 9351 2602  
Email:  
ilektra.spandagou@sydney.edu.au  
Web: <http://www.sydney.edu.au/>

**Assessment Data Use and its Impact on the Learning Outcomes of Students with Special Needs**

#### RECRUITMENT LETTER – (Principals)

Dear (Mr. or Mrs. name of principal):

I write to invite you and your teachers to participate in a research project exploring assessment data engagement at schools and in the classrooms. This study will compare and contrast the way schools utilise assessment data in Australia and in the USA. We hope to interview you, your special resource teacher and teachers in Years 3 - 5 to learn about data practices at your school. We are also interested in understanding how these data practices have contributed to the learning outcomes of your students.

The Participation Information Statement attached provides all necessary information regarding the research objectives and procedures. We hope that it will clarify any questions you may have to make an informed decision regarding your participation.

My associate or I will call your office in a few days to follow-up on any questions you may have about the project.

Thank you for your time and I hope to have an opportunity to learn about data practices at your school.

Sincerely Yours,

Dr. Ilektra Spandagou  
Senior Lecturer of Inclusive Education  
Faculty of Education and Social Work  
University of Sydney

## Appendix B Participant Information Statement



THE UNIVERSITY OF  
**SYDNEY**

Faculty of Education and  
Social Work

ABN 15 211 513 464

**Dr. Ilektra Spandagou**  
Senior Lecturer in Inclusive Education

Room 705  
Education Building A35  
The University of Sydney  
NSW 2006 AUSTRALIA  
Telephone: +61 2 9351 6379  
Facsimile: +61 2 9351 2602  
Email:  
ilektra.spandagou@sydney.edu.au  
Web: <http://www.sydney.edu.au>

### Assessment Data Use and its Impact on the Learning Outcomes of Students with Special Needs

#### PARTICIPANT INFORMATION STATEMENT – (Principal\_AU)

**(1) What is the study about?**

This study will explore how school leaders and teachers in Australia and in the USA use standardised and other test data to make programming and instructional decisions to raise the learning outcomes of students with special needs. These students may include those who have a disability, have a learning difficulty, and are English language learners. Using results from standardised tests, beginning and end of term tests and other interim or alternative assessments, this study will examine the relationship between data practices and student achievement.

**(2) Who is carrying out the study?**

The study is led by Dr. Ilektra Spandagou, Senior Lecturer in Inclusive Education at the University of Sydney. The research associates are Dr. David Evans, Associate Professor of Special Education and Nicole Bien, Ph.D. student in the Faculty of Education at the University of Sydney.

**(3) What does the study involve?**

The investigators hope to learn about data practices through semi-structured one-on-one interviews with school leaders, special resource teachers and Years 3 - 5 classroom teachers. These open-ended interviews explore the role of assessment data in setting school, class, and student goals; and in making curriculum, programming and instructional decisions.

The study will interview school leaders and teachers in two primary schools each in Sydney, Australia and California, USA. The interview topics include: data culture, perception of data, data use in curriculum, program and instructional decisions, and teacher expectations. The interviews will take place at the participating schools. The researchers will seek individual participant consent from teachers.

The views of all participants and records of the participating schools and students will be kept confidential. Pseudonyms will be used to reference all participating schools and individuals.

In addition, the researchers might seek information regarding school demographic profiles and copies of program description, data analytics and other applicable documentation that might surface through the interview process. No personally identifiable data will be collected at any point in time.

**(4) How much time will the study take?**

Each interview is expected to take about 30 minutes.

**(5) Will the interview be recorded?**

Yes, it will be audio-taped for analysis purposes. The tapes will be securely locked in the office of the chief investigator and accessible only by the researchers in this study.

**(6) Can I withdraw from the study?**

Participant Information Statement (Continued)

Being in this study is completely voluntary. You are not under any obligation to consent and if you do consent, you can withdraw at any time without affecting your relationship with The University of Sydney.

You may stop the interview at any time if you do not wish to continue, the audio recording will be erased and the information provided will not be included in the study.

