EVALUATION OF THE FLEXIBLE LITERACY FOR REMOTE PRIMARY SCHOOLS PROGRAM

MAIN REPORT

2015&2016 SCHOOL YEARS

Prepared for:

The Commonwealth Department of Education and Training

Investigators:

Professor Janet Clinton

Georgia Dawson

Project Team:

Patrick Mclaren

Myra Koelle

Lee Kiong Au

Katla Hower

Esther Kee

Date: 24th of July 2017

ISBN: 978-1-76051-256-9 [PDF]
ISBN: 978-1-76051-257-6 [DOCX]

This report was produced by The University of Melbourne Centre for Program Evaluation on behalf of the Australian Government. Funding was provided by the Australian Government Department of Education and Training.

Citation: Clinton, J., Dawson, G.K., McLaren, P., & Koelle, M. (2017). Evaluation of the Flexible Literacy for Remote Schools Program: Main Report. July 2017. Centre for Program Evaluation, the University of Melbourne.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CPE	Centre for Program Evaluation
CYAAA	Cape York Aboriginal Australian Academy
DI	Direct Instruction as specified by NIFDI
DET	Australian Government Department of Education and Training
EDI	Explicit Direct Instruction
EYLND	Early Years Literacy and Numeracy Data
FLFRPSP	Flexible Literacy for Remote Primary Schools Program (also referred to as 'the program')
FTE	Full-time equivalent
GGSA	Good to Great Schools Australia
ICSEA	Index of Community Socio-Educational Advantage
IND	Independent Schools
LBOTE	Language Backgrounds Other Than English
NAPLAN	National Assessment Program Literacy and Numeracy
NT	Northern Territory
NIFDI	National Institute for Direct Instruction
QLD	Queensland
UWA	University of Western Australia
WA	Western Australia
Remote	Covers Remote and Very Remote Schools

EXECUTIVE SUMMARY

BACKGROUND AND PURPOSE

The Flexible Literacy for Remote Primary Schools Program (FLFRPSP) was a policy response to low literacy performance of students living in remote and very remote Australian communities. In developing the program, it was recognised that students in remote and very remote schools faced diverse barriers to learning that were appreciably different from students in metropolitan or regional schools.

In 2014, Good to Great Schools Australia (GGSA) was contracted by the Commonwealth government to implement the FLFRPSP based on the principles of NIFDI's Direct Instruction (DI)[1] or Explicit Direct Instruction (EDI). The program's primary objectives are to:

1. Improve students' literacy abilities and results
2. Increase teacher pedagogical skills in teaching literacy through the use of alphabetic teaching approaches, in particular, DI/EDI

GGSA was responsible for introducing and embedding the pedagogical teaching approaches, EDI or DI, in remote primary schools in NT, QLD, and WA. In 2015, GGSA began working with 33 schools. As at July 2017, there are 35 schools in the program.

In 2015, following an open tender process, the Centre for Program Evaluation (CPE) was contracted by GGSA to quantitatively analyse available data on FLFRPSP to understand its progress in implementation and initial performance indicators, as well as to help GGSA build its internal evaluation capacity. In late 2016 management of the contract moved to the Commonwealth DET and the evaluation scope was broadened.

THEORY AND METHODS

There are two key and inter-related aspects to the methodology driving this evaluation. First, FLFRPSP is educational reform in heterogeneous and complex contexts, and this complexity needs to be factored in to what data is collected, how this data is collected, and how this data is analysed and interpreted. Second, the evaluation takes a longitudinal view, and in so doing considers the data examined in this report as having utility in the current discussion regarding the program, but also having utility as the foundation for ongoing monitoring and evaluation of the program over time. These two concepts are inter-related because educational reform in any complex environment takes time and deserves time. Evaluation plays a critical role in the provision of this time by future-proofing monitoring systems and developing measurement models that are sustainable and where possible standardised so that relevant analysis can be compared over time and in a meaningful way.

The overarching principle of this evaluation is to provide credible information to DET regarding the impact of FLFRPSP on students' literacy outcomes and teacher pedagogical skills. The evaluation utilises a mixed-method convergent parallel design in the collection of data. It is convergent in that it utilises both quantitative data and qualitative data to gain a robust understanding of the program impacts, including drivers and inhibitors of program success for all key stakeholders.

A summary of the specific types of data is presented in Table 1 below.

[1] "DI" in this document will refer to Direct Instruction as specified by NIFDI; this is in contrast to direct instruction as a pedagogy, referred to in this document in full using lowercase letters.

Table 1: ***Specific Data Types and its Uses in Evaluation***

Data	Description	Use In Evaluation
FLFRPSP Program-Based Data	Range of data sets related to student placement in the program, lesson progress, student mastery, records of program activities such as coaching sessions and training, and implementation fidelity indicators	Several indicators of program implementation, dosage, fidelity, and student progress. In this context, fidelity considers the quality of program delivery.
State/Jurisdictional Literacy Data	Early Years Literacy and Numeracy Data (EYLND) – WA Catholic Education Office. PM Benchmark Reading – WA	An indicator of impact of program on literacy outcomes considered proximal to classroom learning
National Assessment Program for Literacy and Numeracy	Annual assessment for students in years 3, 5, 7, and 9. Tests utilised in this evaluation were Reading, Writing, Spelling, Grammar and Punctuation	Potential indicator of the impact of program on literacy outcomes considered more general than state/jurisdictional based literacy measures
Teacher Survey	Survey questions pertaining to teacher perceptions of their interactions with the program and the programs impact	Stakeholder perceptions to be triangulated with quantitative program and literacy achievement data
Secondary Document Analysis	Publicly available reports, information from websites, press releases	Secondary indicators or support for other indicators of student progress, teacher experience, school experience with the program

KEY FINDINGS

NORTHERN TERRITORY

For **students** in NT Government schools there are currently mixed results of early impact. Additional time and data is required to fully understand the programs influence, with NT jurisdictional literacy absent from this analysis. Nonetheless, the results demonstrated that program schools have not regressed on NAPLAN scores over time. As a highlight, in Grammar and Punctuation, the results of lower level students increased towards higher performing students. Furthermore, the individual school analyses showed that for Spelling, Writing and Grammar, and Punctuation, mean change in NAPLAN scores over time were greater than the national average mean score change.

Teacher survey data and Principal feedback on student impact shared the same mixed valence as the statistical indicators, suggesting that some schools are seeing quite significant positive impact from the program on student literacy outcomes and others are not. At the **teacher** and **school** level, turnover is a significant problem, and one that encroaches on implementation and fidelity. Nonetheless, teachers report being well prepared to teach the program and understand the pedagogical approach following initial training. Program data measuring overall school fidelity against four domains indicate that schools are receiving excellent levels of support from GGSA, whereas teacher readiness, a measure of staff being trained, and other activities require further refinements and improvement.

WESTERN AUSTRALIA (GOVERNMENT)

Similarly, there are currently mixed results regarding the program's early impact on **students** in WA Government schools.

More complete literacy data at the jurisdictional level will assist in better understanding the program effects. As would be expected at this point in time, the NAPLAN analysis does not indicate any current impacts from

the program. Nonetheless, there are pockets of promise: steady gains from year 3 to year 5, reduction in the percentage of students below the NMS in several domains, and some schools achieving gains greater than the State average. Furthermore, teacher survey data and Principal feedback on student impact suggests that schools perceive that the program is having a positive effect on student literacy.

At the **teacher** and **school** level, high turnover rates of teaching staff interrupt the program being implemented with a high degree of fidelity. However, teachers report being well prepared to teach the program and have a solid understanding of the program following the training. The program data assessing overall school fidelity suggests that schools receive high levels of support from GGSA, with high numbers of teacher training and observations.

WESTERN AUSTRALIA (CATHOLIC)

Results from WA Catholic education are consistent and robust, and the triangulated evidence from WA suggests that change is occurring in literacy. The jurisdictional literacy data, EYLND, shows early signs of positive impacts from the program. Program schools showed greater gains than control schools (measured as effect size), and 2016 EYLND levels were significantly higher than control schools. The NAPLAN analysis shows that program schools were higher than control schools in mean NAPLAN scores and gains, with the Spelling domain demonstrating a significant difference. The percentage of students below NMS reduced in all domains for Year 3 and in Writing and Grammar and Punctuation for Year 5. Mean gain in NAPLAN scores were higher than the State average for all schools except one across Spelling, Grammar and Punctuation, and Writing domains.

Teacher survey data and Principal feedback on student impact suggested that teachers perceive the program as having a positive effect on student literacy and that they are engaged with the program. At the **teacher** and **school** level, turnover is lower than NT and WA. Interestingly, WA Catholic Education demonstrated higher teacher readiness scores within the school fidelity measure, which may be influenced by smaller turnover rates. Teacher effectiveness measures showed a large initial increase and then stabilisation at a high level. Teachers also reported being well prepared to teach the program and understand the pedagogical approach following initial training. Program data suggested that schools have consistently been implementing the program with a high degree of fidelity.

OVERALL

The FLRPSP appears to be on track to influence the learning lives of students in remote areas of Northern and Western Australia. There is little doubt that the program is having an impact on the literacy levels in participating program schools. Where available, evidence suggests that schools engaged in the program are demonstrating increases in literacy levels across standardized jurisdictional literacy measures. Although currently NAPLAN results for Reading, Writing, Spelling, and Grammar and Punctuation show no statistically significant difference between all control and program schools, the current trajectory for future impact is a positive one.

Education in remote areas of Australia is both complex and complicated. Student attendance, teacher turnover, leadership turnover, and resourcing are key factors that affect not only the day-to-day running of schools but also any initiative being implemented. Almost every teacher and principal who responded to the survey or interview raised the issue of poor student attendance. Teachers and principals both described the program as rigid and inflexible. However, they also acknowledged that it is the rigidity and structure of the program that allows students with irregular attendance to re-enter the program, and more importantly re-enter the learning. In addition, the structure appears to ameliorate high teacher turnover as well. In this

respect, the nature of the program acts as a protective factor for the system wide issues of student attendance and teacher turnover.

PROJECTIONS FOR THE FUTURE

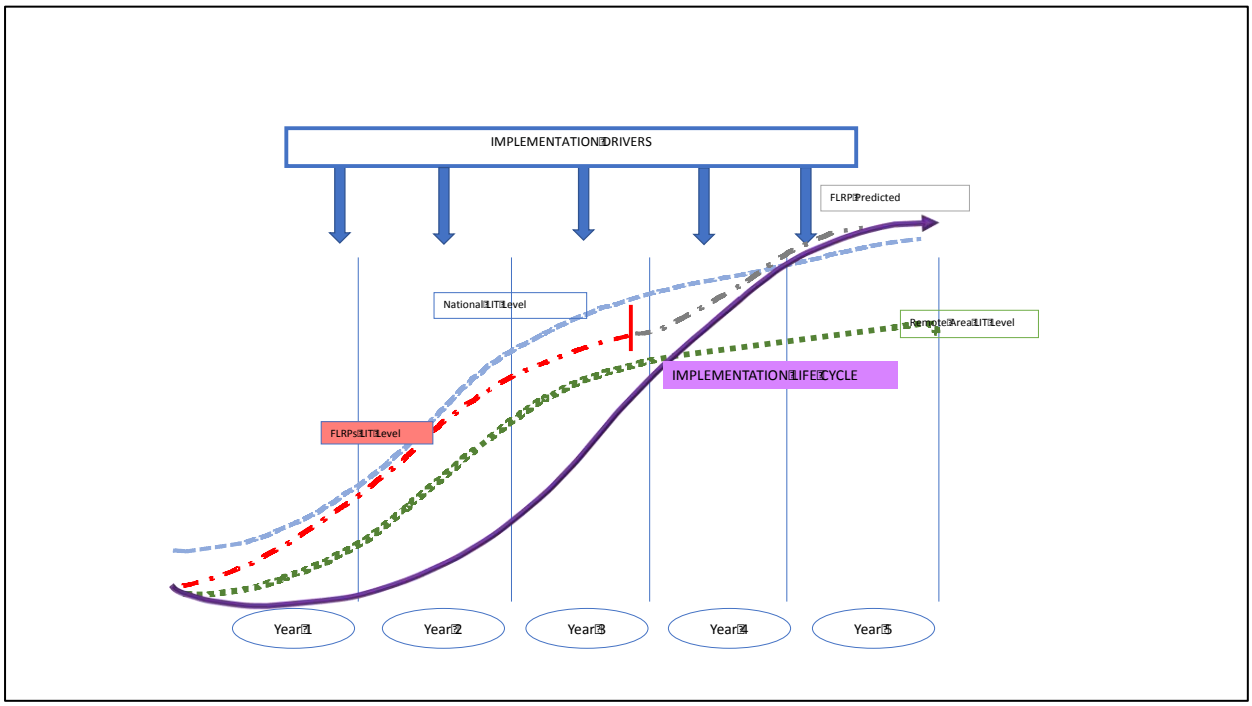
The implementation of FLFRPSP is having an impact on the students, teachers, and schools within the program with, not surprisingly, higher impact – in terms of intensity, complexity, and breadth – associated with more developed implementation processes and practices. The evaluation demonstrates that once training and implementation is achieved, and there is engagement by teachers and students, positive change in literacy learning can occur.

Figure 73, pictured below, demonstrates the life course of the FLFRPSP by plotting the predicted path of implementation as well as the current and predicted paths for literacy change across the various States and Territories and suggests when the program may realise its full potential for outcomes (Year 5). Once implementation reaches higher levels of dosage, impact within the classroom for teachers and students should be visible. This of course assumes that time will be taken to achieve the desired results.

While the program is demonstrating clusters of success, one of the goals here in making an evaluative judgement is to predict the probability of successfully scaling the project. The argument proffered is that some form of explicit instruction (DI/EDI) will positively impact on literacy outcomes for children.

There is little doubt that remote schools in Australia are at a disadvantage. Generally, remote areas experience higher attendance challenges, behavioural issues, proportion of need, teacher attrition, principal turnover, and resource need. Attempting to implement in such circumstances raises considerable challenges; however, education continues and programs are often successful despite the circumstances.

While the context for the program cannot be ignored, it does not totally define the success or failure of the program. Subsequently, it will be essential to understand the relationship between implementation fidelity and the context. Once this is established, scale and implementation of the program can be fully explained.



Program sustainability is a key element for scaling up. To consider the sustainability and scale of the program, a collective approach is necessary. Systems, communities, and schools can disrupt the current challenges that may be restraining the program or causing variable results. The program and its communities may need to be more adaptive to the environment and more agile in relation to the implementation drivers.

The next phase of the evaluation will explore these issues. The results indicate that the implementation of the FLFRPSP has been, to some extent, a complex and intricate solution for how best to enhance quality literacy learning in remote Australia.

INTRODUCTION

FLEXIBLE LITERACY FOR REMOTE PRIMARY SCHOOLS PROGRAM

The Coalition government has invested $23.8 million to improve reading and writing learning outcomes for students in remote primary schools. The FLFRPSP was part of a 2013 election promise and formed a component of the Coalition's policy for schools, Students First. In developing the program, it was recognised that students in remote and very remote schools faced diverse barriers to learning that were pertinently different from students in metropolitan or regional schools.

In 2014, GGSA was contracted by the Commonwealth government to implement the FLFRPSP based on the principles of NIFDI's Direct Instruction. The program's primary objectives are to:

- Improve students' literacy abilities and results
- Increase teacher pedagogical skills in teaching literacy through the use of alphabetic teaching approaches, in particular, DI or Explicit EDI

DI is affiliated with an instructional approach and curriculum materials developed by Siegfried Engelmann and Carl Bereiter in the late 1960s. Direct instruction is the use of explicit teaching techniques to teach specific literacy skills. Teachers who are trained in the direct instruction method will follow a lesson-by-lesson approach to instruction that follows a pre-determined skill. Lessons are sequences and are worked through with students. Teaching is tightly paced and incremental, aiming to maximise time-on-task. Teachers positively reinforce student behaviours and success at each level of attainment.

While still a method of explicit instruction, explicit direct instruction (EDI) gives teachers a step-by-step guide to creating and delivering lessons across all areas of the Australian Curriculum. The model was developed by John Hollingsworth and Dr Silvia Ybarra from DataWORKS. The approach is based on educational theory, brain research, data analysis, and direct instruction. EDI is a strategic collection of instructional practices combined to design and deliver explicit lessons, in this case, literacy lessons.

GOOD TO GREAT SCHOOLS AUSTRALIA

GGSA is a not-for-profit organisation aiming to enable schools to transition from Poor to Fair, Fair to Good, and Good to Great. GGSA targets three educational domains identified as critical for school and student improvement: Great Teachers, Effective Instruction, and Every Child. While all GGSA programs address these overarching objectives, they are tailored to meet the varying needs of schools and students, with schools able to implement some or all elements of the program depending on their specific needs and those of their students. At present, GGSA's focus is addressing the long-term trend of underachievement in Australia, and supporting schools in transitioning from Poor to Fair. Improving teaching practice is seen as particularly important in enabling improvement of low performing schools, and is a key focus of GGSA in working with these schools.

GGSA is responsible for introducing and embedding the pedagogical teaching approaches, EDI or DI, in remote primary schools in NT, QLD and WA. In 2015, GGSA began working with 33 schools. As at July 2017, there are 35 schools in the program, represented in Table 2.

Table 2: *Number of Schools in FLFRPSP by Jurisdiction and Type*

	DI	EDI	Total
NT Government	18	0	18
NT Independent	1	0	1
WA Government	3	1	4
WA Independent	3	1	4
WA Catholic Education	2	5	7
QLD Government	0	1	1
Total	27	8	35

EVALUATION OF THE FLEXIBLE LITERACY FOR REMOTE PRIMARY SCHOOLS PROGRAM

PROJECT BACKGROUND

In 2015, following an open tender process, the CPE was contracted by GGSA to quantitatively analyse available data on FLFRPSP to understand its progress in implementation and initial performance indicators, as well as to help GGSA build its internal evaluation capacity.

The technical report submitted to GGSA in 2016 focused on setting up of the evaluation, including establishing relationships with the project team, building the database, obtaining research application approvals with relevant jurisdictions and some data collection of 2015 data. Importantly, it focussed on assessing the quality of and access to data as well as determining any gaps for collection, and any concerns with evaluation capacity within the GGSA and schools.

The technical report produced in 2016 was limited by the amount of available data, but was able to report on implementation progress such as school fidelity, teaching fidelity, and student progress. In this context, teaching fidelity and school fidelity represent the quality and degree of program delivery. CPE created a customised Microsoft Access® database, containing internal program data only, which will be provided to GGSA at the end of the evaluation to allow them to continue monitoring and evaluating the program. Overall, this first component of the evaluation (a) ensured that GGSA had the capacity for ongoing evaluation, and (b) rigorously analysed available data on the program's early impact on several outcomes including teacher skills and student literacy as measured by internal program data.

In late 2016, management of the contract between CPE transferred to the Commonwealth Department of Education (DET), and the scope of the contract expanded to a full evaluation design, allowing for broader collection of data. This design provided a more robust and nuanced analysis of the impact of FLFRPSP which would guide policy and provide credible evaluative information back to GGSA and the schools in the program. The present evaluation design seeks to determine the impact of the program in achieving its foremost objectives. Specifically, the present evaluation will look at the program's effectiveness in (1) improving the literacy results of students in participating schools, and (2) and improving teachers' pedagogical skills in teaching literacy in the DI and EDI instruction models.

EVALUATION FRAMEWORK AND APPROACH

While the fundamental principle behind evaluation is to make a judgement of the merit and significance of the intervention or program, at the same time, evaluation information should always be utilised for program enhancements and hence add value. The work with FLFRPSP has utilised an adapted Centers for Disease Control and Prevention for Public Health evaluation framework (Centers for Disease Control and Prevention, 1999; see *Figure 1*) to embed principles within the evaluation. The framework promotes a continuous cycle of consultation and feedback between all stakeholders across each stage of the evaluation, and most importantly is underpinned by the evaluation standards of utility, feasibility, propriety, and accuracy. Given the nature of this evaluation and the sensitivities that have been raised in CPE's initial work, these standards are critically important.

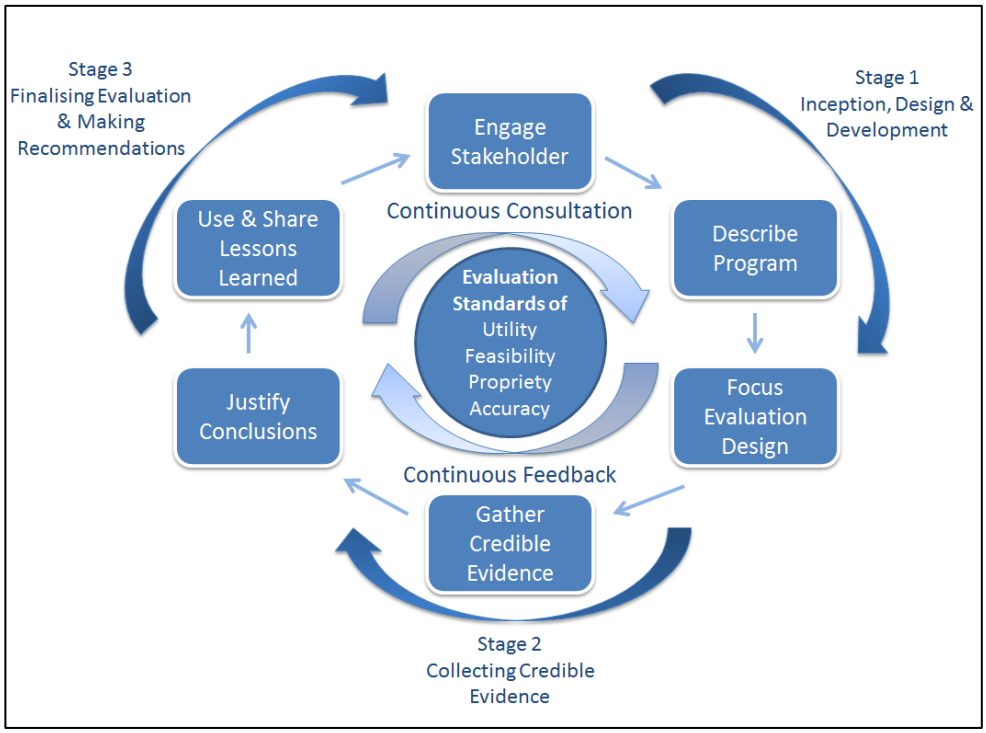


Figure 1. **Evaluation Framework.**

As described by the diagram in *Figure 2*, this evaluation has been an iterative process that commenced with the establishment of an evaluation plan and an assessment of the quality of accessible data. It is noted that information to rigorously evaluate the program was not previously available, hence the decision to establish a scrupulous process for data collection and evaluation. The principles for the evaluation process relate to the evaluation standards that suggest the evaluation must adhere to the principles of utility, propriety, accuracy, feasibility, and accountability (including reproducibility). Throughout the process, the evaluation was designed to add value, ensure evaluation capacity building occurred, and develop evaluative mindsets in key stakeholders.

This phase of the evaluation (Phase 2) is focussed on the short-to-medium rather than the medium-to-long-term outcomes of implementation and impact. Essentially, this phase will focus on data gathering and clarifying the measurement model. Data gathering will determine the current gaps in data and areas where the data needs strengthening. For example, this evaluation phase has not had access to the NT literacy data. Once gathered, each data element is analysed to determine relationships, hence testing the elements of the

program logic. Once the relationships are determined and the model confirmed, the subsequent phase (Phase 3) examines the program's progressive impact through a behavioural insights lens.

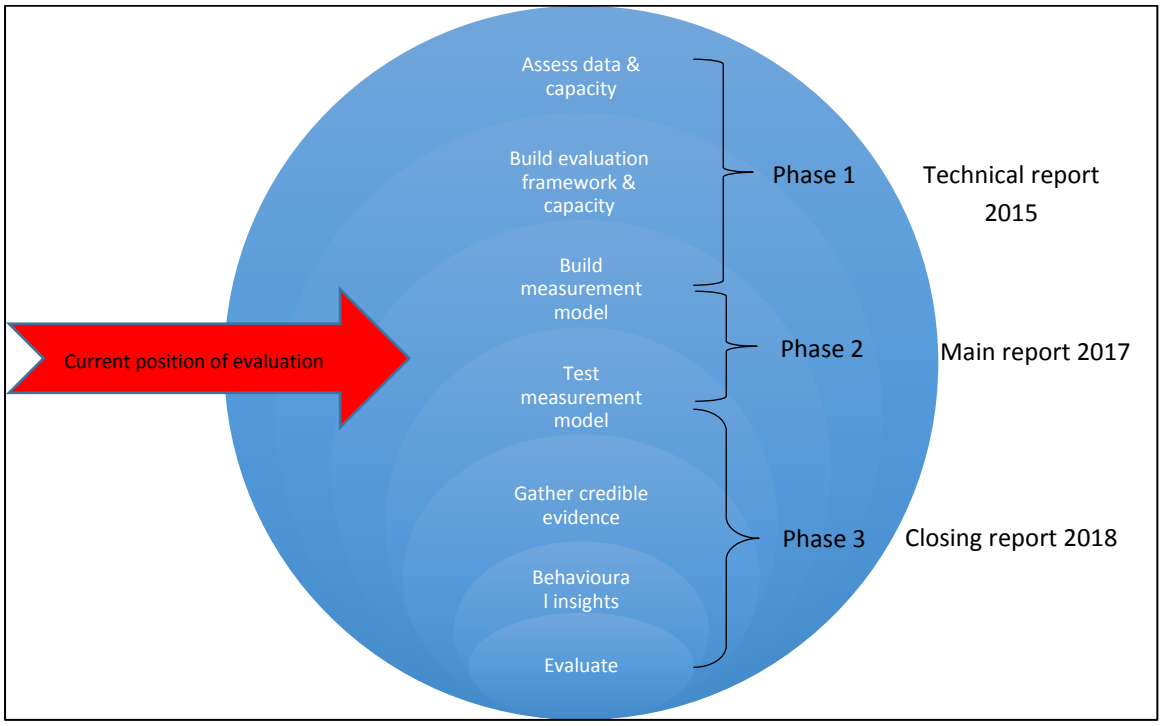


***Figure 2*. Evaluation Progress.**

In making an evaluative judgement at this phase of the evaluation, the program logic specifies the causal links between the implementation of FLFRPSP and literacy outcomes. The theory of change also demonstrates that there are a number of mediating factors that will account for variance in impact. When considering an evaluative judgement based on prediction, it is important to account for moderators and mediators.

The evaluation framework and structure establishes a theory-based approach to ensure that causal inferences can be made over time. Utilising an evaluation heuristic that suggests that the predictive merit and worth of a program (Y) is determined from the sum of a number of parts, such as ax + bx + cx and so on, will be useful when valuing the impact relative to the program fidelity and program drivers. Subsequently, we need to take into account the program theory of change, its fidelity (the degree of implementation and the quality of implementation), and the goodness of fit of the program to the context in order to ensure a clear picture of the causal path over time. Similarly, the drivers of the process of change and evaluation (access to and quality of information available) will influence the outcomes and consequently any judgement that can be made. At this middle phase of program implementation, the strength of relationships needs to be considered and the measurement model refined so as to ensure the evaluation process.

DEVELOPING A MEASUREMENT MODEL

The CIPP model of evaluation (Stufflebeam, 2000) provides an underlying structural lens to view data collection and establish a longitudinal approach to the evaluation. The CIPP model was developed in an educational context and provides a structure that allows for the formative and summative assessment of programs by considering Context, Input, Process, and Product (see *Figure 3*). Each of these aspects is described within the evaluation of the FLFRPSP.

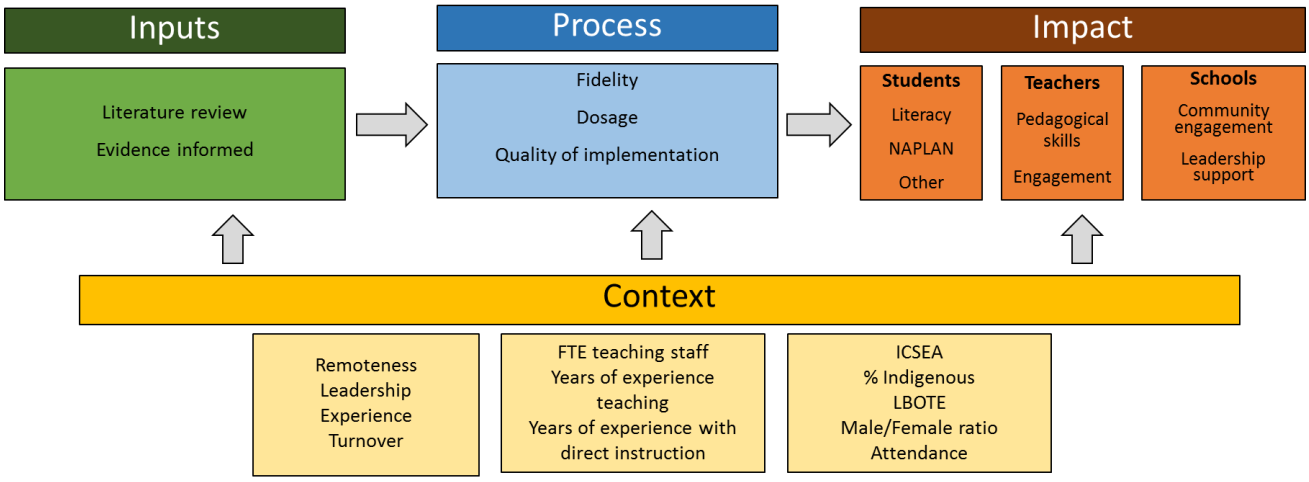


Figure 3. **Diagram of the CIPP Model.**

CONTEXT

Implementing educational reform in a complex realm

In undertaking this evaluation, it is critical that the contexts in which the program is implemented are well considered and understood. One particular challenge is that the contexts are not homogenous at a student, school, or jurisdictional level and this effects the type of data that must be collected. For example, this is the impetus for gathering data and evidence beyond student achievement. The evaluation is not only concerned with where the program is working but equally, what factors might explain why the program works better in some contexts compared to others.

As Snyder (2013) states:

> *"Indeed, what works for one element of one system may not work for other elements even within the same system."*
>
> *"Successful implementation will require a strong focus and commitment on the long-term change of the whole system. The actors will need to talk and work their way through the bumps in the road and build the communicative structures to smooth contextual differences and share experiences learning. This cannot be built into the reform from the outset but must grow with the program as it expands and takes hold."*

Context therefore considers the environment within which FLFRPSP is implemented, seeking to establish the characteristics of the key stakeholders. Consideration of contextual factors will guide the understanding of impact, particularly where impact may differ between schools or jurisdictions. These contextual factors will be captured at the student, teacher, and school level and will provide a more nuanced understanding of what variables are associated with differences in program outcomes between different contexts. These variables are shown in Table 3.

Table 3: ***Contextual Variables at Student, Teacher, and School Level***

Domain level	Contextual variable
Student	ICSEA % Indigenous LBOTE Male to female ratio Attendance
Teacher	Years' experience teaching Years' experience with DI Teacher turnover
School	School remoteness Leadership turnover # FTE teaching staff

INPUT

The input component of this evaluation provides understanding of the program in the context of the problems it is seeking to address, namely improvement of student literacy outcomes in the primary years. To achieve this, a synthesis of the literature on similar programs implemented in similar contexts is provided as input into understanding what is benchmark or standard progress for these types of interventions. The evaluation questions for this component are:

- Based on this synthesis of the literature, what is reasonable progress to expect from FLFRPSP?
- What factors will impinge on progress and impact of FLFRPSP?

PROCESS

The process component of this evaluation focuses on implementation, seeking to understand fidelity, dosage, and quality as well as the barriers and enablers to implementation. Again, these factors are considered at the school, teacher, and student levels.

PRODUCT (IMPACT)

The CIPP model describes Product Evaluation as the measurement, interpretation, and judgement of a program's outcomes (Stufflebeam & Coryn, 2015). Product within the evaluation of the FLFRPSP is referred to as impact from hereon and refers to measured outcomes of the program on the following stakeholders:

- Students – literacy, NAPLAN, engagement, behaviour and wellbeing
- Teachers – pedagogical skills, engagement, attrition
- Schools – engagement with community, all staff attrition

STUDENTS

The primary measure of impact on students is literacy achievement indexed by available standardised jurisdictional literacy data (e.g. EYLND, WA Catholic Education Office). Unfortunately, the provision of this data by jurisdiction is still limited and much of the data is still incomplete, making analysis problematic. Where available, this data was analysed within the program schools over time and compared to the same measures for a set of matched control schools to understand the impact of the program on literacy outcomes for students. In considering impact in the context of these data sources the relevant evaluation questions are:

In schools who have implemented FLFRPSP,

- How does student literacy achievement in 2015 and 2016 compare to the baseline year prior to program implementation (2014)?
- Is literacy achievement greater in program schools when compared to matched control schools?

NAPLAN

NAPLAN data has been collected and analysed as part of the current report. The evaluation team acknowledges the limitations of NAPLAN data for understanding the early impact of FLFRPSP. In examining the potential causal chain between the program and NAPLAN outcomes, a more longitudinal approach needs to be taken. It has been subject to some initial analysis; however, at this point in the evaluative process, caution is recommended in attributing any outcomes (positive or negative) in NAPLAN to the program.

BEYOND LITERACY: BROADER CONCEPTIONS OF IMPACT ON STUDENTS

Recognising that student improvement in academic outcomes can be paralleled by other cognitive and non-cognitive outcomes, the evaluation also seeks a broader understanding of impact on students. This broadened understanding includes engagement, behaviour, and wellbeing, and will be captured by several methods including teacher and principal perception of student engagement, observation of lessons, and a survey of students in the program to gain their perceptions of these factors.

Initial findings on these additional outcomes are presented in the current report. Further outcomes will be included in the closing report in 2018.

TEACHERS

Impact from the perspective of teachers in schools implementing the program include both direct program impact such as pedagogical skill in DI/EDI, as well as broader impact on engagement in teaching generally, such as turnover/retention and wellbeing. Teacher pedagogical skill is captured through FLFRPSP data. Other perception-based measures of impact have been determined through a survey of teachers with preliminary results presented in this report and a broadened scope of perspectives to be presented in the closing report in 2018.

SCHOOLS

The overall notion of impact with respect to schools encompasses a range of constructs. These constructs include: principal skills in program implementation, school and community engagement with DI/EDI, and overall staff turnover/retention. Some of these constructs have been measured and included within this report via interviews with a subset of school Principals. Additional data sources will be collected for the closing report based on case study, further interviews, and observation. These will provide for a richer understanding of impact at the school level as well as providing important contextual information about program impacts.

THIS REPORT

PURPOSE OF THIS REPORT

The outputs of this evaluation are three reports (see *Figure 2)* with this Main Report focusing on analysis of data to demonstrate early changes in student literacy outcomes and to begin capturing perspectives from Principals and teachers. More specifically, the present report outlines the changes in students' literacy results since they began participating in the program, over time, and relative to control schools. In addition to student

performance data, the present report presents results from both interviews with Principals and surveys completed by a subset of teachers in program schools. The survey data relates to teachers' perceptions of the program, such as student engagement and performance, as well as its impact on their pedagogical skills and practice. The closing report due early 2018 will build on these findings, and provide more nuanced understanding of the program's efficacy and the perspectives of key stakeholders.

STRUCTURE OF THIS REPORT

This report began with an executive summary and introduction, which gave an overview of the evaluation objectives and outcomes. Following this introduction, the report detailed the evaluation framework and approach. It now presents the evaluation methodology which includes the evaluation questions, specific methods, and data analysis plan. This is followed by the results section, which details at student, teacher and school level for each program (DI and EDI) and for each jurisdiction (where possible):

1) demographic information,
2) program implementation progress,
3) standardised student literacy outcomes,
4) impact evidenced through other data sources, and
5) overall summary

All results are considered in the subsequent summary and discussion, where the findings are critically examined in the context of related research and theory.

LITERATURE REVIEW

Following is the result of a brief and rapid review of relevant literature on direct instruction pedagogies, their efficacy generally, and more specifically on a range of academic and allied academic domains. It also considers factors that affect implementation and special considerations for Indigenous populations. Please note that use of the acronym DI only applies to the NIFDI program of explicit instruction. Where the term direct instruction is used in full, it is referring generally to explicit instruction pedagogies.

INTRODUCTION

Direct instruction as a pedagogical approach has been widely examined and reviewed in recent years. This literature review seeks to examine the most prominent existing analyses and studies on the implementation of direct instruction in a number of contexts since 2000. The review is split into the following parts: Meta-Analyses, Studies, Factors Affecting Implementation, and Indigenous Australian Considerations.

DEFINITIONS

Direct instruction is a general term for a set of teaching practices based on the explicit explanation of materials and explicit direction of students. Behavioural and cognitive goals are made explicit to students (Luke, 2014) and content is made clear through demonstrations, participation, guided practice, and discussion (Hattie, 2008).

Direct instruction as developed by Engelmann (Engelmann & Carmine, 1982), is a more specific model based on a behaviourist approach to learning. Engelmann's direct instruction utilises a scripted, highly structured approach "that follows a pre-determined skill acquisition sequence."

EDI is a more specific application of direct instruction, and was developed by Hollingsworth and Ybarra (2009). EDI takes a more flexible approach to teaching and allows teachers to modify their instruction based on student needs.

META-ANALYSIS (EFFICACY)

In a mega-analysis combining four meta-analyses consisting of a total of 304 individual studies, Hattie (2008) determined that direct instruction produced an overall moderate effect size of *d* = .59. Effect sizes varied between studies, depending on the subject being taught. For example, the effects of direct instruction were larger for reading (*d* = .89), than for mathematics (*d* = .50). Other meta-analyses have demonstrated effect sizes of varying magnitudes, ranging from small (*d* = .21; Borman, Hewes, Overman, & Brown, 2003) to moderate *(d* = .55; Haas, 2005). Haas (2005) analysed teaching methods for mathematics and found that direct instruction had the largest effect for both low-ability (*d* = .84) and high-ability students. Borman et al. (2003) found that direct instruction models, while having an effect size of *d* = .21 overall, had a smaller effect size when only studies with comparison groups were analysed (*d* = .15).

The authors were unable to identify any meta-analyses on direct instruction research conducted post-2008.

STUDIES

While there has been a lack of meta-analyses in the past 10 years, several studies have examined the efficacy of a direct instruction approach in various contexts.

LITERACY

Studies measuring literacy have found generally positive results. Shippen, Houchins, Steventon, and Sartor (2005) found that 55 students enrolled in two different direct instruction programs all made significant gains, but more capable students made greater gains than lower-achieving students. Rebar (2007) found that first-graders being taught with the National Institute for Direct Instruction (NIFDI) model increased literacy outcomes significantly faster than the comparison schools, and by the end of the project were above the national mean as well as above the mean of the comparison schools. In total, the project had more than 41,223 students enrolled. From a sample size of 104 Grade 4 students, Lencioni (2013) found that lower performing readers benefitted from direct instruction with independent practice of comprehension strategies, whereas higher performing readers benefitted from direct instruction with cooperative learning groups to practice comprehension strategies.

However, evidence is not exclusively positive; Gaston, Martinez, and Martin (2016) found direct instruction to be less effective than other teaching methods aimed at literacy.

MATHS

Although most studies focus on literacy outcomes, Al-Makahleh (2011) found that direct instruction significantly increased the maths outcomes of Grade 4 and 5 students with learning difficulties. direct instruction was found to be less effective than other teaching methods aimed at mathematics (Firdaus, Wahyudin, & Herman, 2017).

WELLBEING

One Melbourne study conducted a-based social and emotional skills program, *You Can Do It! Early Childhood Education Program*, on 99 prep and Grade 1 students in Melbourne (Ashdown & Bernard, 2012). At the end of the 10-week intervention, according to teacher reports, the program had a significant positive effect on the students, increasing levels of social-emotional competence and well-being, as well as reducing problem behaviour in Grade 1 students. Additionally, lower-achieving Grade 1 students showed an increase in reading achievement.

SPECIAL EDUCATION AND LEARNING DIFFICULTIES

Direct instruction has been effective in increasing academic outcomes for students with learning difficulties. Hattie (2009) demonstrates that the impact of direct instruction is similar for mainstream students and those with learning difficulties.

Al-Makahleh (2011) found significant differences in mathematics outcomes between an experimental and a control group of Grade 4 and 5 students with learning difficulties. Cadette, Wilson, Brady, Dukes, and Bennett (2016) found direct instruction to be effective for teaching students with autism how to answer 'wh-'-type questions.

direct instruction has also been shown to be more effective than other teaching methods for increasing literacy in students with learning difficulties, including Simultaneous Prompting Procedure (Celik & Vuran, 2014), Activity-Based Intervention (Botts, Losardo, Tillery, & Werts, 2014), and follow-in labelling and incidental exposure to vocabulary (Lund & Douglas, 2016).

However, direct instruction was found to be less effective than Strategy Instruction among students with Intellectual Disability (Blik, Harskamp, & Naayer, 2016); Strategy Instruction resulted in more student autonomy and better academic achievement.

Overall, the use of direct instruction methods in special education seems promising, although study designs in this area would benefit from greater population sizes and more strictly experimental designs.

ATTENDANCE

Some researchers have suggested that implementation of direct instruction programs may increase student attendance as a result of increased student well-being. Minyerri school in the NT, which has run a direct instruction program since the beginning of 2016, saw an attendance average of 88% for Term 1, 2016, up from 76% for Term 1, 2015 (Chandler, 2016b). The Cape York Aboriginal Australian Academy (CYAAA) project, however, found that student attendance over the three-year trial either decreased or remained the same (ACER, 2013; McCollow, 2012).

OTHER

While many studies with a pre- and post-test design found moderate to large effect sizes, it should be noted that these single-group studies did not have a control group and are therefore lack methodological rigor (Borman et al., 2003). Generally, most studies would benefit from more rigorous methodology and greater sample sizes.

FACTORS AFFECTING IMPLEMENTATION

ATTENDANCE

Attendance has been identified as a factor negatively affecting implementation of direct instruction programs (Wolgemuth, Savage, Helmer, Lea, & Harper, 2011). Teacher reports identify tardiness and low attendance as reasons students' skills are not progressing (Grossen, 2004).

TEACHERS

Qualitative evidence suggests that teacher training in direct instruction programs is an important part of the success of any direct instruction-based intervention (Dow, 2011; Grossen, 2004; Kanfush, 2014; Lowe & Balcher, 2012). Similarly, the often highly-scripted and inflexible nature of direct instruction interventions does not allow for teachers to modify definitions or explanations for misunderstandings; Dow (2011) shares the

example of a teacher explaining the rule that "all mammals have hair", to which a student replies that a spider must then be a mammal: "The scripted correction procedure didn't cover explaining to 8 year-old, on-edge, at-risk, Indigenous low-progress readers, whose first language is not English, that the rule does not exclude other types of animals from having hair, and is not a definitive rule about mammals by itself" (Dow, 2011, p. 57). Teachers may also not maintain fidelity to scripted lessons (Kanfush, 2014).

COMMUNITY

It has been argued that children from different cultural and economic backgrounds require specialised or modified versions of direct instruction (Gaston et al., 2016).

This is also true for Australian rural and Indigenous school contexts. The community and greater culture is also cited as a potential obstacle when the culture of the community is different to or unfamiliar with that of the school (Luke, Dooley, & Woods, 2011; Stewart, 2002), as this creates another barrier to content comprehension. Students must bridge the cultural and contextual gap as well as learn the educational content.