**(7) Will anyone else know the results?**

All aspects of the study, including results, will be strictly confidential and only the researchers will have access to the information collected. The results will form the basis of a Ph.D. thesis. In addition, a summary report of the study may be submitted for journal publication or presented at an academic conference. However, individual participants will not be identified. Instead, pseudonyms will be used in these reports.

**(8) Will the study benefit me?**

This research may illuminate different data practices in each country and at each school given cultural, environmental and policy differences in each context. These differences may be considered as feedback to the participating school community. They may also offer participants a point of reflection for their own data practices in helping students achieve minimum or higher standards.

**(9) Can I tell other people about the study?**

Yes, you can.

**(10) What if I require further information about the study or my involvement in it?**

When you have read this information, the research associate, Nicole Bien will discuss it with you further and answer any questions you may have. If you would like to know more at any stage, please feel free to contact Nicole Bien, Ph.D. Student at [nbie6288@uni.sydney.edu.au](mailto:nbie6288@uni.sydney.edu.au) (Email), or Dr. Ilektra Spandagou, Senior Lecturer at the University of Sydney, +61 2 9351 6379 (Telephone) or [ilektra.spandagou@sydney.edu.au](mailto:ilektra.spandagou@sydney.edu.au) (Email).

**(11) What if I have a complaint or any concerns?**

Any person with concerns or complaints about the conduct of a research study can contact The Manager, Human Ethics Administration, University of Sydney on +61 2 8627 8176 (Telephone); +61 2 8627 8177 (Facsimile) or [ro.humanethics@sydney.edu.au](mailto:ro.humanethics@sydney.edu.au) (Email).

*This information sheet is for you to keep*

## Appendix C Participant Consent Form



THE UNIVERSITY OF  
**SYDNEY**

**Faculty of Education and  
Social Work**

ABN 15 211 513 464

**Dr. Ilektra Spandagou**  
*Senior Lecturer in Inclusive Education*

Room 705  
Education Building A35  
The University of Sydney  
NSW 2006 AUSTRALIA  
Telephone: +61 2 9351 6379  
Facsimile: +61 2 9351 2602  
Email:  
ilektra.spandagou@sydney.edu.au  
Web: <http://www.sydney.edu.au/>

### PARTICIPANT CONSENT FORM (All Participants)

I, .....[PRINT NAME], give consent to my participation in the research project

**TITLE: Assessment Data Use and its Impact on the Learning Outcomes of Students with Special Needs**

In giving my consent I acknowledge that:

1. The procedures required for the project and the time involved have been explained to me, and any questions I have about the project have been answered to my satisfaction.
2. I have read the Participant Information Statement and have been given the opportunity to discuss the information and my involvement in the project with the researcher/s.
3. I understand that I can withdraw from the study at any time, without affecting my relationship with the researcher(s) or the University of Sydney now or in the future.
4. I understand that my involvement is strictly confidential and no information about me will be used in any way that reveals my identity.
5. I understand that being in this study is completely voluntary – I am not under any obligation to consent.



Participant Consent Form (Continued)

6. I understand that I can stop the interview at any time if I do not wish to continue, the audio recording will be erased and the information provided will not be included in the study.

7. I consent to:

- |                        |     |                          |    |                          |
|------------------------|-----|--------------------------|----|--------------------------|
| i) Audio-taping        | YES | <input type="checkbox"/> | NO | <input type="checkbox"/> |
| ii) Receiving Feedback | YES | <input type="checkbox"/> | NO | <input type="checkbox"/> |

If you answered YES to the "Receiving Feedback" question, please provide your details i.e. mailing address, email address.

**Feedback Option**

Address: \_\_\_\_\_


Email: \_\_\_\_\_

Signed:.....

Name: .....

Date: .....