PARENTS

Kanfush (2014) reported parent perceptions of direct instruction, and found that while parents did not have a clear understanding of what constituted the direct instruction approach, they were nonetheless very satisfied with the program being used to teach their children reading skills and found that their children were progressing. The lack of parent understanding may result in parents being less effective in supporting the program and reinforcing the instruction (Kanfush, 2014).

OTHER

While it may be presumed that other factors such as school leadership, socio-economic status, age, and region affect implementation of direct instruction programs, the authors were unable to identify any other factors explicitly described in the research.

INDIGENOUS AUSTRALIAN CONSIDERATIONS

As the academic achievement gap between Indigenous students in Australia and privileged, non-Indigenous students has continued to widen (De Bortoli & Thomson, 2010; Ewing, 2011; Luke et al., 2011), attention has turned towards improving the outcomes of underachieving students. This has resulted in direct instruction-oriented programs such as the Schoolwise Program in NSW (Wheldall, Beaman, & Langstaff, 2002), which compared academic improvement between Aboriginal and non-Aboriginal students, and the CYAAA project, which has demonstrated mixed results overall (Australian Council for Educational Research, 2013).

There is debate around the implementation of literacy interventions among Indigenous students, especially regarding the factors that affect the success of implementation of such interventions. Several researchers have suggested that Indigenous students require different education strategies on the basis that existing models of education are suited to a different cultural context (Dow, 2011; Luke et al., 2011; Stewart, 2002). Luke et al. (2011) argue that instead of focusing on comprehension and academic performance, teaching models for Indigenous students should focus on "bridging community cultural practices and epistemologies with systematic introductions to the specialised genres and registers of school and institutional texts" (p. 154). This aligns with findings by De Bortoli and Thomson (2010) that certain cognition, beliefs, and actions, for example a student's interest in learning and belief in their own ability, are associated with learning outcomes. Language concerns are also made apparent; students whose first language is often not Australian English must engage with materials that are not made for their language, adding another barrier to their learning and to implementation of any programs (Dow, 2011). One critique of direct instruction programs for Indigenous

students is that a direct instruction lesson which is highly scripted and fast-paced does not allow for specific corrections or wider explanations when misunderstandings (often culturally-based) occur (Dow, 2011).

The Schoolwise Program, which included 14 Aboriginal students and 20 non-Aboriginal students in an intensive MULTILIT program over 20 weeks, found that both groups showed significant and major gains on all measures used ($p < .001$; Wheldall et al., 2002). Differences between Aboriginal and non-Aboriginal students both pre- and post-test were not statistically significant. This result contradicts the idea that Indigenous students require a different approach to non-Indigenous students; however, the sample population was quite small and the study did not incorporate a control group.

The CYAAA Initiative was designed to improve rural and Indigenous student outcomes in three schools in QLD. The model was divided into three separate learning domains: Class, in which literacy, numeracy, and English were taught using a direct instruction methodology; Club, in which children were provided with artistic, musical, and sporting activities; and Culture, in which Indigenous culture and language were taught (ACER, 2013). Quantitative data for the Initiative was quite limited, and it is therefore inconclusive whether the Initiative had an impact on student learning, although existing data shows that a greater proportion of students are at or above the national literacy standards. Qualitative data from evaluation, however, indicates that the Initiative has resulted in improvements in student learning, especially literacy, improvements in staff professional development, and has increased community attendance and involvement in the Initiative. In addition, "student behaviour has improved" and teachers have "higher expectations of their students" (ACER, 2013, p. 9). Student attendance, however, decreased since the beginning of the Initiative. Overall, qualitative data shows a variety of positive outcomes for CYAAA, although quantitative data remains largely inconclusive (ACER, 2013; McCollow, 2012).

CONCLUSION

There is a growing body of positive evidence for the efficacy of direct instruction in a number of educational domains. Meta-analyses examining efficacy of teaching methods have found direct instruction to be one of the most effective methods, with effect sizes ranging from $d = .21$ to $d = .59$ (Borman et al., 2003; Hattie, 2008). Studies investigating the effectiveness of direct instruction have also found it to be an effective method for increasing literacy, maths, and wellbeing outcomes. Direct instruction is also an effective method when used for students with learning difficulties; however, direct instruction is often less effective than other teaching methods for this population. A number of factors affect the efficacy and implementation of direct instruction programs, including student attendance, teacher training, and parental involvement. Direct instruction in Indigenous and rural Australian contexts is complicated by the differences in culture and social context, as well as language use. The efficacy of direct instruction among Indigenous populations is unclear, but seems to be positive. Finally, although direct instruction in general seems to be generally effective, studies in this area would benefit from more rigorous study design and larger sample sizes.

METHODOLOGY

RATIONALE

The overarching principle of this evaluation is to provide credible information to DET regarding the impact of FLFRPSP on literacy outcomes in students and teacher pedagogical skills. In so doing, CPE has endeavoured to:

- Leverage and build on evaluation work to date, including data collected
- With DET, identify additional sources of data (interviews with key stakeholders, other academic assessments) that will inform impact

- Continue to focus on capacity building within GGSA and where possible in schools around data management and monitoring
- Focus on sustainability of the evaluation work beyond the life of the evaluation through frameworks and processes

As described, aims are underpinned by the Centers for Disease Control and Prevention evaluation standards adopted from the Joint Committee on Standards for Educational Evaluation (1999).The standards consider four factors: utility, feasibility, propriety and accuracy. Further, CPE has collected this data with propriety adhering to both university and jurisdictional ethical requirements and guidelines.

There are two key and inter-related aspects to the methodology driving this evaluation. First, FLFRPSP is educational reform in heterogeneous and complex contexts, and this complexity needs to be factored in to what data is collected, how this data is collected, and how this data is analysed and interpreted. Second, the evaluation takes a longitudinal view, and in so doing considers the data examined in this report as having utility in the current discussion regarding the program, but also having utility as the foundation for ongoing monitoring and evaluation of the program over time. These two concepts are inter-related because educational reform in any complex environment takes time and deserve time. Evaluation plays a critical role in the provision of this time by future-proofing monitoring systems and developing measurement models that are sustainable and where possible standardised so that relevant analysis can be compared over time and in a meaningful way.

METHODS

The evaluation utilises a mixed-method convergent parallel design in the collection of data. It is convergent in that it utilises both quantitative data and qualitative data to gain a robust understanding of the program impacts, including drivers and inhibitors of program success for all key stakeholders. The parallel aspect reflects the fact that the two data types have been collected together. Following is a description of all data sources drawn from in the evaluation of FLFRPSP, their characteristics, participants represented, and their role in contributing to the findings in this report.

DESCRIPTION OF THE DATA SOURCES

FLFRPSP PROGRAM-BASED DATA

Type: Quantitative

Description: a wide range of data sets related to student placement in the program, lesson progress, student mastery, records of program activities such as coaching sessions and training, and implementation fidelity indicators

Use in evaluation: provides several indicators of program implementation, dosage, fidelity, and student progress

Participants represented by the data: Students, teachers, school leadership

STATE/JURISDICTIONAL LITERACY DATA

Type: Quantitative, standardised

Descriptions: EYLND – WA Catholic Education Office; PM Benchmark Reading – WA

Use in evaluation: an indicator of impact of program on literacy outcomes considered proximal to classroom learning

Participants represented by the data: Students

NATIONAL ASSESSMENT PROGRAM FOR LITERACY AND NUMERACY

Type: Quantitative, standardised

Description: Annual assessment for students in years 3, 5, 7, and 9. Tests utilised in this evaluation were Reading, Writing, Spelling, Grammar and Punctuation for program and control schools for the assessment years 2014, 2015, 2016.

Two analyses have been conducted using NAPLAN information: changes to student participation rates in NAPLAN assessments and analysis of change in NAPLAN results over time in program schools and compared to control schools.

Use in evaluation: a potential indicator of the impact of program on literacy outcomes considered more general than state/jurisdictional based literacy measures

Participants represented by the data: Students

TEACHER SURVEY

Type: Quantitative and qualitative, purpose designed

Description: Survey questions pertaining to teacher perceptions of the following domains (with # of questions in brackets) – general demographics (9), knowledge of the program (1), training (4), implementation (6), fidelity (13), attitudes towards the program (6), perceptions of student outcomes (10), teacher self-efficacy (1), job satisfaction (3) and wellbeing (1).

Use in evaluation: Stakeholder perceptions as indicators of aforementioned measurement domains to be triangulated with quantitative program and literacy achievement data

Participants represented by the data: Teachers (respondents), students

SECONDARY DOCUMENT ANALYSIS

Type: Qualitative

Description: Publicly available reports, information from websites, press releases

Use in evaluation: Secondary indicators or support for other indicators of student progress, teacher experience, school experience with the program

Participants represented by the data: Students, teachers, school leadership, schools

ANALYSIS OF PROGRAM, OUTCOME, AND PROGRAM DRIVERS

SEQUENCE OF ANALYSES

Data was analysed following this sequence.

- Student Literacy
    - Descriptive analysis
    - Program Schools vs Control Schools Demographic Comparison
    - Paired 2014 (pre-FLFRPSP implementation) and 2016 (post-FLFRPSP implementation) NAPLAN Score Testing between Program Schools vs Control Schools
    - Unpaired average change in NAPLAN scores between 2014 (pre-FLFRPSP implementation) and 2016 (post-FLFRPSP implementation)
    - State Level Literacy Change Analysis[2]
    - Change in National Minimum Standard (NMS) over years 2014, 2015 and 2016 for Grades 3 and 5
    - NAPLAN participation rates by year and grade level.
- Teacher Survey
    - Descriptive analysis
- Document Analysis

DATA PREPARATION

A stringent data analysis process was adhered to. This section will describe each step in detail; primarily the process entailed:

- Summary of data received
- Summary of data management
- Missing data analysis
- Missing data management

ANALYSIS PROCEDURES UTILISED

- **Missing Completely at Random Test**

Little's Missing Completely at Random Test was utilised in the analysis process; this is a statistical process for analysing missing data within a multivariate dataset (Little, 1988). Missing values are analysed to determine whether they are systematically absent from a dataset, or missing completely at random. If the result of this test is not significant (i.e. $p > .001$), the missing data are considered to be randomly distributed across all cases analysed. As such, it is appropriate to use imputation methods, such as Expectation Maximisation, to predict missing values.

- **Expectation Maximisation**

This is a process for estimating missing values. Values are calculated utilising the missing data's relationship to other variables (Expectation) which are then analysed to determine if the imputed value is most likely to be an accurate reflection of the missing data (Maximisation). This is a common statistical procedure which is considered to be one of the most rigorous methods for managing missing data, particularly if the data missing is random, or if less than 5% of data is missing.

[2] Where data was available.

- **Multiple Imputation**

Multiple Imputation is a statistical method used to overcome the problem of missing data points within a data set. Studies which have a large amount of data missing, and must exclude those participants with some data points missing decreases their statistical power and increases the probability of bias. In the present evaluation, many students are missing one or more data point, and we have chosen to use multiple imputation to increase the sample size and improve the power of analysis.

A common solution to missing data points is list wise deletion, whereby those cases with values missing are discarded. In the case of the present evaluation, that would mean that students who were missing one or more data point would be excluded from the analysis. However, there may be systematic differences between those students who have data and those missing data, and excluding all students who have missing data points may distort the sample, leading to biased and inaccurate results. Relatedly, excluding large amounts of data may weaken the analysis.

Another method for replacing missing values is single-imputation or imputation (such as Expectation Maximisation), where missing values are replaced with probable estimates based on the available data. Unlike multiple imputation, only one value is given as an approximation of the missing value. This implies a higher level of precision, and may inflate the confidence intervals and significant tests. Because of these shortfalls, multiple imputation was selected as the most appropriate method for accounting for missing data within the present evaluation.

Multiple imputation overcomes these issues by replacing missing data values with a set of plausible values estimated by other information within the data set. Unlike other models of imputation, where a missing value is replaced by one other, the two or more options given in multiple imputation demonstrates the uncertainty of the estimated value.

There are three stages of multiple imputation:

1- **Imputation -** The process of imputing alternative values and creating a complete dataset. In the case of multiple imputation several analyses are run, and multiple imputations are assigned.
2- **Analysis -** Each of the imputation results is analysed separately.
3- **Pooling –** The analysis results are pooled into one multiple imputation result.

Multiple imputation has been found to be more accurate than single imputation at estimating what the results would have been if values were not missing.

- **Multivariate analysis of variance (MANOVA) and Analysis of Variance (ANOVA)**

ANOVA and MANOVA is a family of statistical tests used to compare the means of different groups and determine if there is a statistically significant difference between them. There are several ANOVA models, each of which is only appropriate for certain variables in particular contexts. In some instances in the present report, one or more assumption were violated and a non-parametric test, Welsch's ANOVA, was used. Welsch's ANOVA can be used when the assumption of homogeneity of variance has been violated.

- **Paired Samples t-test**

A paired sample t-test, also called a repeated measures t-test, compares the means of two related groups on the same dependent variables to determine if there is a significant difference between them. These tests are typically used to compare the same group at different time-points, such as in the present evaluation, where student results were compared at different stages of the program.

- **Independent Samples t-test**

The independent t-test compares the means of two unrelated groups on the same dependent variables to determine if there is a significant difference between them. These tests are typically used to compare the intervention and non-intervention groups, such as in the present evaluation, where student results were compared between program and control schools.

- **Effect Sizes**

Effect sizes represent the magnitude of the effect, such as the degree of difference between the two groups or two time points. Importantly, an effect size is a standardised measure of an effect, and is therefore comparable across studies. Effect sizes are good supplementary information to statistical significance tests. While significance tests solely indicate whether there is a difference between groups or variables, effect sizes offer information about the magnitude and direction of that difference. There are a number of different measures of effect size, which are applicable to particular statistical tests. Over-reliance on significance testing does not provide an accurate picture of the between-groups difference.

SCHOOL COMPARISON ANALYSIS

The Australian Government DET provided CPE with a list of 255 schools originally targeted as eligible for the FRFRPSP. The criteria for school selection are:

- Bottom 30% nation performance measures by aggregate 2012 and 2013 NAPLAN data that show 30 per cent or more students below the national minimum stand in reading and writing,

and either or both

- High student vulnerability in the first year of schooling (measured by Australian Early Development Index collection)
- High student socio-educational disadvantage (measured by Index of Community Socio-economic Advantage)

The evaluation team used this list and consulted the myschool.com website to collect as much information as possible for matching schools including, location (remoteness), staff numbers, school Community Socio-Educational Advantage (ICSEA), student enrolment numbers, percentage of Indigenous students, and percentage of students with language background other than English (LBOTE). The data was used to identify like or similar schools to form the control groups. The mean values of school factors critical to implementation, such as the number of teachers and ISCEA values, are presented for both control and program schools in Table 4.

Preliminary assumption checking revealed that data was not normally distributed, as assessed by a Shapiro-Wilk test ($p > .05$); however, MANOVA is robust to this violate. One univariate and multivariate outlier was detected, as assessed by boxplot and Mahalanobis distance ($p > .001$); while MANOVA can be robust to violation of this assumption, the analysis will be run with and without the outlier. A linear relationship was not detected, as assessed by scatterplot, thus, sensitivity testing is recommended. However, no multicollinearity was detected (as seen in Table 1); and there was homogeneity of variance-covariance matrices, as assessed by Box's M test ($p = .342$).

A multivariate analysis of variance (MANOVA) was conducted and showed no statistically significant difference in school level variables between the control and program schools $F(6, 53) = 17.68$, $p > .05$; Wilk's $\Lambda = 0.376$,

$\eta^2 = 0.05$ (see table below for individual differences)[3]. Thus, schools were considered statistically similar and it is not necessary to control for any school level variables in the analysis the following NAPLAN analysis[4].

Table 4: ***School Comparison MANOVA School Variable Correlation***

	Fulltime Teachers	Fulltime Non-teachers	ICSEA	Student Enrolment	Percentage Indigenous	Percentage LOTE
Fulltime Teachers	1	**.543****	**.385****	**.858****	**-.403****	-0.217
Fulltime Non-teachers		1	**.468****	**.445****	**-.452****	**-.339****
ICSEA			1	**.602****	**-.829****	**-.593****
Student Enrolment				1	**-.631****	**-.319***
Percentage Indigenous					1	**.654****
Percentage LOTE						1

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

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

Table 5: ***Difference Between Control and Program Schools***

	Control Schools	Program Schools
Fulltime Teachers	10.87 (7.60)	10.07 (5.03)
Fulltime Non-teachers	4.90 (4.59)	5.71 (5.14)
ICSEA	661.93 (123.71)	667.67 (98.66)
Student Enrolment*	113.10 (111.43)	103.28 (62.48)
Percentage Indigenous	91% (1.9%)	92% (1.8%)
Percentage LOTE	76% (3.2%)	82% (2.7%)

*Full time equivalent student enrolment.

[3] The same results were obtained with or without outliers.

[4] Sensitivity testing using a series of One-way ANOVA confirms these results.

RESULTS

NORTHERN TERRITORY GOVERNMENT SCHOOLS

The NT has 206 schools, which include Government, Non-Government, and Special schools across Primary and Secondary years (source: ACARA 2017). Of these, 19 schools are currently part of FLFRPSP with 18 classified as Government schools and one, TIWI College, an independent school. All NT schools involved in the program run DI.

Selection of schools into the program for the NT was managed centrally by the NT Department of Education, who nominated eligible schools to participate in the program. Since the program's inception, 5 additional NT schools have joined the program and 2 have withdrawn. According to GGSA, there are three main reasons for withdrawal from the program across all states and jurisdictions:

- Lack of capacity to implement the program
- Funding
- Lack of community support

NT Government schools locally assess literacy using Progressive Achievement Tests – Reading (PAT-R). PAT-R tests assess reading comprehension, vocabulary knowledge, and spelling, and are normed on Australian students. Assessment of comprehension covers Foundation to Year 10 and assessment of spelling ranges from Year 2 to Year 10. PAT-R is a widely-used assessment in Australian schools to determine reading progression of students.

PAT-R aligns with the foundational skills required to read. In particular, it provides assessment of comprehension, both literal and inferential, contextual word knowledge or vocabulary, phonics, and phonemic awareness (spelling).

DEMOGRAPHIC PROFILE

Table 6 summarises relevant demographic information regarding FLFRPSP schools in the NT and compares them with the same data for all schools in the program. NT program schools have a lower mean ICSEA value than the mean for all program schools. Furthermore, NT program schools have a higher percentage of Indigenous students and students with a LBOTE. Attendance rates for schools in the program are in line with the attendance rates for all program schools. All schools are classified geographically as Very Remote.

Table 6: ***Mean Values for Key Demographic Variables for NT Program Schools with All Program School as a Comparison***

	Teachers	Staff	ICSEA	Total students	Girls	Boys	Indig %	LBOTE %
NT Gov	9.84	3.79	634.21	102.42	50.47	52.05	99.00	91.00
PROG	9.94	5.81	666.32	100.71	49.19	52.06	92.00	83.00

Table 7: *Attendance Rates for NT Program Schools*

	2014			2015			2016		
	Days present	Days absent	Attendance Rate	Days present	Days absent	Attendance Rate	Days present	Days absent	Attendance rate
NT Gov	119	50	70%	122	46	73%	124	48	72%
Overall	111	44	72%	110	44	73%	108	44	71%

STUDENT LEVEL DATA

IMPLEMENTATION

The DI program is timetabled to be delivered for 2.5 hours per day. This allows for a benchmark lesson progress rate across all groups of 1.0 lessons per day. Program data tracks the lessons taught so that an average lesson per day can be calculated at the school level. *Figure 4* shows this information and indicates the average lesson delivery per day ranged from 0.4 lessons to 0.9. This data represents the number of lessons that schools complete each day, on average, and not the hours spent on instruction. It is possible that, for a range of reasons, teachers are unable to complete the full lesson within the time available. For example, students who have been absent and returned to school may need to be re-assessed to determine their lesson level. Hence, this will take time away from normal lesson teaching. It is also possible that some teachers choose to spend less than the scheduled time on DI, preventing them from completing a full lesson each day. While this data shows that schools are not reaching full fidelity in terms of lesson completion, it doesn't explain why, with further insight offered by the teacher survey.

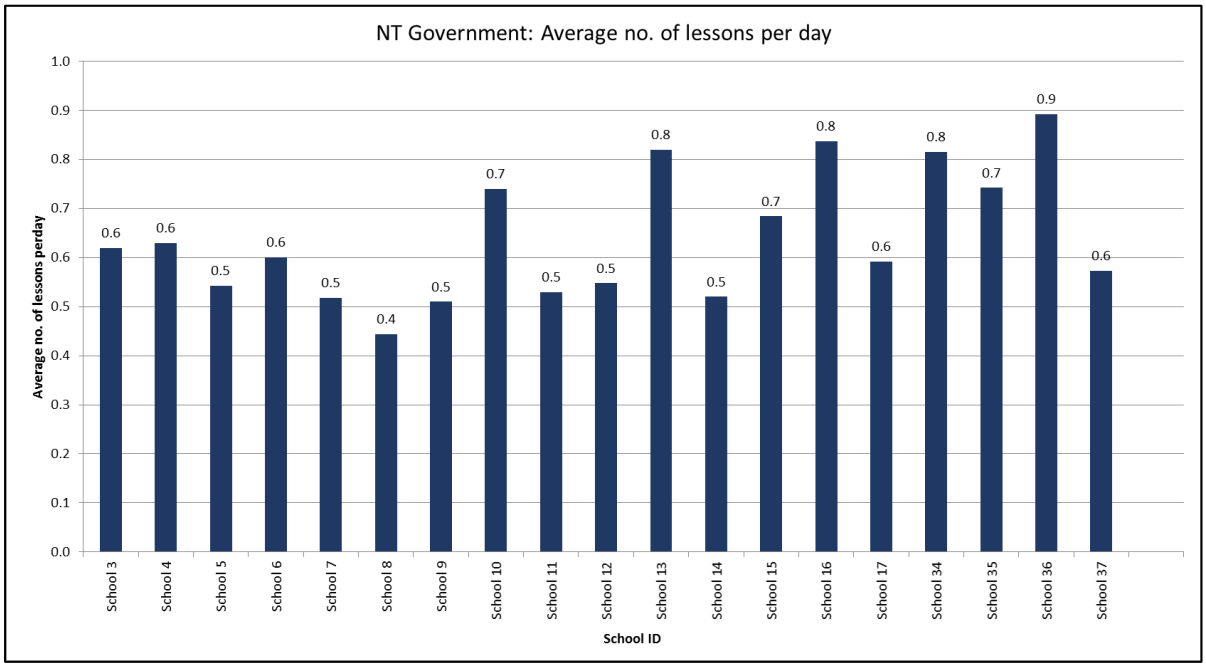


Figure 4. **Average Number of Lessons Delivered Per Day for NT Program Schools.**

The teacher survey asked respondents how long they were teaching literacy per day (on average) with all respondents indicating at least 2 hours and up to 3 hours. The majority indicated that literacy was taught for 2.5 hours per day. Respondents were also asked to indicate if they were able to cover all the lesson content in the time available. Of the 9 respondents, 4 indicated that they had enough timetabled time to cover lesson content and 5 indicated that they did not have enough time to cover the content. Those who responded 'no' to this question were asked to indicate what factors prevented them from completing lesson. Responses included irregular student attendance, lack of student engagement, behaviour management, teaching assistants who were not proficient, teacher training, and skill level.

STUDENT PROGRESS

DI program data for students provides indicators of student progress against mastery assessments. This information has been collated at the individual school level and is presented in *Figure 5*. This graph indicates that half of schools are testing at least 70% of students to determine whether they should progress to the next lesson level. Where students are not tested regularly, these figures appear to be impacted by irregular student attendance. Of those students tested, a majority of students are passing the assessments; however, this measure ranges from 57% to 98%.

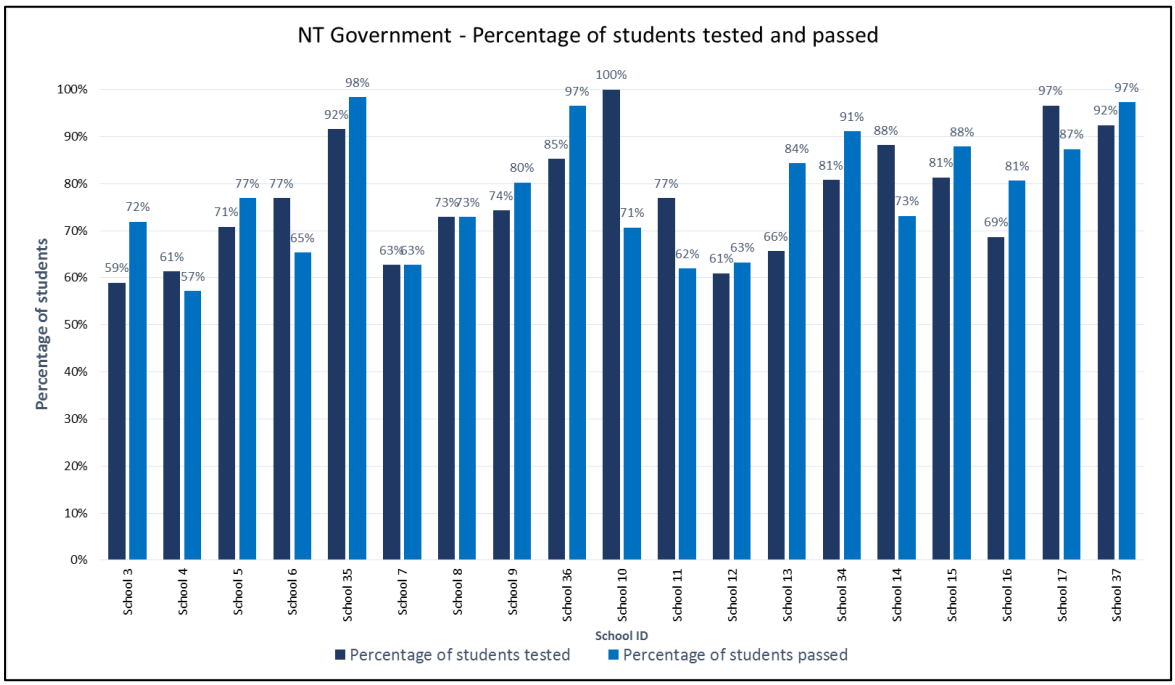


Figure 5. **NT Program Schools Percentage of Students Tested for Mastery and Percentage of Students Passed Mastery Assessment.**

IMPACT: STANDARDISED ASSESSMENT

The following section presents the results of standardised assessments that are considered indicators of impact on student outcomes.

JURISDICTIONAL LITERACY DATA

The NT Government did not provide jurisdictional literacy data (PAT-R results) to the evaluation team, hence, the focus of impact from standardised assessments for NT schools in this report is on NAPLAN.

NATIONAL ASSESSMENT PROGRAM FOR LITERACY AND NUMERACY (NAPLAN)

NAPLAN assessment scores formed part of four separate analyses:

- Cohort tracking compared with control schools between 2014 (pre-FLFRPSP implementation) and 2016 (post-FLFRPSP implementation)
- Change in each year level NAPLAN average scores over time (e.g. Year 3 NAPLAN for years 2014, 2015 and 2016)
- Percentage of students below National Minimum Standard (NMS) over years 2014, 2015 and 2016 for Grades 3 and 5
- Average change in NAPLAN scores between 2014 (pre-FLFRPSP implementation) and 2016 (post-FLFRPSP implementation)
- NAPLAN participation rates by year and grade level.

Each of these analyses is now presented for NT program schools.

NAPLAN: PROGRAM SCHOOLS AND CONTROL SCHOOLS

To explore change in NAPLAN over time, within and between the program and control schools, a series of independent *t*-tests were conducted comparing NAPLAN data from year 3 in 2014 and year 5 in 2016, for the NAPLAN domains of Reading, Writing, Spelling, and Grammar and Punctuation. For this analysis, change variables were created which indicate the change in NAPLAN scores between 2014 and 2016 (Change Variable = 2016 – 2014).

After this analysis was conducted and to better display results, mean scale scores for each domain were displayed for 2014 and 2016 compared to the National Average and Very Remote National Average; these graphs can be found in *Figure 6*.

Preliminary assumption testing did detect some assumption violations, with the presence of minor outlier, and skewness/kurtosis between greater then ±1; however, there was homogeneity of variances for the NAPLAN scores for control and program schools, as assessed by Levene's test for equality of variances. Due to these assumption breaches sensitivity testing will also be conducted using Mann Whitney U tests. Further, as multiple analysis were conducted a bonferroni correction was applied to adjust for Type 1 error (detecting a difference when none is present), significance values have been set to less than 0.01.

Overall, some differences were observed between control and program schools within the NT Government Schools cohort across the four NAPLAN domains. More growth was detected between 2014 and 2016 for controls schools (M = 89.10, SD = 91.90) compared to program schools (M = 47.61, SD = 100.66) on NAPLAN reading change $t(184) = 2.91$, $p < .01$, Cohen's $d = 0.43$[5]. This was also observed for NAPLAN writing change

[5] However, this was not confirmed with sensitivity testing (U = 4582, p = .68)

$t(188) = 3.92$, $p < .001$, Cohen's $d = 0.62$[6] with control schools ($M = 116.63$, $SD = 70.68$) compared to program schools ($M = 70.68$, $SD = 76.95$) experiencing higher growth. However, there were no differences on spelling change $t(195) = -0.96$, $p = .34$, and grammar and punctuation change $t(195) = -0.29$, $p = .77$. These results suggest that both the control and program schools progressed at a comparable rate for students in year 3 in 2014 with the exception of NAPLAN writing.

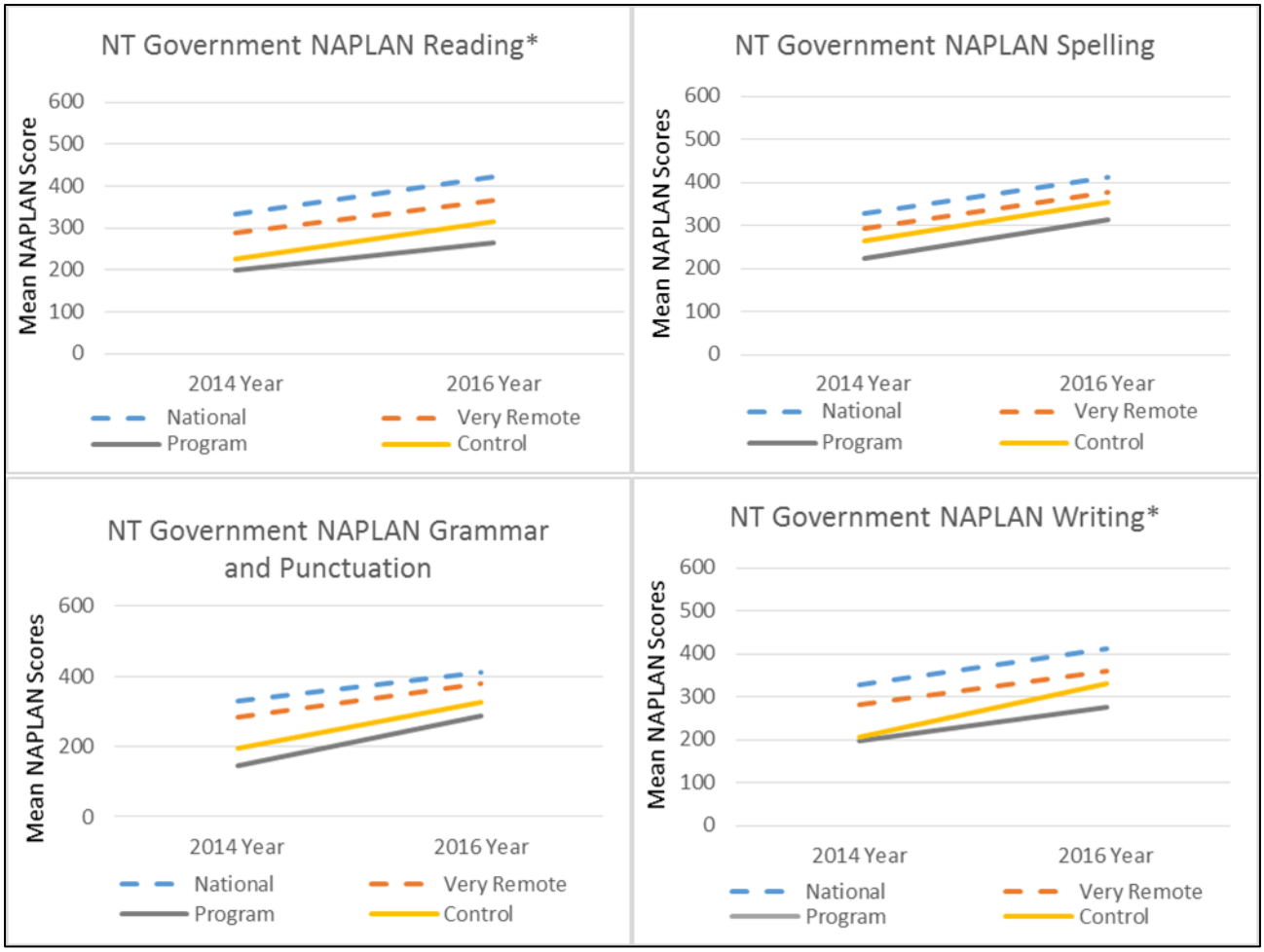


Figure 6. **Change in Mean NAPLAN Scores for NT Government Program Schools Against Control, National Average and Very Remote Average.**

NAPLAN: VARIABILITY

To explore changes in student NAPLAN scores over time mean scores for each NAPLAN subscale were created for Year 3 and 5 for each school for 2014, 2015 and 2016. A One-way ANOVA was conducted to compare scores over this period. Preliminary assumption testing detected outliers, as assessed by inspection of a boxplot, and breaches the assumption of normality, as assessed by Shapiro-Wilk's test ($p > .05$), as well as values greater then ±2. As such, sensitivity testing was also conducted using Kruskal-Wallis H tests.

The results indicate no significant difference[7] over 2014, 2015 and 2016, for any NAPLAN domains in the NT Government cohort; reading $F(2, 38) = 0.47$, $p = .63$; writing $F(2, 38) = 0.31$, $p = .73$; spelling $F(2, 38) = 0.75$, $p = .48$; or grammar and punctuation $F(2, 38) = 2.511$, $p = .10$. Despite there being no significant differences, for

[6] Mann Whitney U test confirms these results ($U = 2977$, $p < .001$)

[7] Confirmed by Kruskal-Wallis H tests.

NAPLAN Reading, Writing, and Grammar and Punctuation, overall variability in these scales reduced over time, which reflects a decrease in the gap between higher and lower scoring students (see *Figure 7*).

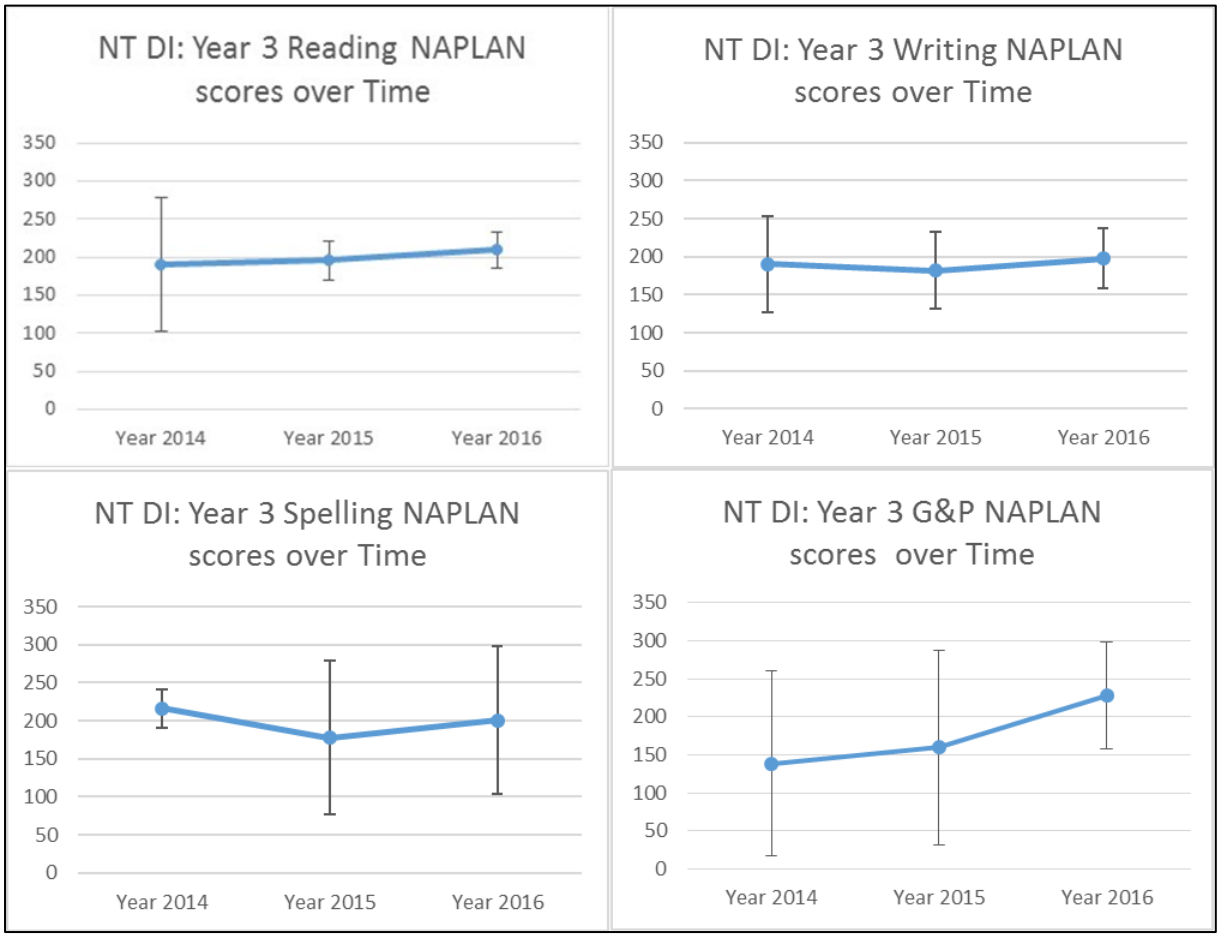


Figure 7. **NT DI Mean NAPLAN Scores 2014 to 2016.**

NAPLAN: PERCENTAGE OF STUDENTS BELOW NATIONAL MINIMUM STANDARD

Analysis of the change in the percentage of students below national minimum standard (NMS) for years 2014, 2015 and 2016 and for Grades 3 and 5 are presented in *Figure 8* and *Figure 9*. For Grade 3, the results are mixed. Reading and Grammar and Punctuation showed an initial decrease and then stabilised at this percentage level. However, both Writing and Spelling demonstrated an increase in percentage of students below NMS in 2016. In general, Grade 5 demonstrated quite high percentages of students below NMS. For Reading, there was an initial decrease in the percentage in 2015 but this percentage increased in 2016. Overall these results do not suggest that there have been any material improvements in the percentage of students below NMS. It should be noted however, that this analysis is based on relational, cohort reliant data and may reflect overall changes to NAPLAN scores at a national level. Further, at this stage of the program, it would not necessarily be expected that these figures would appreciably change.

Figure 8. **Percentage of Students Below NMS for Years 2014 to 2016, Grade 3.**

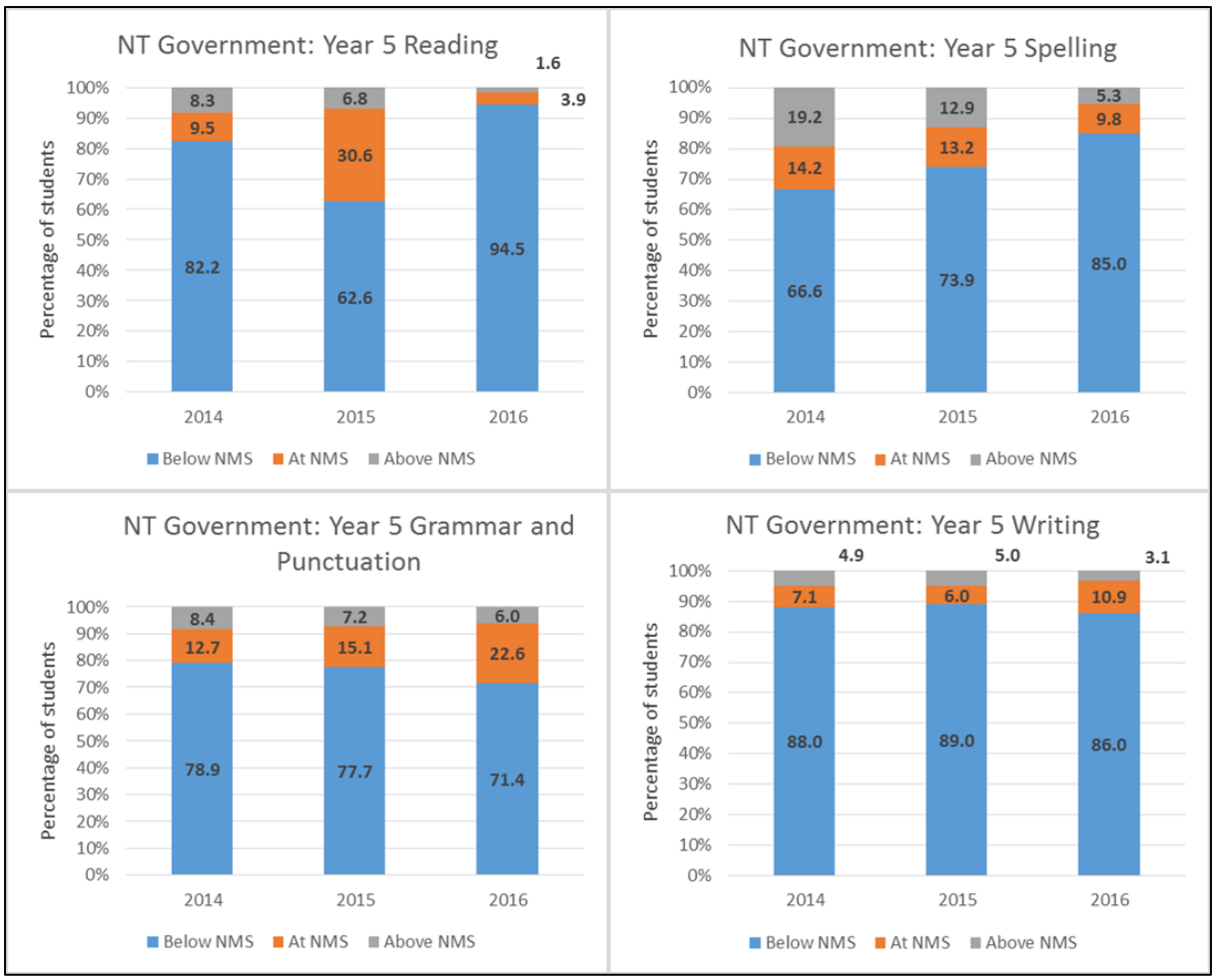


Figure 9. **Percentage of Students Below NMS for Years 2014 to 2016, Grade 5.**

NAPLAN: INDIVIDUAL SCHOOL ANALYSIS

At the individual school level, NAPLAN analysis focused on the mean (average) change in scores between Year 3 and Year 5 for the same cohort (unmatched). This was undertaken for all NAPLAN assessments and is presented in *Figure 10*. A majority of schools in the program in NT demonstrated positive mean change in NAPLAN scores with several exhibiting scores above the NT state average, depending on the relevant NAPLAN assessment. These results reflect expected outcomes for NAPLAN after 2 years of program implementation for increases between Year 3 and Year 5.