## Appendix D Interview Guide



**THE UNIVERSITY OF SYDNEY**

ABN 15 211 513 464

**Faculty of Education and Social Work**

---

**Dr. Ilektra Spandagou**  
Senior Lecturer in Inclusive Education

Room 705  
Education Building A35  
The University of Sydney  
NSW 2006 AUSTRALIA  
Telephone: +61 2 9351 6379  
Facsimile: +61 2 9351 2602  
Email:  
ilektra.spandagou@sydney.edu.au  
Web: <http://www.sydney.edu.au/>

**Assessment Data Use and its Impact on the Learning Outcomes of Students with Special Needs**

**INTERVIEW GUIDE (Semi-Structured)**

Thank you for agreeing to participate in this interview. It will take approximately 30 minutes. I will ask you a series of guiding questions, which have no right or wrong answers. Your answers are strictly confidential and will be kept anonymous. Your participation is voluntary and you can withdraw from this interview at any time. Finally, I will tape-record this interview for data analysis and it will be kept securely at the University of Sydney.

**A. Background Information**

1. School Name: \_\_\_\_\_
2. Participant Gender:
  - a. Female..... 1
  - b. Male..... 2
3. Specialisation:
  - a. Principal or vice principal ..... 1
  - b. Special resource teacher.....2
  - c. Third grade class teacher.....3
  - d. Fourth grade class teacher..... 4
  - e. Fifth grade class teacher.....5
  - f. Sixth grade class teacher.....6
  - g. Other..... 7
4. Years at current school: \_\_\_\_\_
5. Years teaching current grade level at your current school: \_\_\_\_\_
6. Years as an educator: \_\_\_\_\_

**B. Interview**

1. Goal Setting and Definition for Success:
  - a. What are the academic targets in relations to learning for students with special learning needs (students with a disability, English language learners, indigenous students), How are these targets guided and set and who sets them?
  - b. What are the indicators used to measure success for each student subgroup and how often are they evaluated?
  - c. Do these students participate in the NAPLAN?
2. Data Engagement:
  - a. In general, what type of assessment data is used to diagnose the learning needs and inform programming decisions for students with special needs?

Assessment Data Use and Learning Outcomes  
Version 3, AU, July 2012

Page 1 of 2

- b. How do you and your school use assessment data to align instructional programs or decisions with the learning goals discussed above?
- c. Please discuss the usefulness of each assessment type available to you?
- d. Is there a regular forum where teachers and administrators engage with data?
- e. What expectations do you have from data engagement?
  - i. Are they meant to be instructional, evaluative or predictive?
- f. Please provide examples of how data has been applied to support programs directly related to improving the academic results of students with special education needs?
- g. Do you have a sample of a data dashboard that includes analysis of data, student targets, and associated actions? Are you (teachers and school) held accountable for the targets you have set or are they there to serve as a guiding post?
- h. Are there insights about students that you believe could benefit programming and instructional efforts that are not within the scope of the current assessment data?
- i. What are some challenges in general with data-focused practices?

**3. Collaboration:**

- a. How do teachers collaborate with one another and the special resource or special education teachers? Who is ultimately responsible for the achievements of students with special needs?
- b. What happens when a group of students consistently do not progress?
- c. How do you make use of the insights from data to engage or collaborate with parents?

**4. Impact:**

- a. What added value do you see from data-driven practice in relations to the subgroup of students who perform below proficiency level?
- b. What benefits do data-driven practices have on the way teachers work together and separately to improve the achievement of students especially those with diverse learning needs?
- c. Do you feel empowered or restricted by data? Why?

Appendix E Colour-code Worksheet of Student Achievement at Almond

**Class List Report for:**  
**Exam: ELA - Multiple Choice 1 of 3 - Grade 5 (Adm1-2011/2012)**

Page 1 of 8

October 31, 2011

**Report Options**

Schools: All  
 Grades: All  
 Teachers: All  
 Courses: All  
 Gender: All  
 Ed Programs: All

Ethnicities: All  
 Custom Groups: All  
 Roster: 2011-2012 Fall, School Year  
 Test Delivery: All  
 # Students: 28