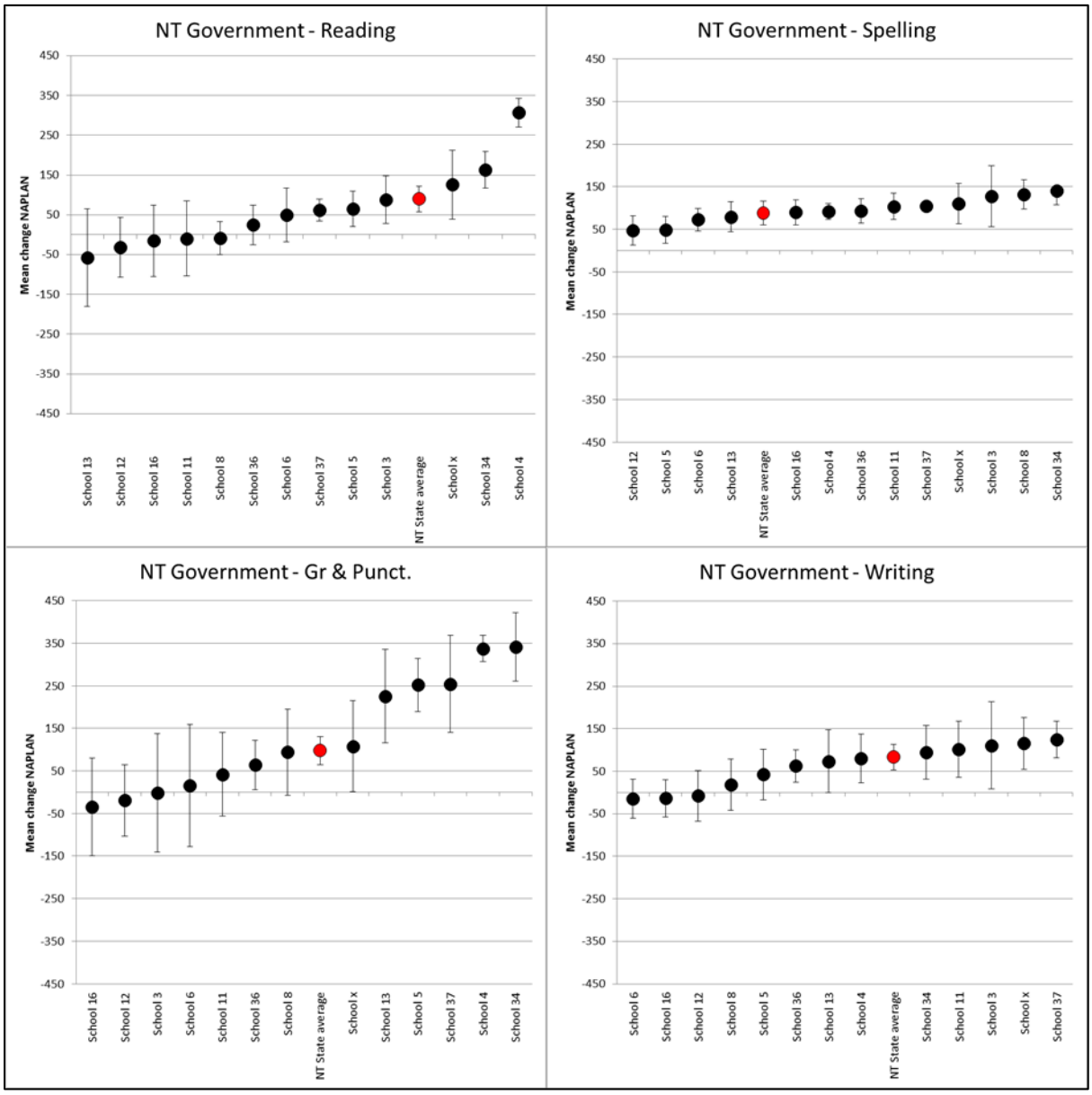


Figure 10. **NT Schools NAPLAN Change Year 3(2014) to Year 5(2016) by School.**

An average effects size for all NT Government Schools (n = 13) was also calculated based on the difference between on the mean (average) change in scores between Year 3 (2014) and Year 5 (2016). Schools demonstrated a medium to large effect size for change over two years; NAPLAN Reading Hedge's g = 1.11; NAPLAN Writing Hedge's g = 0.94; NAPLAN Spelling Hedge's g = 2.31; NAPLAN Grammar and Punctuation Hedge's g = 1.12.

NAPLAN: PARTICIPATION RATES

NAPLAN participation rates were analysed to determine if increases were evident since FLFRPSP commenced in 2015 (see *Figure 11*). Results did not indicate a positive trend over the years 2014 to 2015 for NT program schools. Nonetheless, it should be noted that for the NT DI schools, in 2016 participation was higher than the state average in all assessments. Writing participation was equivalent to the state average.

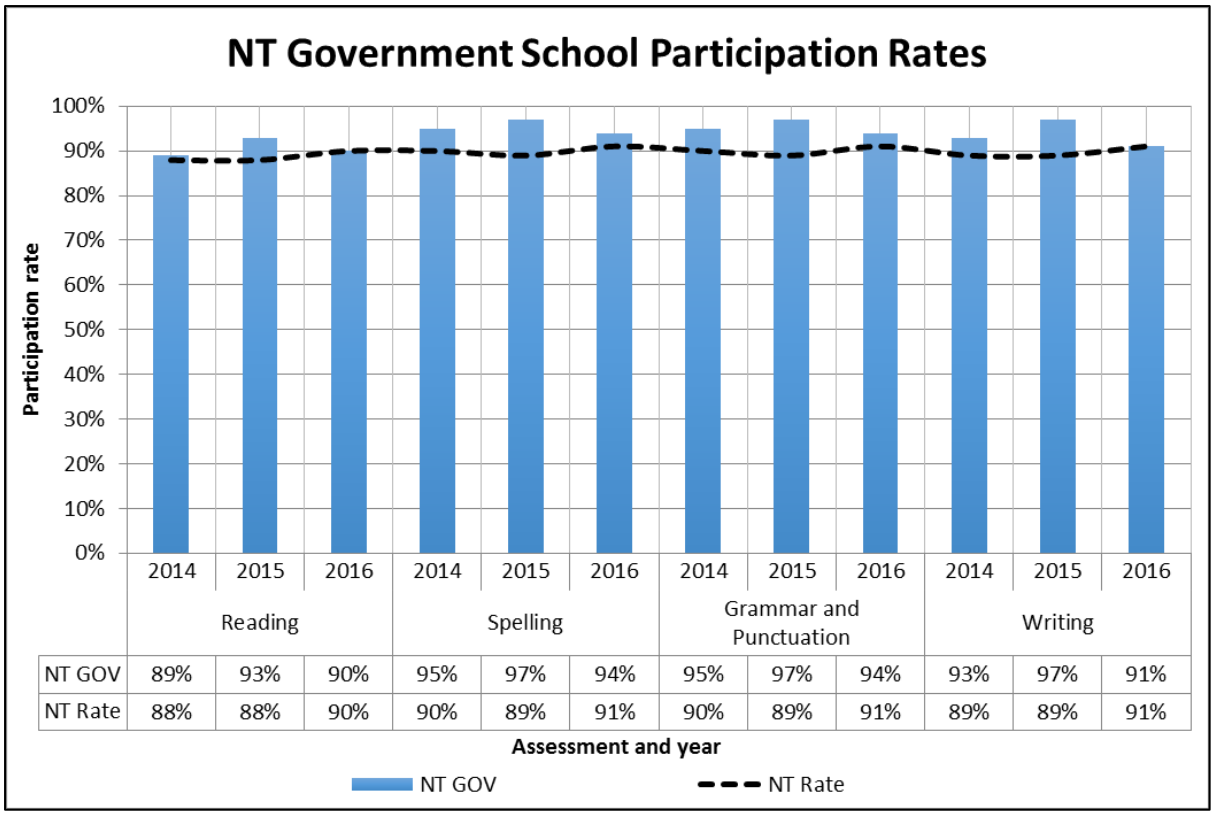


Figure 11. **NAPLAN Participation Rates by Assessment Type and Year for NT Program Schools.**

IMPACT: OTHER DATA SOURCES

TEACHER SURVEY

Nine NT teachers participated in a survey. The evaluation team intends to continue interviewing teachers and attending school visits and observations to gather a larger number of perspectives and experiences of school staff in the program.

All teachers were currently using the DI approach to teach literacy in their schools. The teachers surveyed were typically very experienced, with five teachers having worked in schools for over 15 years, two for 10-15 years, and one between 2 and 5 years.

As part of the survey, teachers were asked to provide their perceptions of the impact of FLFRPSP on students in their school. These questions considered factors such as literacy skills, engagement, wellbeing, and whether students liked the program. *Figure 12* shows the results of these questions for NT schools. The majority of teachers (78%) felt the program had improved students' literacy and that students were engaged with and generally liked the program. One teacher commented,

"It has done what it claims to be able to do. It has enabled the teachers to raise the literacy levels at our school. We are producing competent confident readers."

In terms of other outcomes for students, a majority of teachers felt the program had improved student wellbeing and, in general, had made a difference for students in their learning. Some of the positive outcomes of the program identified by teachers were reduced negative behaviours, increased confidence in other learning areas, less need for scaffolding, and pride in showing their families what they can do. Teachers who felt that the program had not improved other outcomes for students reported that students had less time to do writing tasks, and in one instance, a teacher reported that they felt that student behaviour had suffered because of the program.

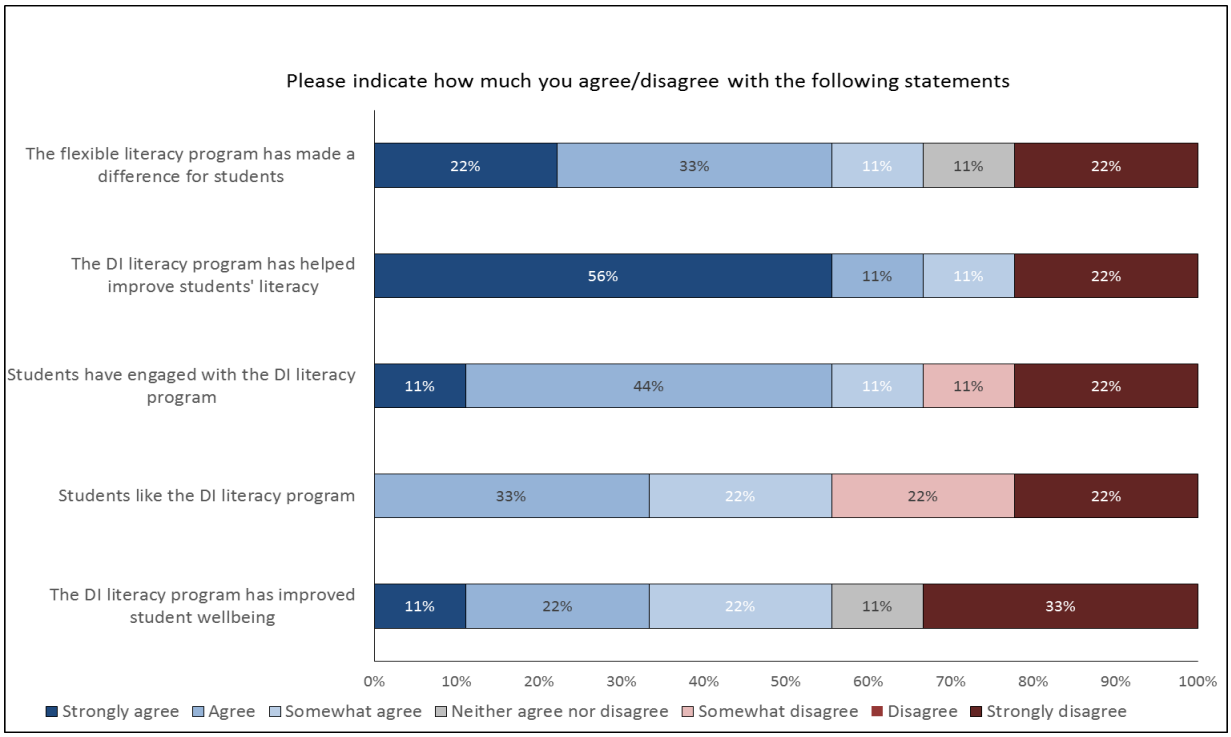


Figure 12. **Teacher Perception of the Impact of the Program on Students (n=9).**

PRINCIPAL INTERVIEWS

Five Principals were interviewed by the evaluation team about their perceptions of the impact of the program on student literacy and other outcomes. The evaluation team intends to continue interviewing Principals and attending school visits and observations to gather a larger number of perspectives and experiences of school staff in the program.

In general, the NT Principals felt that the program was helping students in terms of literacy skills, but that this was reliant on attendance. That is, there were stronger gains in students who attended school all the time. Those students with intermittent attendance found it harder to reach mastery and often the schools did not have the teaching resources to catch these students up. Nonetheless, two Principals made the comment that the structure of the program did allow for those students who attended less regularly to understand the routine and fit back into it even if they were not achieving as much as they should. One Principal commented,

"The ones that are coming every day, yes it has, but I am pushing every moment especially in the early years – so we can see a huge difference in the kids that are coming up in the early years."

Another Principal made the comment that DI was a ***"strong pedagogical approach to literacy"*** and felt that it had benefited reading results and spoken English.

However, one Principal felt that the program was very prescriptive and not a great fit for their students, commenting,

> ***"The kids who are progressing well are the ones who come to school every day and who would progress regardless. …It's not a good fit for my community or my school."***

In terms of other impacts such as wellbeing and engagement in school, Principals noted that they had seen improvements in classroom behaviour and the students responding positively to routine and familiarity after initial settling-in periods.

DOCUMENT ANALYSIS

Secondary documentation related to schools in the NT suggested mainly positive outcomes, such as literacy progress, student engagement, proficiency in English, and attendance. Two schools noted negative factors that impacted their implementation of DI. Tiwi School cited attendance as "the biggest challenge" in implementing DI, and another found that DI was "pretty tricky for the NAPLAN testing because it doesn't fit in with that we're doing" (Terzon, 2015). Regardless of these negative factors, both schools overwhelmingly reported positive results with DI/EDI.

Literacy progress. Several schools reported increased literacy outcomes as a result of DI, but the information was non-specific. One Principal observed increased outcomes since DI was implemented (Considine, 2016b). Another school saw excellent results in the first 18 months of implementation (Considine, 2016c). Hard work and dedication from teachers resulted in students "making leaps and bounds towards their education", even passing on their learning to their families. One school found that DI was very helpful in increasing students' proficiency in spoken and written English(Aikman, 2016). The principal of another school noted that DI resulted in solid, gap-free learning, as "it's really explicit and there's no room for waffle"(Terzon, 2015). Still another school reported "wonderful progress", with one student reading in front of the whole school for the first time.

Student engagement. Aspects of student engagement mentioned in the literature seemed to focus on student confidence and attention in NT schools. One school saw students become "more confident and engaged in classroom activities", according to their principal (Chandler, 2016a). Another principal believed that her students had become driven by the successes they achieve in lessons every day. Students there have also benefitted from having a routine that allows them to settle quickly and learn. Tiwi School found that students are able to concentrate for longer periods of time. One school found the structure and continuity of DI to be helpful in facilitating student learning and attention; students from there were forced to attend a school in Darwin when their school was evacuated during a cyclone. As the Darwin school also utilised DI, these students were able to continue their lessons seamlessly due to DI's consistency (Education HG News Team, 2015).

Proficiency in English. Although DI/EDI is targeted at English-language literacy, proficiency specifically in Australian English or Aboriginal English was a prominent theme in the NT documentation, as many students' first language is an Indigenous language. Increasing proficiency in English is a priority in NT schools. The Indigenous Education Strategy Newsletter, Issue 3, reports principal and teacher reports of DI's benefits: increased English use, increased English proficiency, increase of on-task behaviour, and little or no loss of learning after holidays (Considine, 2016a). One school noted that students have shown a growing ability to switch between their native language, Aboriginal English, and Standard Australian English.

TEACHERS

For teachers in the NT program schools, the average years of experience is 11.64 years, indicating a relatively experienced workforce. However, teacher turnover is high in NT program schools. *Figure 13* and *Figure 14* show turnover rates by schools in the program for both teachers and teaching assistants for the 2015 and 2016 school years. On average in 2016, schools lost over half their staff of teachers (58%) and three quarters of their teaching assistants (77%). These figures were even higher in 2015, 161% and 168% respectively. This presents a significant challenge for schools and school leadership, and was also raised directly by teachers and principals participating in the survey and interviews. Staff attrition is a major factor influencing the implementation and the impact of the program in NT.

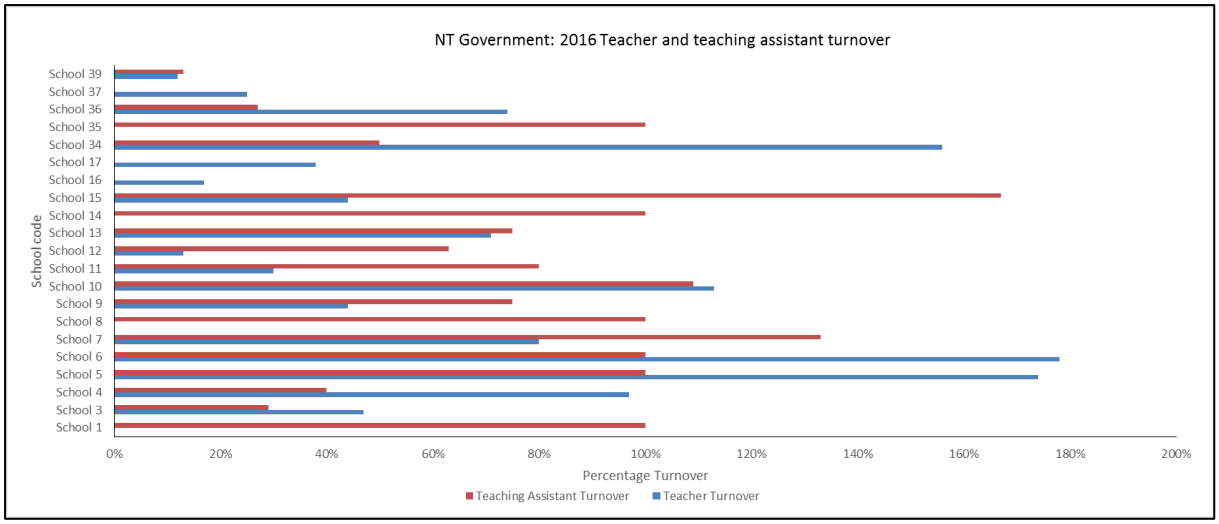


Figure 13. **Teacher and Teaching Assistant Turnover 2016 for NT Program Schools.**

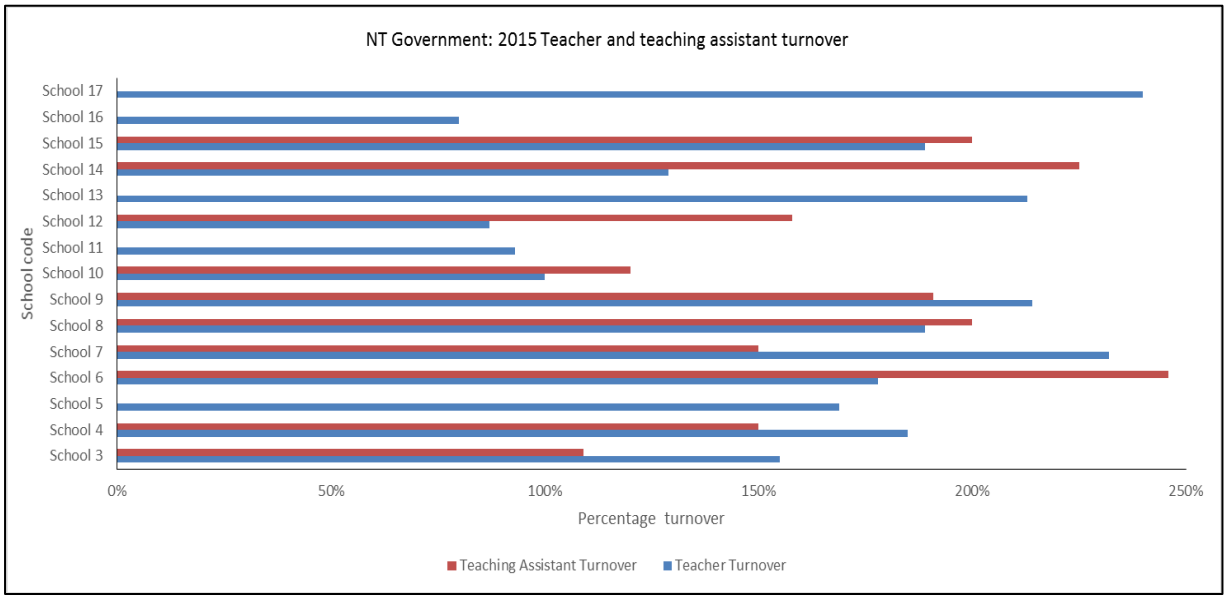


Figure 14. **Teacher and Teaching Assistant Turnover 2015 for NT Program Schools.**

The DI program data provides a measure of teacher effectiveness across three dimensions: classroom organisation, instructional delivery, and behaviour management. Classroom organisation refers to the recording of data, material, and use of allocated time. Instructional Delivery is a broader concept and considers the use of specific instructional activities, including use of remedies, assessments, correction of errors, use of repetition, pacing, and use and effectiveness of signals. Behaviour Management considers between-task transitions, teacher engagement during independent and group work, motivations, and positive behavioural support.

Figure 15 shows the development of teacher effectiveness as measured by the program over Terms 1, 3 and 4 in 2015 and Terms 2, 3 and 4 in 2016. Note that data was not available for Term 2 2015 and Term 1 2016. Teacher effectiveness demonstrated a large increase in 2015 on all three measures, in particular, classroom organisation. This is also reflected in Principal interview data, which indicated that implementation was difficult in the first Term, but became easier and more streamlined as the year progressed. In 2016, teacher effectiveness appears to have plateaued. These results may reflect the significant turnover of staff experienced by program schools and therefore the challenge of maintaining teacher effectiveness scores.

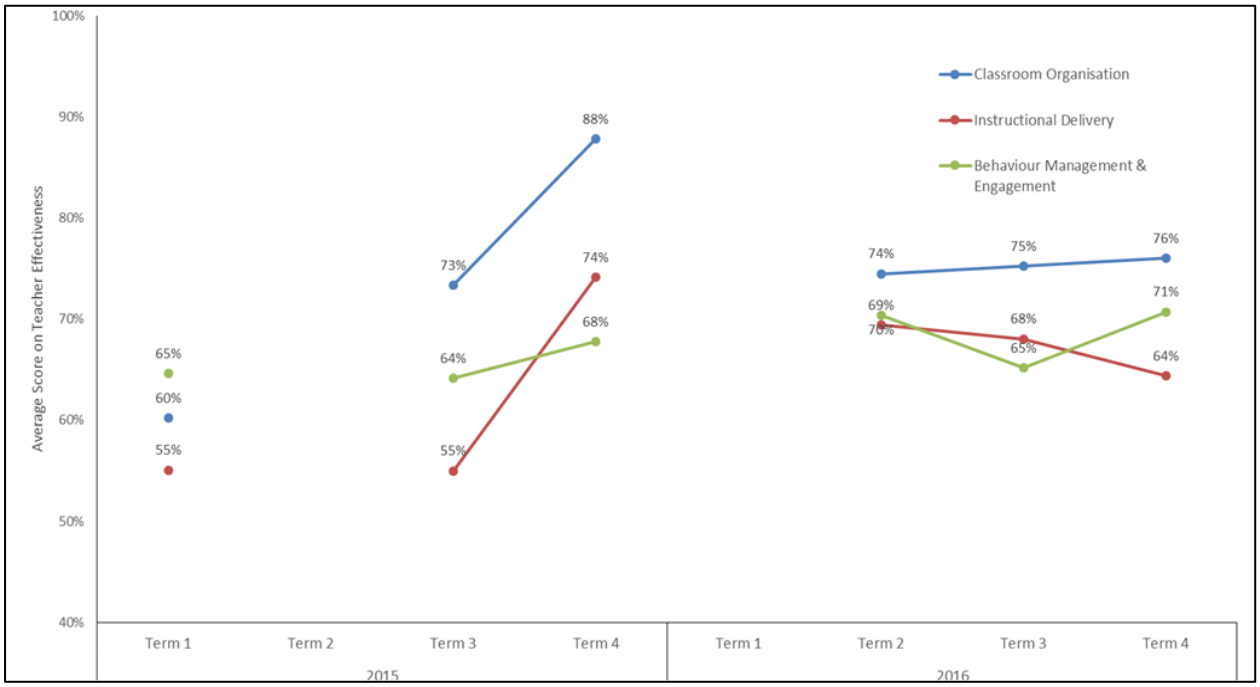


Figure 15. **Teacher Effectiveness Measures from DI Program Data for NT Program Schools.**

As part of the survey, teachers were asked to rate their perception of the impact of the program on their pedagogy, their readiness to teach following training, and their knowledge in how to teach literacy. *Figure 16* and *Figure 17* show the results of these questions. The majority of teachers were clear regarding their role within the FLFRPSP program, the specific pedagogical approach and the specific tasks and activities that they were required to instruct. With respect to other impacts of the program on their teaching, half of the respondents felt that the FLFRPSP program improved their teaching generally and in literacy, and felt it was a good match for their school.

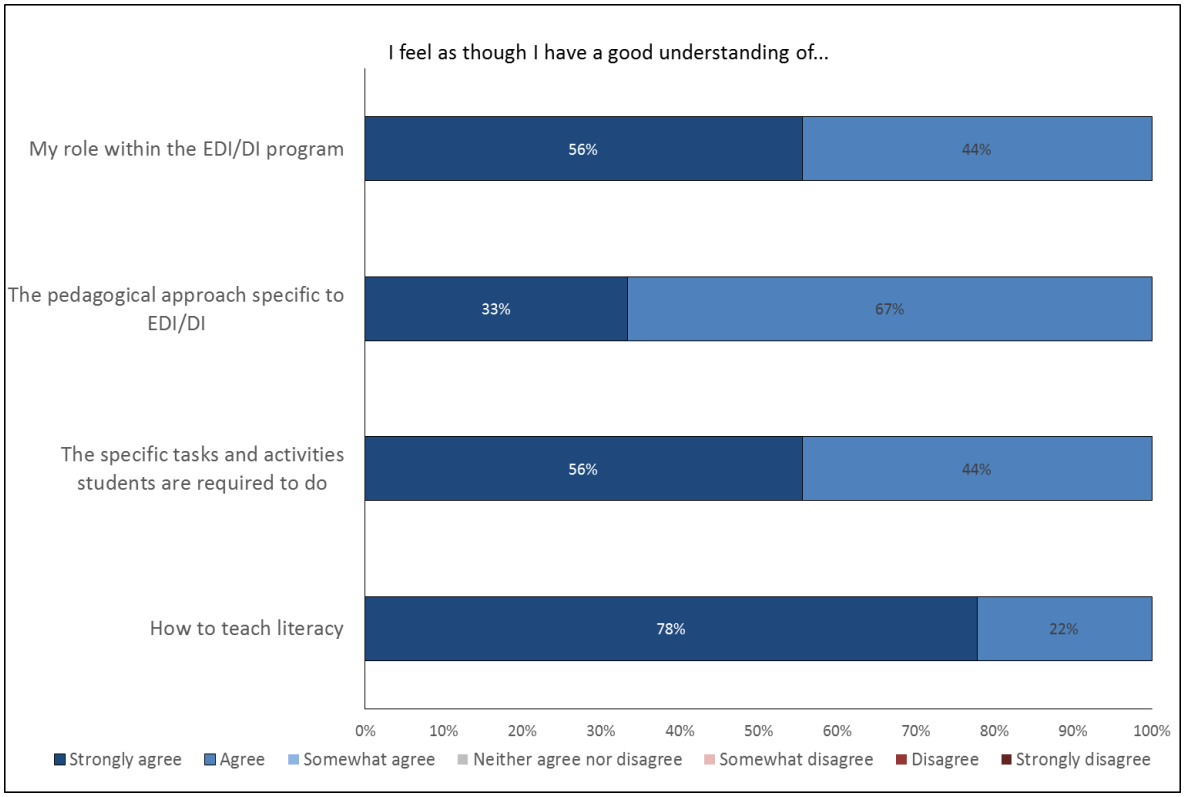


Figure 16. **Teacher Perception of the Program's Impact on their Role and Work as Teachers (n=9).**

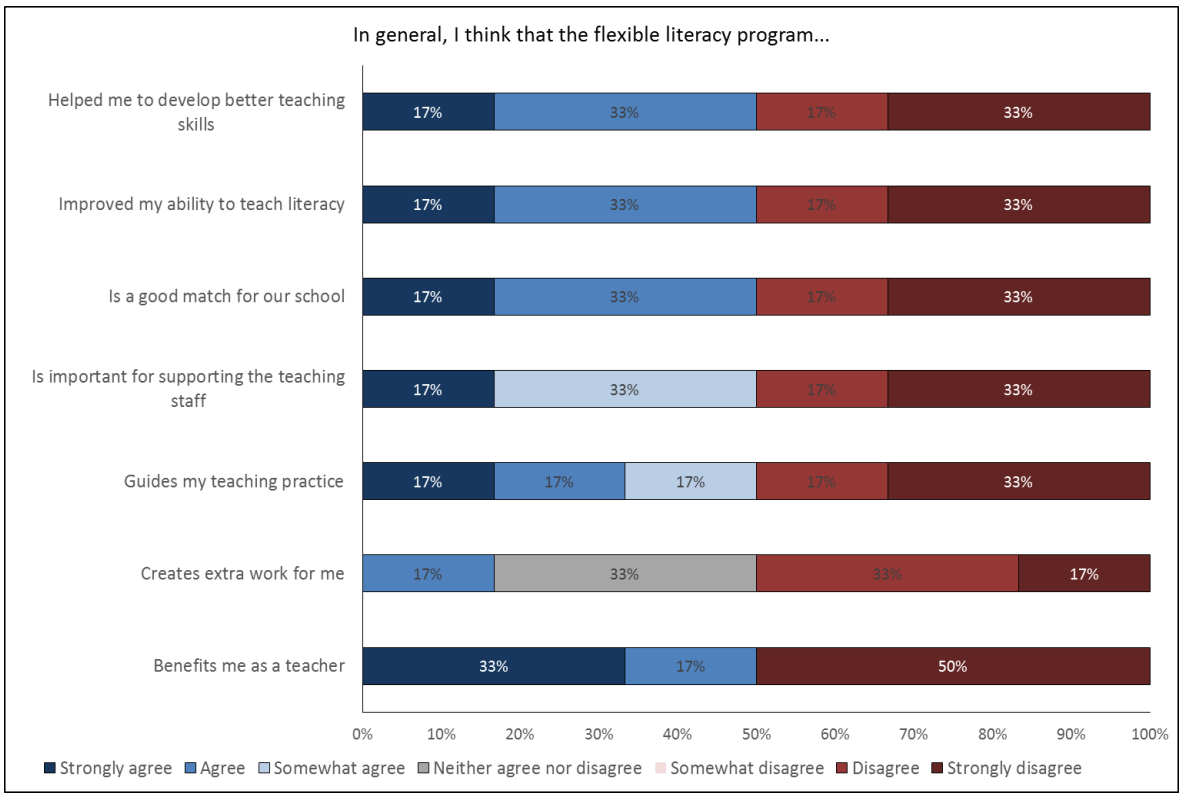
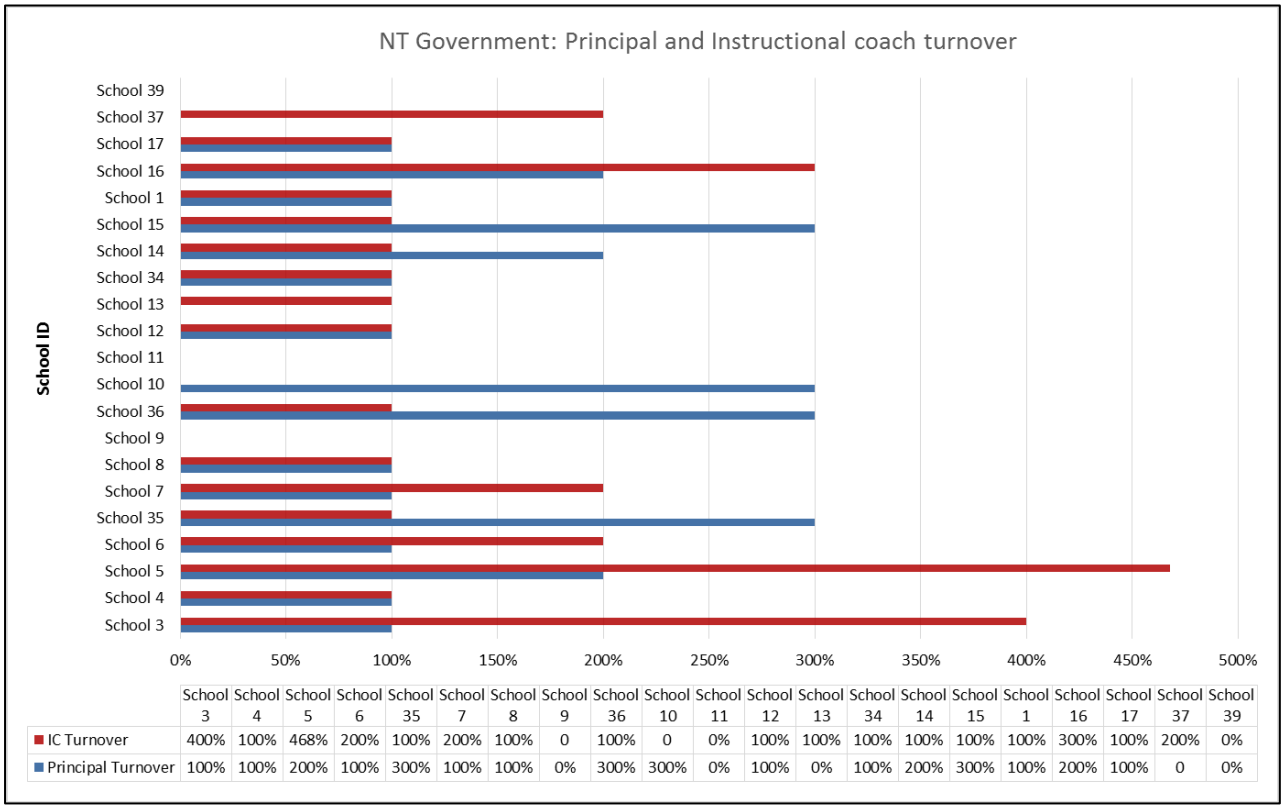


Figure 17. **Teacher Perception of the Program's Impact on their Pedagogy (n=9).**

SCHOOL LEVEL DATA

In addition to several indicators of program impact and implementation, relevant demographics have been considered at the school level. Demographic data includes publicly available information such as that available on the MySchool website as well as provided information such as turnover in leadership staff. Program data includes measures of school fidelity (explained further below) and program support provided by GGSA through training and observations. Finally, perceptions of impact from teachers and principals are also included where relevant to the school as a whole.

Figure 18 shows turnover rates for principals and instructional coaches by DI school in the NT between 2015 and 2017. Instructional coaches are school-based staff that provide on-site coaching support to teaching personnel. Note that in the data table '0' refers to no data whereas '0%' reflects no turnover. Turnover of these leadership roles with respect to the program is high across most schools with the majority replacing a leadership role at least once per year. This has implications for the FLFRPSP in terms of continuity, advocacy, and implementation.



	School 3	School 4	School 5	School 6	School 35	School 7	School 8	School 9	School 36	School 10	School 11	School 12	School 13	School 34	School 14	School 15	School 1	School 16	School 17	School 37	School 39
IC Turnover	400%	100%	468%	200%	100%	200%	100%	0	100%	0	0%	100%	100%	100%	100%	100%	100%	300%	100%	200%	0%
Principal Turnover	100%	100%	200%	100%	300%	100%	100%	0%	300%	300%	0%	100%	0%	100%	200%	300%	100%	200%	100%	0	0%

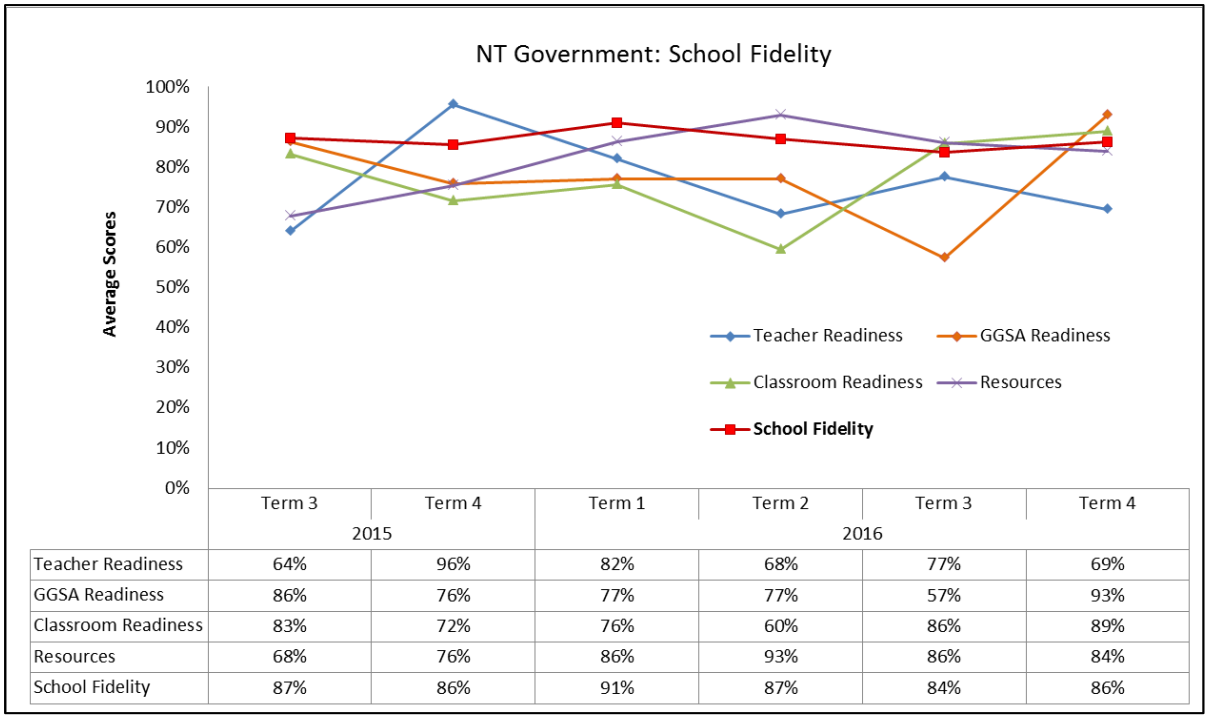
Figure 18. **Principal and Instructional Coach Turnover 2015 to 2017 for NT Program Schools (where provided).**

IMPLEMENTATION

At the school level, understanding of implementation is measured primarily with DI program data. School fidelity comprises four factors: teacher readiness, GGSA readiness, classroom readiness, and resources that are then combined into an overall score. Table 8 describes each of these factors in more detail and *Figure 19* shows scores for all factors and overall across 2015 and 2016. There is some variability across all four factors over time. GGSA readiness was consistently high and indicates that support and monitoring from GGSA is well implemented. Teacher readiness scored lowest of the four factors, which again may reflect teacher turnover rates, which have a flow on effect to other school-based factors such as classroom readiness and resource readiness.

Table 8: *Description of School Fidelity as Measured by DI Program Data*

Factor	Number of items	Description of underlying items
Teacher readiness	3	Teaching staff are trained, teachers rechecked, LPC and STS form completion by staff
GGSA readiness	4	Communication systems established, process established to discuss data, GGSA visits are planned, calls to review data occur weekly
Classroom readiness	6	Students are placement tested, students placed into instructional groups, group sizes within guidelines, sufficient time schedule for each group, classroom arrangement adequate for effective instruction, staff assigned to groups
Resource readiness	2	Teachers use full allocated time, schools have instructional materials to deliver DI program
Overall school fidelity	4 scales	Average of the above 4 School readiness measure scores



	Term 3 2015	Term 4 2015	Term 1 2016	Term 2 2016	Term 3 2016	Term 4 2016
Teacher Readiness	64%	96%	82%	68%	77%	69%
GGSA Readiness	86%	76%	77%	77%	57%	93%
Classroom Readiness	83%	72%	76%	60%	86%	89%
Resources	68%	76%	86%	93%	86%	84%
School Fidelity	87%	86%	91%	87%	84%	86%

Figure 19. **School Fidelity Measures for NT Program Schools 2015 and 2016.**

IMPACT

Several questions in the survey of teachers addressed the potential impact of the program at a school level. These questions focused on the drivers of implementation and aspects of the program that were easy or difficult to implement.

NT FLFRPSP program teaching staff identified several positive program drivers including the support provided from GGSA and NIFDI, the data generated by the program as evidence of learning and progress, and the week-long training program at the beginning of each year. One teacher also mentioned that support from a regional

literacy officer was helpful. Drivers that did not support the program included poor student attendance, lack of community consultation, and lack of regular teaching assistants.

The principal comments from the interviews were similar to teachers in terms of positive views of GGSA and NIFDI support. Several felt that achieving the prescribed 2.5 hours per day was a significant challenge with some schools reducing the time allocated to be more realistic about what could be achieved. Some schools were also grappling with waning community support and were making adaptations to the program to accommodate community views. Despite these challenges, nearly all principals conceded that the program was working, particularly when students attended regularly and staff had bought in to the pedagogical approach.

PROGRAM SUPPORT

Program support refers to implementation and monitoring assistance provided by GGSA, in-school coaches or program developers (NIFDI, DataWorks). Training data is the average number of training sessions per school, and includes behaviour training, DI program training, and support training for teachers, teaching assistants, and school leadership variously. Observations are the average number of observations per school, and consist of two-minute observations, five-minute observations, and extended observations. These observations were conducted by an implementation manager, principal, instruction coach, teacher coach, or teaching principal. The NT data shows there to be a relatively high number of both training and observations. Despite variation in the observation numbers across the terms, the number of training sessions remained stable.

Figure 20 shows the average number of observations and average teacher training for NT program schools in 2015 (Terms 3, 4) and 2016. Training data is the average number of training sessions per school, and includes behaviour training, DI program training, and support training for teachers, teaching assistants, and school leadership variously. Observations are the average number of observations per school, and consist of two-minute observations, five-minute observations, and extended observations. These observations were conducted by an implementation manager, principal, instruction coach, teacher coach, or teaching principal. The NT data shows there to be a relatively high number of both training and observations. Despite variation in the observation numbers across the terms, the number of training sessions remained stable.