**Student Performance**

Student	Overall Performance	Grade Five Reading: 1.3 (Section 1 - Multiple Choice: 5, 8, 25) (3 pts. possible)	Grade Five Reading: 1.5 (Section 1 - Multiple Choice: 4, 7, 28) (3 pts. possible)	Grade Five Reading: 2.3 (Section 1 - Multiple Choice: 10, 11, 13) (3 pts. possible)	Grade Five Reading: 2.4 (Section 1 - Multiple Choice: 9, 12, 14) (3 pts. possible)	Grade Five Reading: 3.2 (Section 1 - Multiple Choice: 6, 17, 19) (3 pts. possible)
<b>GROUP AVERAGE</b>	<b>Approaching 23.5(83.33%)</b>	<b>Below 1.5(33.33%)</b>	<b>At or Above 2.0(66.67%)</b>	<b>Below 1.0(33.33%)</b>	<b>At or Above 2.0(66.67%)</b>	<b>At or Above 2.1(71.43%)</b>
1	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
2	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
3	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
4	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
5	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
6	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
7	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
8	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
9	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
10	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
11	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
12	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
13	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
14	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
15	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
16	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
17	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
18	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
19	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
20	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
21	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
22	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
23	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
24	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
25	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
26	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
27	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)
28	Approaching 23.5(83.33%)	Below 1.5(33.33%)	At or Above 2.0(66.67%)	Below 1.0(33.33%)	At or Above 2.0(66.67%)	At or Above 2.1(71.43%)

**Appendix F Pacing Guide at Kukui**


School						
At-a-Glance Pacing Map for Second Grade – October 2011						
Benchmark/Skill	Lesson/Materials	Monday	Tuesday	Wednesday	Thursday	Friday
<b>LA Standards</b>	<b>Trophies</b> Mr. Putter and Tabby Fly The Plane <b>Focus Skill</b>	<b>26</b> Day 1 1.1	<b>27</b> Day 2 1.2	<b>28</b> Day 3 1.3 SOQ Assembly	<b>29</b> <b>GLAD</b>	<b>30</b> Day 4 1.4/1.5 Q1 Ends School Fair 10/1
<b>Math Standards</b> 2.G.1	<b>Investigations</b> Unit 2 Shapes, Blocks, and Symmetry Unit 2 Session 1- Features 2D/3D Shapes 1.1- Geoblock Faces 1.2- Working with geoblocks 1.3- Sorting geoblocks 1.4- Addition combo and ways to fill 1.5- Combing shapes					
<b>LA Standards</b>	<b>Trophies</b> Hedgehog Bakes a Cake <b>Focus Skill</b>	<b>10</b> Day 1 2.1	<b>11</b> Day 2 2.2	<b>12</b> Day 3 2.3	<b>13</b> Day 4 2.4	<b>14</b> Day 5 2.5
<b>Math Standards</b> 2.G.1	<b>Investigations</b> Unit 2 Session 2 What Is a Rectangle 2.1- Assessment: Sorting shapes by # of sides 2.2- Sorting Quadrilaterals 2.3- Ordering rectangles 2.4- Building rectangles 2.5- Drawing rectangles					
<b>LA Standards</b>	<b>Trophies</b> Hedgehog Bakes A Cake <b>Focus Skill</b>	<b>17</b> Day 1 2.6	<b>18</b> Day 2 2.7	<b>19</b> Day 3 2.8	<b>20</b> Day 4 2.9	<b>21</b> Day 5 3.1
<b>Math Standards</b> 2.G.1	<b>Investigations</b> Unit 2 Session 2 2.6- How Many Rectangles? 2.7- Assessment: Is it a rectangle? 2.8- Making boxes from rectangles 2.9- Assessment: Faces of a geoblock Unit 2 Session 3 Symmetry 3.1- Mirror Symmetry					
<b>LA Standards</b>	<b>Trophies</b> Lemonade For Sale <b>Focus Skill</b>	<b>24</b> Day 1 3.2	<b>25</b> Day 2 3.3	<b>26</b> Day 3 3.4	<b>27</b> Day 4 3.5	<b>28</b> Day 5 1.1
<b>Math Standards</b> 2.G.3	<b>Investigations</b> Unit 2 Session 3- Symmetry 3.2- Copy tiles 3.3- Paper folding and Cutting 3.4- Is it Symmetrical? 3.5- End of Unit Assessment Unit 3 Stickers, Number strings, and story problems Unit 3 Session 1 1.1- Does order Matter?					

Appendix G Data Team Process at Hibiscus

ry School

**Data Team Process: Review, Reflect, & Respond**

*Follow the Steps to Complete the W.E.S. Data Analysis Protocol Sheet*

REVIEW	REFLECT	RESPOND
<p><b>Step 1: Collect the Data</b></p> <p>Before the Meeting: Collate Classroom data</p> <p>Complete the data collection sheet and compute the totals, averages, percentages for your class</p> <p><b>Step 2: Chart &amp; Analyze the Data</b></p> <p>Chart your data on the grade level grid</p> <p>Assist the Statistician in computing totals, averages, and percentages</p> <p>Examine the data display and record your thoughts regarding: Strengths, Challenges, and Trends</p>	<p><b>Step 3: Establish a Common Goal</b></p> <p>Based on the Team analysis, create, or review/revise a SMART Goal to address an area of need</p> <p>(Specific, Measurable, Achievable, Relevant, Time Bound)</p> 	<p><b>Step 4: Select Instructional Strategies</b></p> <p>Identify high-yield strategies that will result in student achievement gains. Agree on one or two strategies that everyone will try</p> <p><b>Step 5: Determine Results Indicators</b></p> <p>Determine what is the observable evidence (besides the next assessment) that will indicate strategies are being implemented and learning is taking place</p> <p>Decide on a date for the next assessment and date for the next analysis meeting</p>
<p><b>After the Meeting:</b> Implement selected strategies. Collect evidence for results indicators. Administer the next assessment. Bring data to the next meeting</p>		

**Appendix H Learning Community Meeting Protocol at Hibiscus**

**Teacher Learning Community Meetings**

Location: Production Room



Date	Grade	Time
August 15, 2011	3	9:35 a.m. to 11:00 a.m.
	K	11:30 a.m. to 1:10 p.m.
August 16, 2011	4	9:35 a.m. to 11:00 a.m.
	1	11:30 a.m. to 1:10 p.m.
August 18, 2011	5	9:35 a.m. to 11:00 a.m.
	2	11:30 a.m. to 1:10 p.m.

**AGENDA**

What	Why	How
Welcome / Celebrations	As we share celebrations with one another and participate in inclusion activities, we strengthen the foundation of our TEAM.	Welcome / Celebrations ☺
Academic Plan Revisions and Priorities	To build and understanding of how specific Academic Plan Enabling Activities connect to TLC activities	• Review Revisions to 2011-12 Academic Plan
Overview of Progress Monitoring Data System	To build an understanding of the process and protocol that will be used in the Progress Monitoring Data System	• Overview of Progress Monitoring & Data Teams • Introduce Data Analysis Process Cycle • Introduce Data Analysis Protocol • Complete p. 1 of the Protocol
SMART GOAL	To build an understanding of SMART Goals and collaborate on writing a goal for Reading Street assessment data	• Introduce components of a SMART Goal • Use page 1 of the Data Analysis Protocol to collaboratively write a SMART Goal for RS assessment data

**Things to Bring:**

Reading Street Baseline Results (Grades 2, 3, 4, 5)  
Reading Street Groupings based on DIBELS (Grades K & 1)

Your Smile! ☺

**Appendix I Instructional Leadership Team Meeting Agenda at Kukui**

Type of Meeting: <u>Team Meeting</u>		Date: <u>8/30/11</u>		Time: <u>2:30-3:30</u>	
Location: <u>Conference Room Office</u>		Notes: _____		Person Responsible: _____	
Desired Outcome: _____		Facilitator: _____		_____	
Topic	ID/A	Desired Outcome	Person Responsible	Time	
TMT	I	Quote: Someone's sitting in the shade today because someone planted a tree a long time ago. Summarizer: Ana Marie		2:30-2:40	
Data	A	Id # of students needed to make AYP using the AYP calculator		2:40-2:50	
Data	A	Complete Safe harbor calculator		2:50-3:00	
Data	A	Update and Identify Full Academic Year students/subgroups		3:00-3:10	
Communication	I	HSA goals identified/posted throughout the school		3:10-3:15	
Data	I	Review 1 <sup>st</sup> month BM testing scores Jane: math Tina: reading		3:15-3:20	
Planning	A	Teachers begin to create custom groups on the BM system Plan & train new teachers		3:20-3:30	