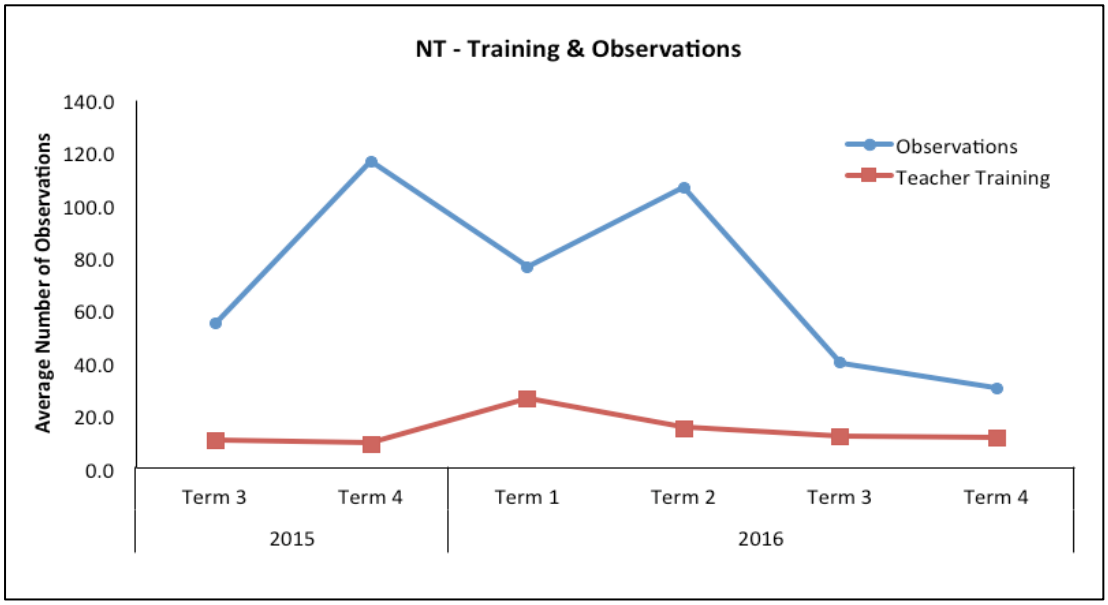


Figure 20. **Average Number of Training Sessions Attended and Coaching Observations by Term for NT Government Schools.**

Teaching staff were also asked in the survey their perceptions of their readiness to teach the program following training. *Figure 21* displays the results of this question for teaching staff in NT Government schools. The majority of staff agreed that they were ready to implement the program, had enough resources, and understood the potential benefits of the program following training. Results were more varied with regard to feeling supported to implement the program. This question did not specify the source of that support. Given that comments made by teachers were positive towards the support from GGSA, this is likely to reflect support from the school. Please note, this last comment applies only to the teachers who responded to the survey and does not necessarily reflect the view of all teacher in NT program schools.

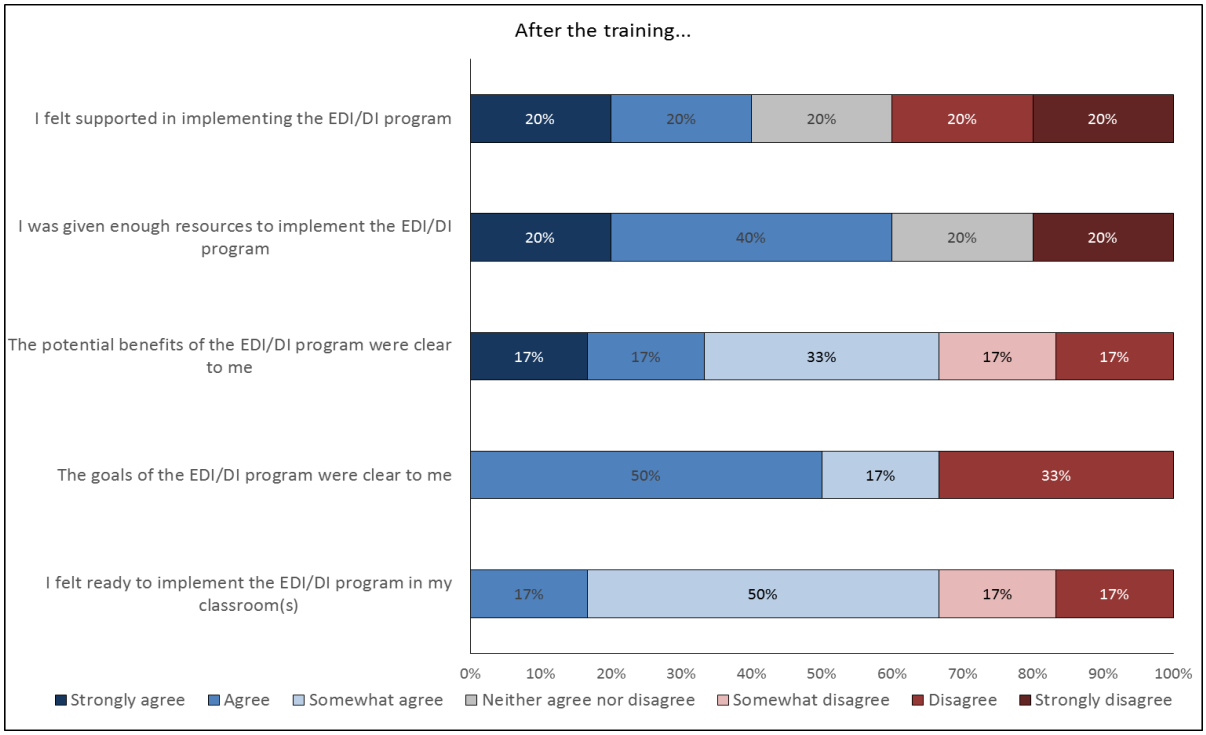


Figure 21. **Teachers' Perception of their Understanding of the Program Following Initial Training (n=9).**

OVERALL SUMMARY OF COLLECTIVE RESULTS FOR NT GOVERNMENT

In summary, there are variable results from NT Government program schools at the student, teacher, and school level.

For students in NT Government schools there are currently mixed results of early impact suggesting that additional data and time is required to fully understand the impact of the program across all NT Government schools. A major hindrance to understanding the impact of the program was the absence of jurisdictional data to indicate changes in student literacy skills. This data is critical to understanding program impact because the content of the PAT-R assessments is most reflective of student classroom learning content. NAPLAN is a more distal indicator of literacy growth and achievement and a measure that requires time to demonstrate impact. Nonetheless, there are aspects of NAPLAN that have utility in the current analysis. First, the results have demonstrated that program schools have not regressed on NAPLAN scores over time. This was evidenced in NAPLAN variability analysis, which showed small (non-significant) gains (other than Spelling) in Year 3 results since the program began. In addition, this analysis showed that variability in scores in NAPLAN reduced over time (for all assessments apart from Spelling), suggesting that students' skills are becoming more consistent within cohorts. As a highlight, in Grammar and Punctuation, the lower level students were increasing skill levels towards the higher performing students. The individual school analyses showed that for Spelling, Writing and Grammar and Punctuation, mean change in NAPLAN scores over time were greater than the

national average mean score change. NAPLAN participation rates demonstrated relative consistency over time and were above State averages.

Teacher survey data and Principal feedback on student impact shared the same mixed valence as the statistical indicators, suggesting that some schools are seeing quite significant positive impact from the program on student literacy outcomes and others are not. Key drivers that arose from most respondents related to the importance of regular student attendance and behaviour management in contributing to the programs overall impact.

At the teacher and school level, turnover is a significant problem, and one that encroaches on implementation and fidelity. Nonetheless, teachers report being well prepared to teach the program and understand the pedagogical approach following initial training. However, they do not necessarily consider that working with the program has influenced their ability to teach more generally, suggesting that their learnt skills in DI/EDI are a discrete set of instructional activities. Program data measuring overall school fidelity against four domains indicate that schools are receiving excellent levels of support from GGSA, whereas teacher readiness, a measure of staff being trained, and other activities require further refinements and improvement.

Identified drivers[8]:

- Lack of regular student attendance (-)
- Behaviour management (-)
- Structure of the program allowed for absent students to re-enter (+)
- Teacher turnover (-)

HEAT MAP OF NT GOVERNMENT PROGRESS IN FLFRPSP

Figure 22 illustrates for each data source the extent to which schools in the jurisdiction are operating below, at, or above expected levels for this stage of the evaluation. This is based on review of all the available evidence and an evaluative judgment by the evaluation team. It provides a visual summary of the status of the jurisdiction and will inform the next phase of the evaluation; it also provides a basis on which progress can be monitored and re-assessed. Overall, it indicates that NT Government program schools are meeting expectations for this phase of the evaluation and has varied results. Further refinements are required including the evaluation team working with NT DET to obtain credible jurisdictional literacy data.

[8] Enablers are identified with a positive sign, while barriers are identified with a negative sign.

	Implementation	Impact		
	Program data	Proximal primary data	Distal primary data	Other primary data
Students	Lesson progress and mastery assessments	No PAT-R data provided	NAPLAN	Teacher and Principal report
Teachers	Training and observation (GGSA provided)	Teacher effectiveness program measures	N/A	Teacher and Principal report
Schools	School fidelity	N/A	N/A	Teacher and Principal report

Figure 22. **Heat Map for NT Program Schools.**

The rating criteria for each cell in the heat map are as follows:

Above expectations for this phase of the evaluation	Meets expectations for this phase of the evaluation (mixed results – some aspects require attention while others are good.	Below expectations for this phase of the evaluation (requires substantial improvement)	Not enough information available	Not applicable N/A

WESTERN AUSTRALIA GOVERNMENT AND INDEPENDENT SCHOOLS

Four WA Government schools are implementing FLFRPSP; 3 DI and 1 EDI. One additional school was involved in the program until Term 2 2016 and taught EDI. The demographics and other relevant data for this school are not included in the current analysis.

WA Government schools assess literacy progress using PM Benchmark Reading, which assesses instructional and independent reading levels through meaningful texts.

DEMOGRAPHIC PROFILE

Table 9 summarises relevant demographic information regarding FLFRPSP schools in WA Government and compares them with the same data for all schools in the program. WA Government schools have a lower mean ICSEA value than the mean for all program schools. Furthermore, WA Government schools have a higher percentage of Indigenous students but a lower percentage of students with a LBOTE. Attendance rates for schools in the program are also lower compared with overall program information. All schools are classified geographically as Very Remote.

Table 9: *Mean Values for Key Demographic Variables for WA Government Schools in the Program with all Program Schools Data as a Comparison*

	Teachers	Staff	ICSEA	Total students	Girls	Boys	Indig %	LBOTE %
WA Gov	9.25	9	660.50	83.75	37.5	46.5	90.50	77.50
PROG	9.94	5.81	666.32	100.71	49.19	52.06	92.00	83.00

Table 10: *Attendance Rates for WA Government Program Schools*

	2014			2015			2016		
	Days Present	Days Absent	Attendance Rate	Days Present	Days Absent	Attendance Rate	Days Present	Days Absent	Attendance rate
WA Gov	94	51	61%	93	53	63%	82	57	59%
Overall	111	44	72%	110	44	73%	108	44	71%

IMPLEMENTATION

As mentioned previously, the DI program is timetabled to be delivered for 2.5 hours per day. This allows for a benchmark lesson progress rate across all groups of 1.0 lesson per day. Program data tracks the lessons taught so that an average lesson per day can be calculated at the school level. *Figure 23* shows this information for the three DI schools in the WA Government group and indicates the average lesson delivery per day was 1.0, 0.70 and 0.70 lessons respectively. This data represents the number of lessons that schools complete each day, on average, and not the hours spent on instruction. It is possible that, for a range of reasons, teachers are unable to complete the full lesson within the time available. For example, students who have been absent and returned to school may need to be re-assessed to determine their lesson level. Hence, this will take time away from normal lesson teaching. While this data shows that schools are not reaching full fidelity in terms of lesson completion, it does not explain why, with further insight offered by the teacher survey.

The teacher survey asked respondents how long they were teaching literacy per day (on average) with all respondents apart from one indicating at least 2.5 hours and up to 3 hours. Respondents were also asked to indicate if they were able to cover all the lesson content in the time available. Of the 4 respondents, 3 indicated that they had enough timetabled time to cover lesson content. The respondent who did not think they had enough time to cover a lesson also indicated that they had a complex group of students who had literacy skills across several grade levels despite being in the same school year (and age group).

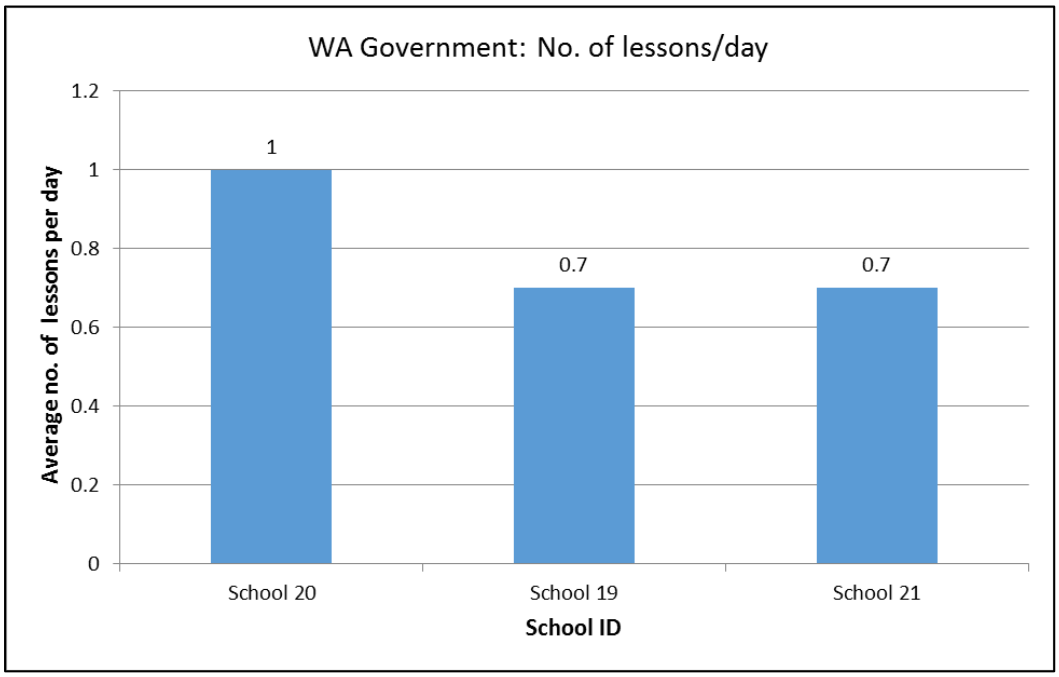


Figure 23. **Average Number of Lessons Delivered Per Day for WA Government Program Schools.**

DI program data for students provides indicators of student progress against mastery assessments. This information has been collated at the individual school level and is presented in *Figure 24*. This graph indicates for School 19 that ¾ of students are tested, with 70% passing and for School 20, 50% are regularly tested with 79% passing. School 21 assessed 61% of students and 74% passed. Again, percentage of students tested is likely impacted by student attendance rates.

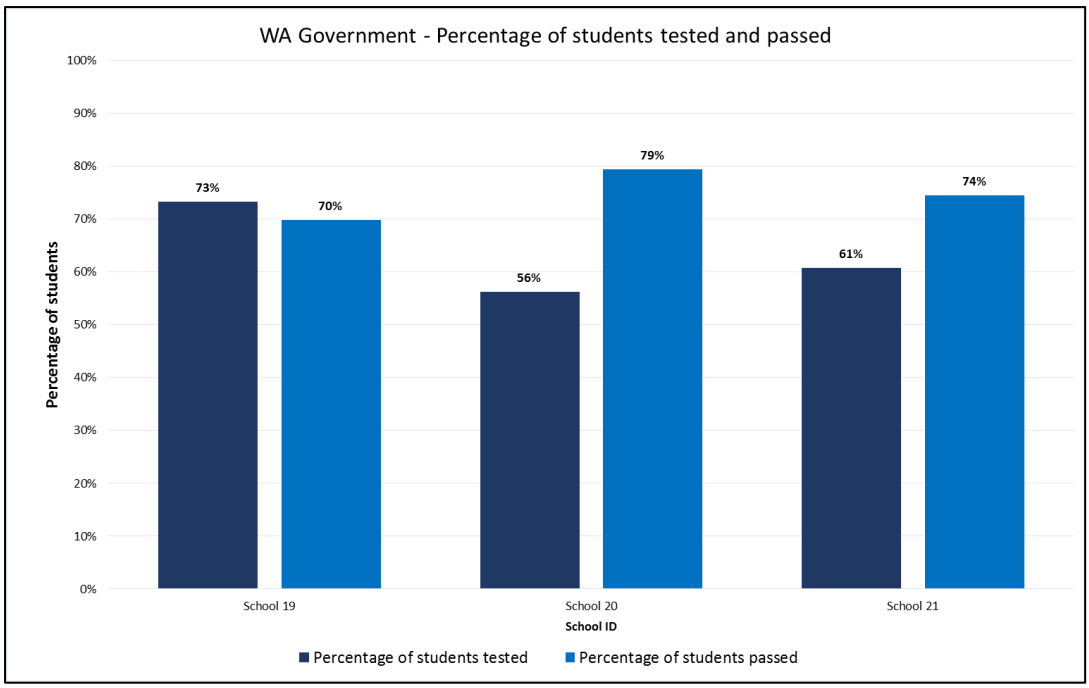


Figure 24. **Percentage of Students Tested for Mastery and Students Passed.**

EDI schools measure student mastery in Weeks 5 and 10 of each term. Percentage scores have been used in this analysis because total scores can differ depending on year level. *Figure 25* graphically demonstrates average student mastery scores for students in EDI WA Government schools over Terms 3 and 4 2015 and for all four terms in 2016. Despite an initial dip in average mastery score in Term 4 2015, scores have remained consistent in 2016 for one school. A full data set was not provided for the second school (in blue) as they exited the program.

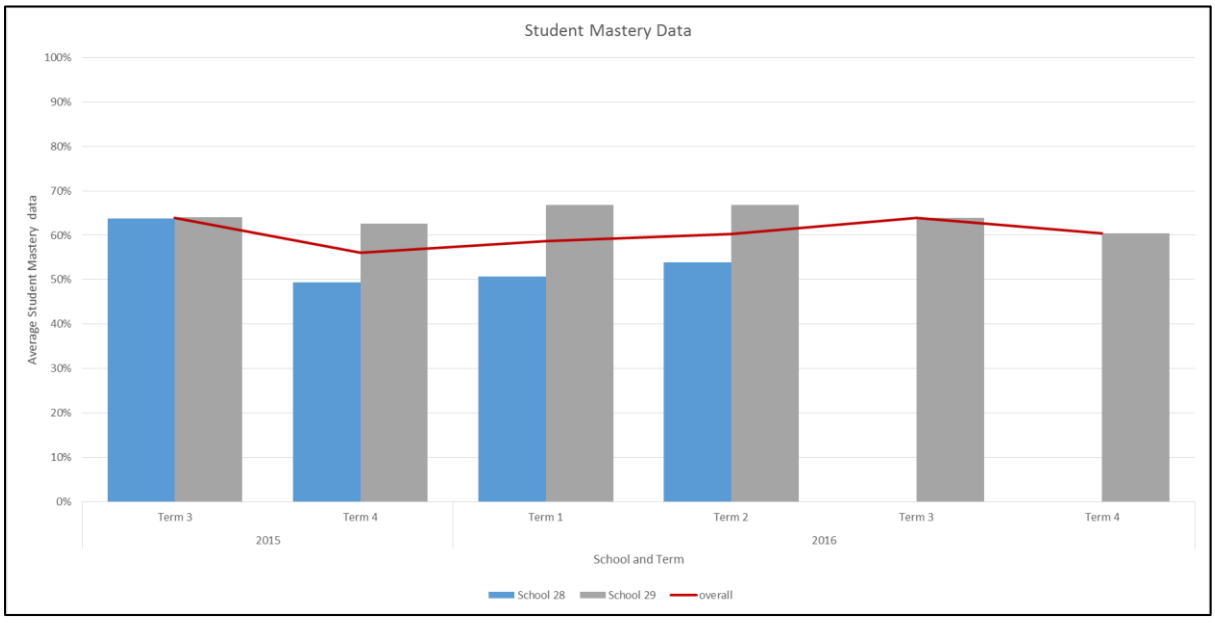


Figure 25. **Student Mastery Scores by Term for 2 WA Government Schools Delivering EDI.**

IMPACT: STANDARDISED ASSESSMENTS

The following section presents the results of standardised assessments that are considered indicators of impact on student outcomes.

JURISDICTIONAL LITERACY DATA

WA Government schools assess student literacy using PM Benchmark Reading Levels that assess instructional and independent reading levels through meaningful texts.

PM Benchmark data was provided to the evaluation team by WA Department of Education and Training. A small group of students from two schools had sufficient data (at least 2 time points) to determine effect size change in reading levels. The results of this analysis are presented in *Figure 26* and *Figure 27*. Please note that the black line indicates the effect size across all students in the relevant school. The degree of change varied within each school and grade with some students demonstrating considerable change and others minimal. There is insufficient data to make definitive statements regarding literacy progress in WA Government program schools. Nonetheless, the evaluation team will continue to work with WA DET to gather more complete data sets to further support the evaluation of the program.