**Appendix J Evaluation of Student Performance and Action Plan Part I at**

**Almond**

<p><b>2011-12</b></p>	<p><b>Data CONFERENCE TEMPLATE</b></p>		
<p>Teacher: <u>          </u></p>	<p>Date <u>11/1/11</u></p>		
<p><b>Objective:</b></p> <ul style="list-style-type: none"> <li>• To use data and standards to construct strategic goals and action plans for student learning</li> </ul>			
<p><b>Outcomes:</b></p> <ul style="list-style-type: none"> <li>• To review standards needing re-teaching and preview new key standards</li> <li>• To identify students needing challenge, support, and intervention</li> <li>• To align teachers, principal, and support staff on goals, monitoring, action plan and next steps.</li> </ul>			
<p><b>Materials Needed</b></p> <ul style="list-style-type: none"> <li>• Classroom data <i>Note: this includes but is not limited to benchmark assessment data.</i></li> <li>• Grade level standards for the content area</li> <li>• Standards pacing guide and assessment blueprint and/or copy of the next assessment</li> </ul>			
<p><b>Standards-based goals</b>                  Reference <b>data</b> to determine standards need to be re-taught to the whole class.                  Reference the <b>pacing guide</b> to determine what standards you will be teaching next.                  This will inform your planning and support you in sharing goals with your partner teacher(s).</p>			
<p><b>Whole class standards for re-teaching</b></p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none;"> <p>prepositional phrases</p> <p>appropriate conjunctions</p> <p>combining sentences/revision</p> </td> <td style="width: 50%; border: none;"> <p>author's choice of words</p> <p>synonym/antonym</p> </td> </tr> </table>		<p>prepositional phrases</p> <p>appropriate conjunctions</p> <p>combining sentences/revision</p>	<p>author's choice of words</p> <p>synonym/antonym</p>
<p>prepositional phrases</p> <p>appropriate conjunctions</p> <p>combining sentences/revision</p>	<p>author's choice of words</p> <p>synonym/antonym</p>		
<p><b>Analysis of why students didn't learn standard</b></p> <ul style="list-style-type: none"> <li>* insufficient practice</li> <li>* misunderstanding the question</li> <li>* not looking at meaning/context, missing the big picture</li> </ul>			
<p><b>Instructional Plan – What techniques will you use to address these standards?</b></p> <ul style="list-style-type: none"> <li>* daily skill-based practice</li> <li>* more emphasis on words in context getting the main idea/theme/point</li> </ul>			

**Appendix K Evaluation of Student Performance and Action Plan Part II**

at Almond

↓

**Student-based goals**  
Specific groups of students with similar strengths/needs:

	<b>Benchmark or above</b> At grade level/needs challenge	<b>Strategic</b> Specific skills needed to move into P/A	<b>Intervention</b> Acceleration needed to move out of low levels
<b>Who are the students?</b>	Eduardo John Fernando Maya Jose Brenda Grace	Bassel Sarah Michael Roman Rikelm <del>Julia</del> Julia	? Julia? Kristen Cis. K Christopher Isael an
<b>What are their needs?</b> content cluster strands, skills, standards, etc.	Independent inquiry and application Research Presentations sophisticated language Revision	Develop and build on ideas Increased sophistication Main idea/point Author's techniques	Basic skills & work habits Producing authentic language
<b>What is your instructional plan?</b>  <b>How and when will you structure small group intervention?</b>	Lit Circles Ind. Projects	Group Reading Skill practice Greater independence	Guided reading Independent reading/responding Skill practice
<b>What support do you need from support staff and your team teacher(s)?</b>		Ms P ← Desiray Eric Christina Cis.	Ms. P ← Juan Isael

**Teacher Reflection**  
What next steps will you take as a result of this data analysis?

- Daily skill-based practice
- Increase group/guided reading/responding
- Work on authors' techniques ← words → revision  
- context → fig. lang. → 420: 25