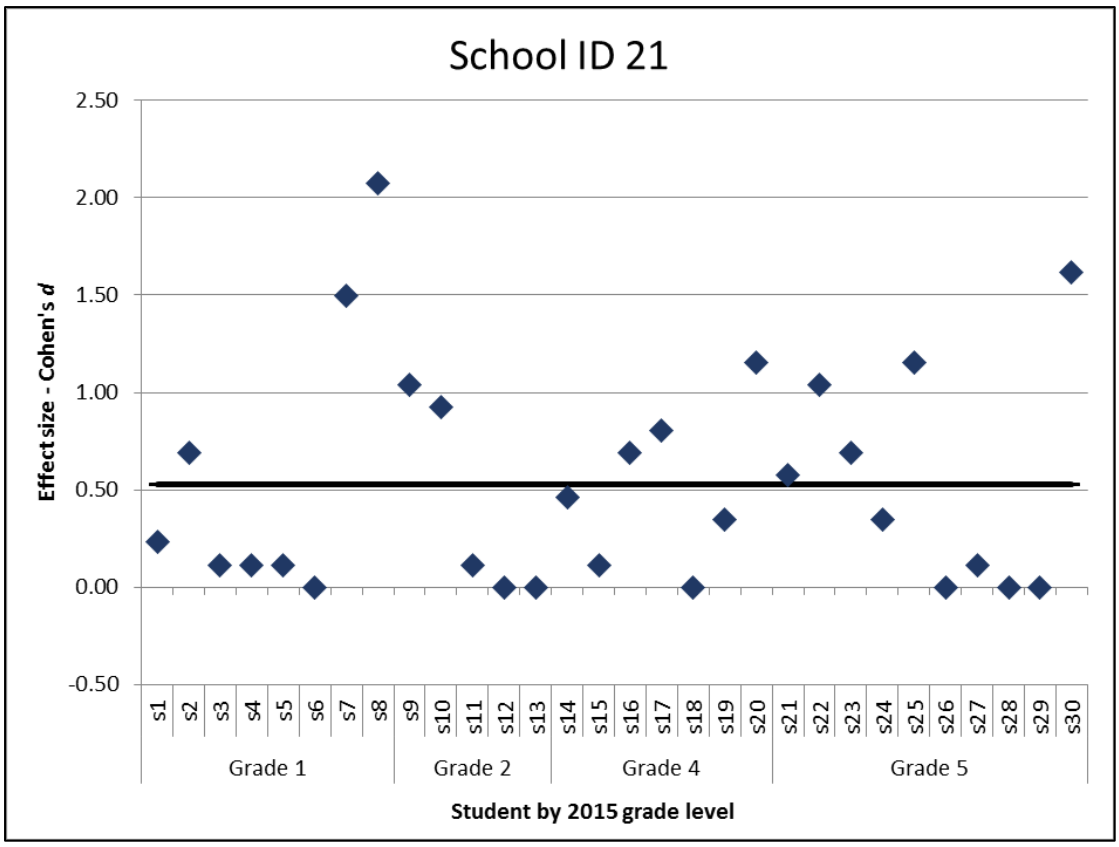


Figure 26. **Effect Sizes for School ID 21 on PM Benchmark Scores Between 2015 and 2016.**

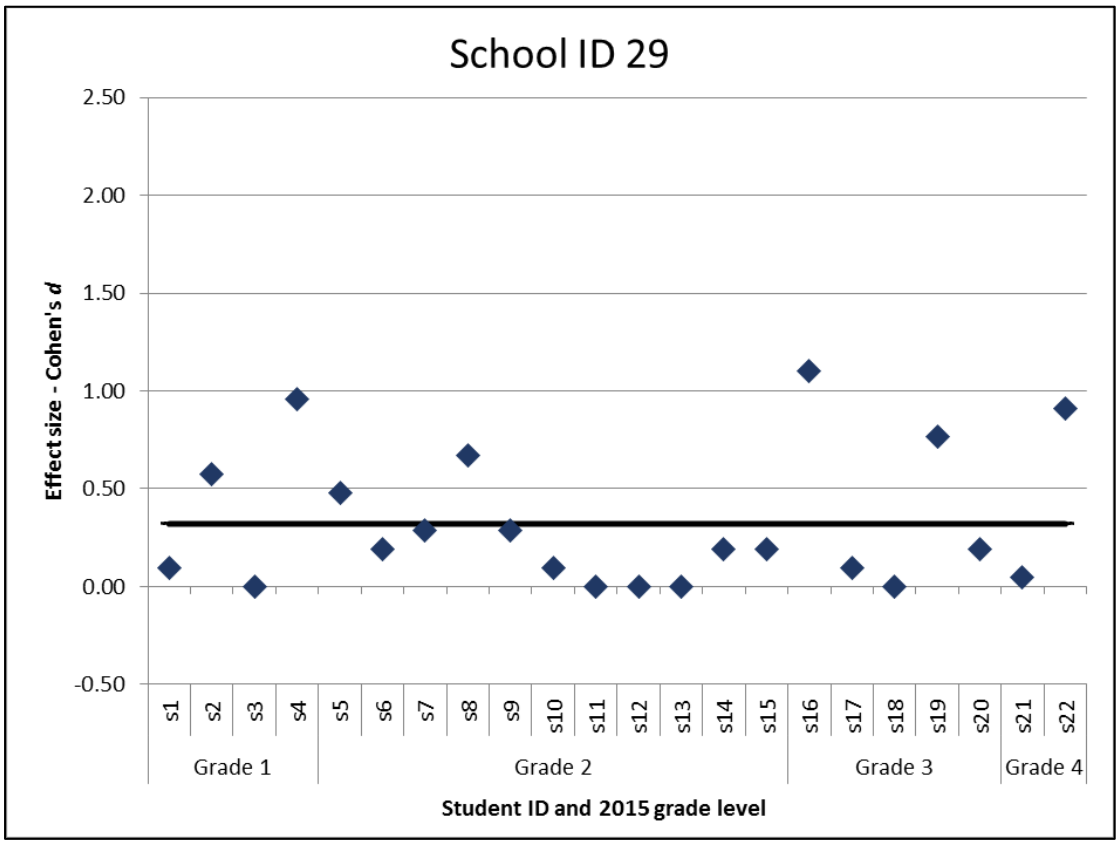


Figure 27. **Effect Sizes for School ID 29 on PM Benchmark Scores Between 2015 and 2016.**

NATIONAL ASSESSMENT PROGRAM FOR LITERACY AND NUMERACY (NAPLAN)

NAPLAN assessment scores formed part of four separate analyses:

- Cohort tracking compared with control schools between 2014 (pre-FLFRPSP implementation) and 2016 (post-FLFRPSP implementation)
- Change in each year level NAPLAN average scores over time (e.g. Year 3 NAPLAN for years 2014, 2015 and 2016)
- Percentage of students below National Minimum Standard (NMS) over years 2014, 2015 and 2016 for Grades 3 and 5
- Average change in NAPLAN scores between 2014 (pre-FLFRPSP implementation) and 2016 (post-FLFRPSP implementation)
- NAPLAN participation rates by year and Grade level.

Each of these analyses is now presented for WA Government program schools.

NAPLAN: PROGRAM SCHOOLS AND CONTROL SCHOOLS

To explore change in NAPLAN over time, both within and between the program and control schools, a series of independent *t*-tests were conducted comparing NAPLAN data from year 3 in 2014, and year 5 in 2016, for the NAPLAN domains of Reading, Writing, Spelling, and Grammar and Punctuation. For this analysis, change variables were created which reflect the change in NAPLAN domains scores between 2014 and 2016 (Change Variable = 2016 – 2014).

After this analysis was conducted, to better display results, mean scale scores for each domain were displayed for 2014 and 2016 compared to the National Average and Very Remote National Average; these graphs can be found in *Figure 28*.

Preliminary assumption testing did detect some assumption violations, with the presence of minor outlier, and skewness/kurtosis between greater than ±1; however, there was homogeneity of variances for the NAPLAN scores for control and program schools, as assessed by Levene's test for equality of variances. Due to these assumption breaches sensitivity testing is also conducted using Mann Whitney U tests. Further, as multiple analysis were conducted a bonferroni correction was applied to adjust for Type 1 error (detecting a difference when none is present), significance values have been set to less than 0.01.

Overall, no differences were detected in overall growth on NAPLAN between control and program schools within the WA Government School cohort between 2014 and 2016 on NAPLAN reading $t(43) = 0.39$, $p = .70$.; NAPLAN writing $t(42) = 0.18$, $p = .37$; NAPLAN spelling $t(45) = -1.36$, $p = .18$; or NAPLAN Grammar and Punctuation $t(45) = -1.37$, $p = .18$. These results suggest that both the control and program schools progressed at a comparable rate for students moving from year 3 in 2014 to year 5 in 2016.

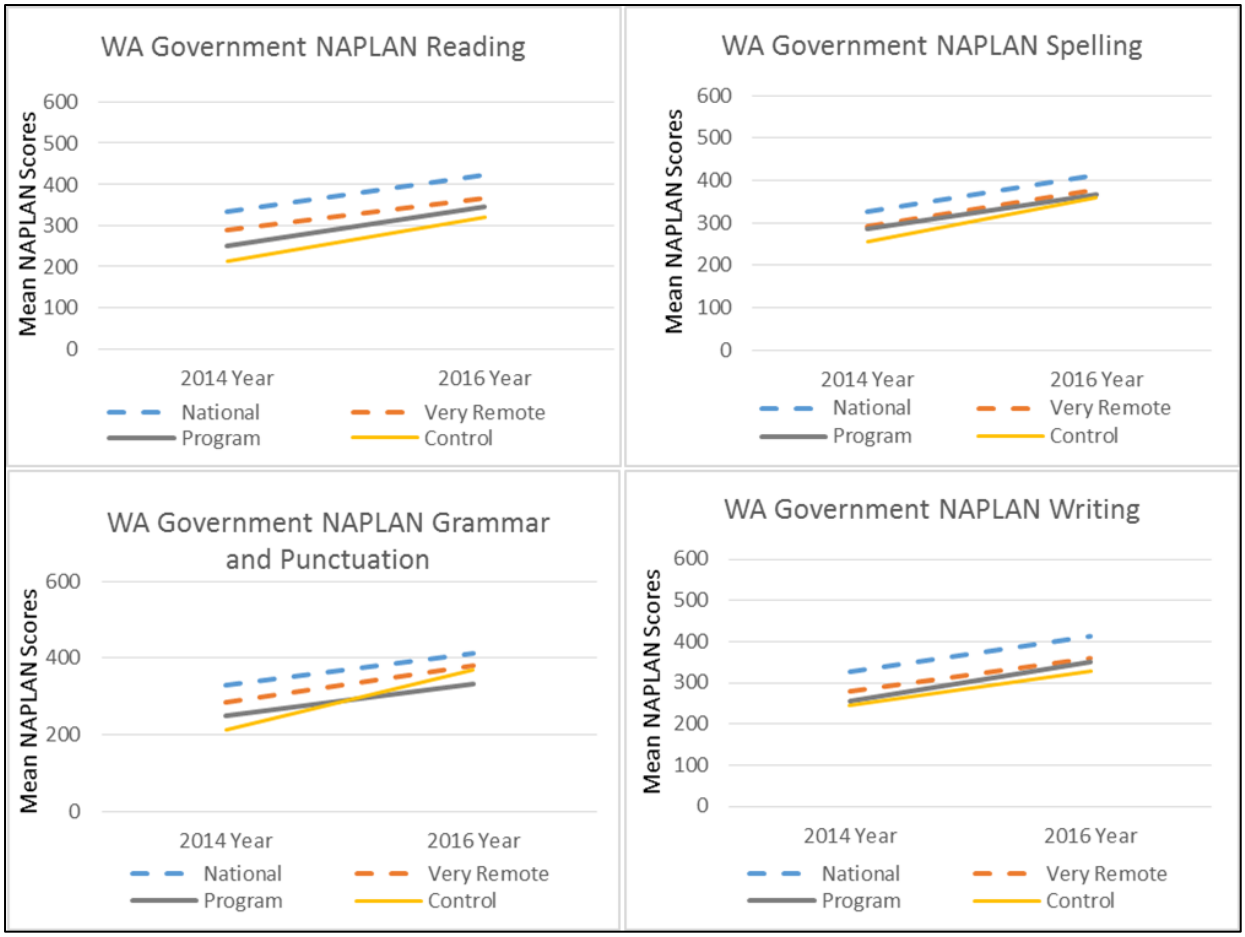


Figure 28. **Change in Mean NAPLAN Scores for WA Government Program Schools Against Control, National Average and Very Remote Average.**

NAPLAN: VARIABILITY

To explore changes in student NAPLAN scores over time mean scores for each NAPLAN domain were created for Years 3 and 5 for each school for 2014, 2015, and 2016. A One-way ANOVA was conducted to compare scores over this period. Preliminary assumption testing detected outliers, as assessed by inspection of a boxplot, and breaches the assumption of normality, as assessed by Shapiro-Wilk's test ($p > .05$), as well as values greater then ±2. As such, sensitivity testing was also conducted using Kruskal-Wallis H tests.

The results indicate no significant difference[9] over 2014, 2015 and 2016, for any NAPLAN domains in the WA Government Schools cohort; reading $F(2, 27) = 0.11$, $p = .90$; writing $F(2, 27) = 0.05$, $p = .95$; spelling $F(2, 27) = 1.04$, $p = .37$; or grammar and punctuation $F(2, 27) = 0.60$, $p = .56$. Despite there being no significant differences, for NAPLAN Reading, Spelling, and Grammar and Punctuation, in year 3, overall variability in these domains reduced over time, which may reflect a decrease in the gap between higher and lower scoring students lower scoring students.

[9] Confirmed by Kruskal-Wallis H tests.

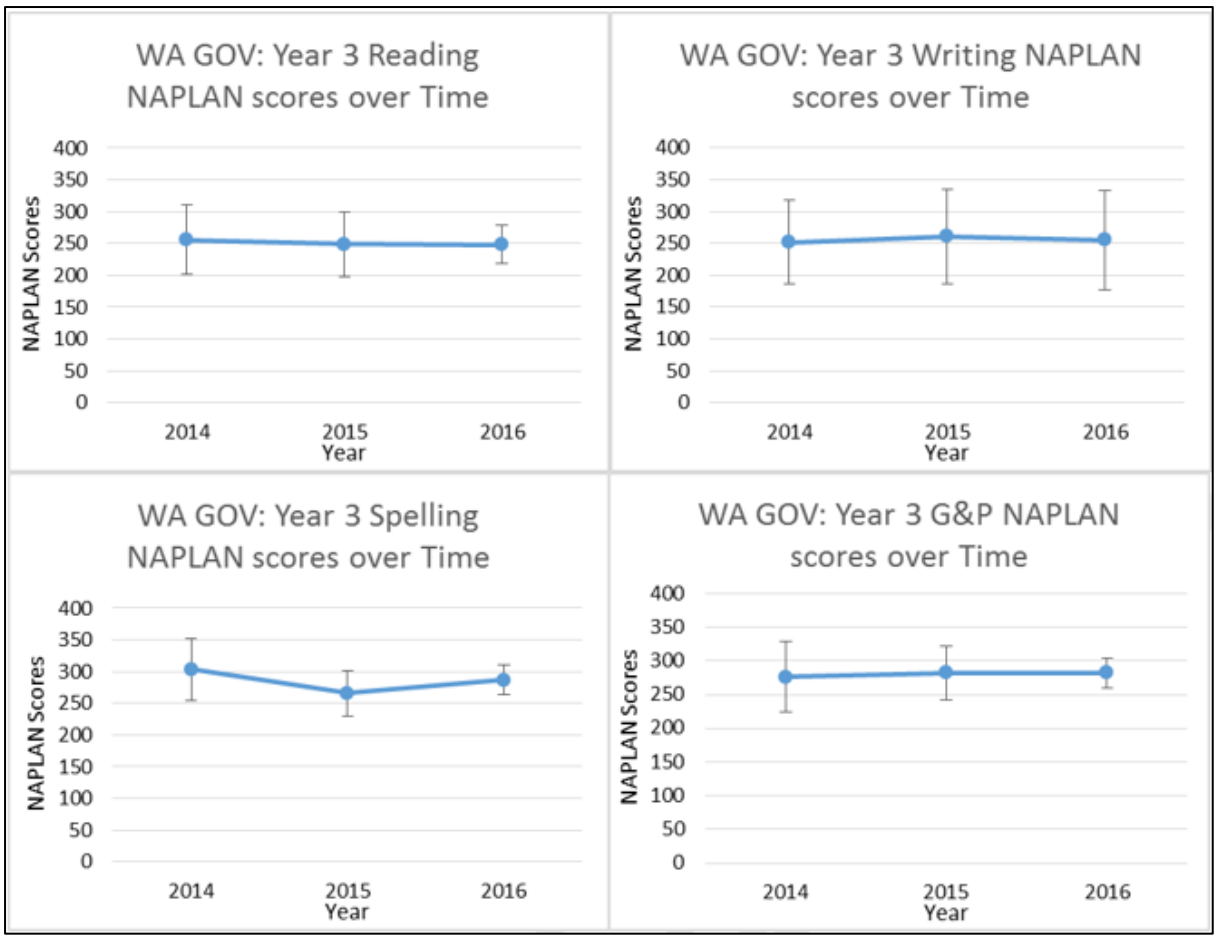


Figure 29. **WA Government Mean NAPLAN scores 2014 to 2016 for Year 3.**

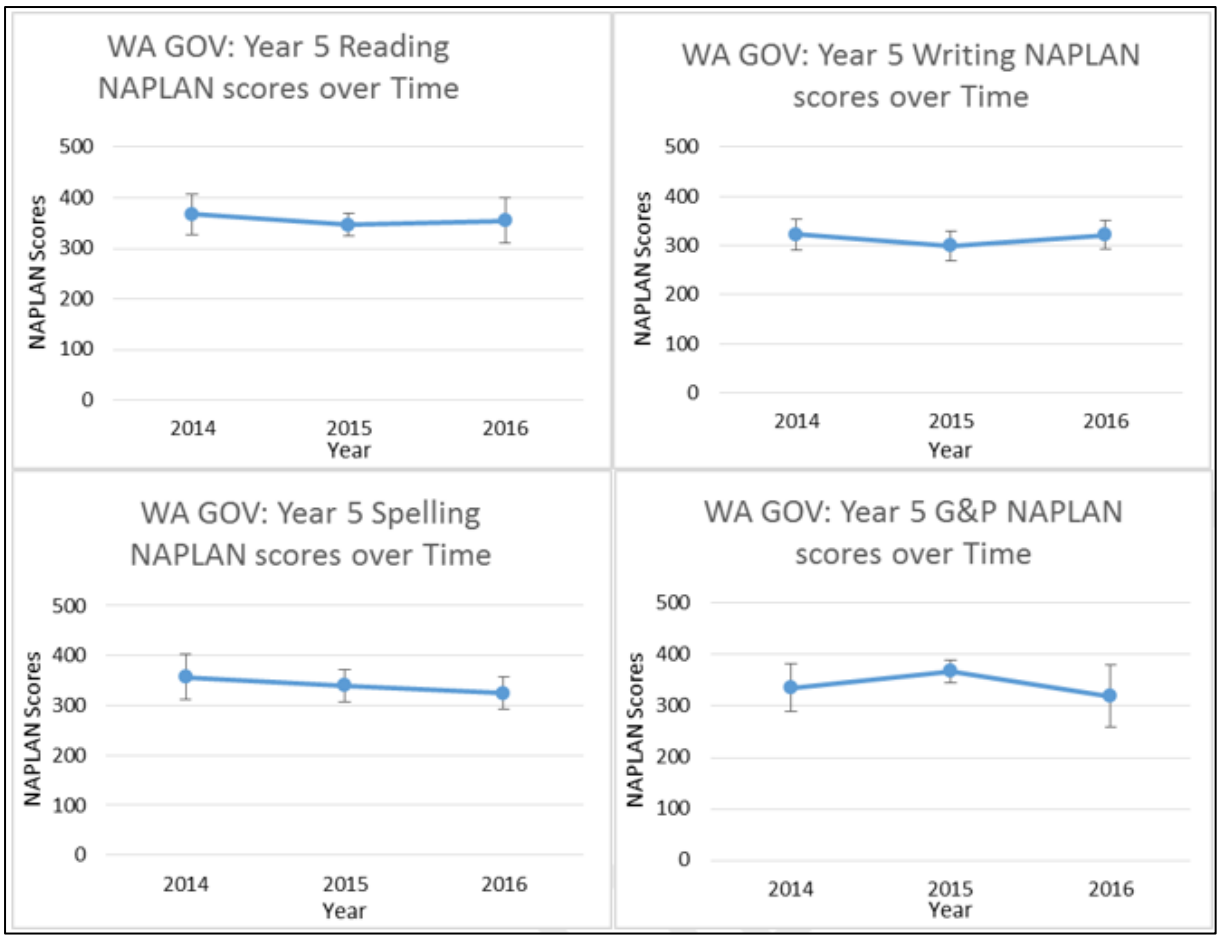


Figure 30. **WA Government Mean NAPLAN Scores 2014 to 2016 for Year 5.**

NAPLAN: PERCENTAGE OF STUDENTS BELOW NATIONAL MINIMUM STANDARD

Analysis of the change in the percentage of students below the national minimum standard (NMS) for years 2014, 2014, and 2016 and for Grades 3 and 5 are presented in *Figure 31* and *Figure 32*. Overall, there was a decrease in the percentage of students below the National Minimum Standard for students in year 3 across all fours domains with values lower in 2016 compared to 2014. However, the results were mixed for students in year 5. Writing and Grammar and Punctuation have shown a decrease from 2014 to 2016, while the number of students falling below the National Minimum Standard has increased for Reading and Spelling; although reading did decrease in 2015, it did not remain at this level. Overall, these results do not suggest that there have been any material improvements in the percentage of students below NMS. It should be noted, however, that this analysis is based on relational cohort-reliant data and may reflect overall changes to NAPLAN scores at a national level. Further, at this stage of the program, it would not necessarily be expected that these figures would reach statistical significance.

Figure 31. **Percentage of Students (Year 3) Below NMS for WA Government schools.**

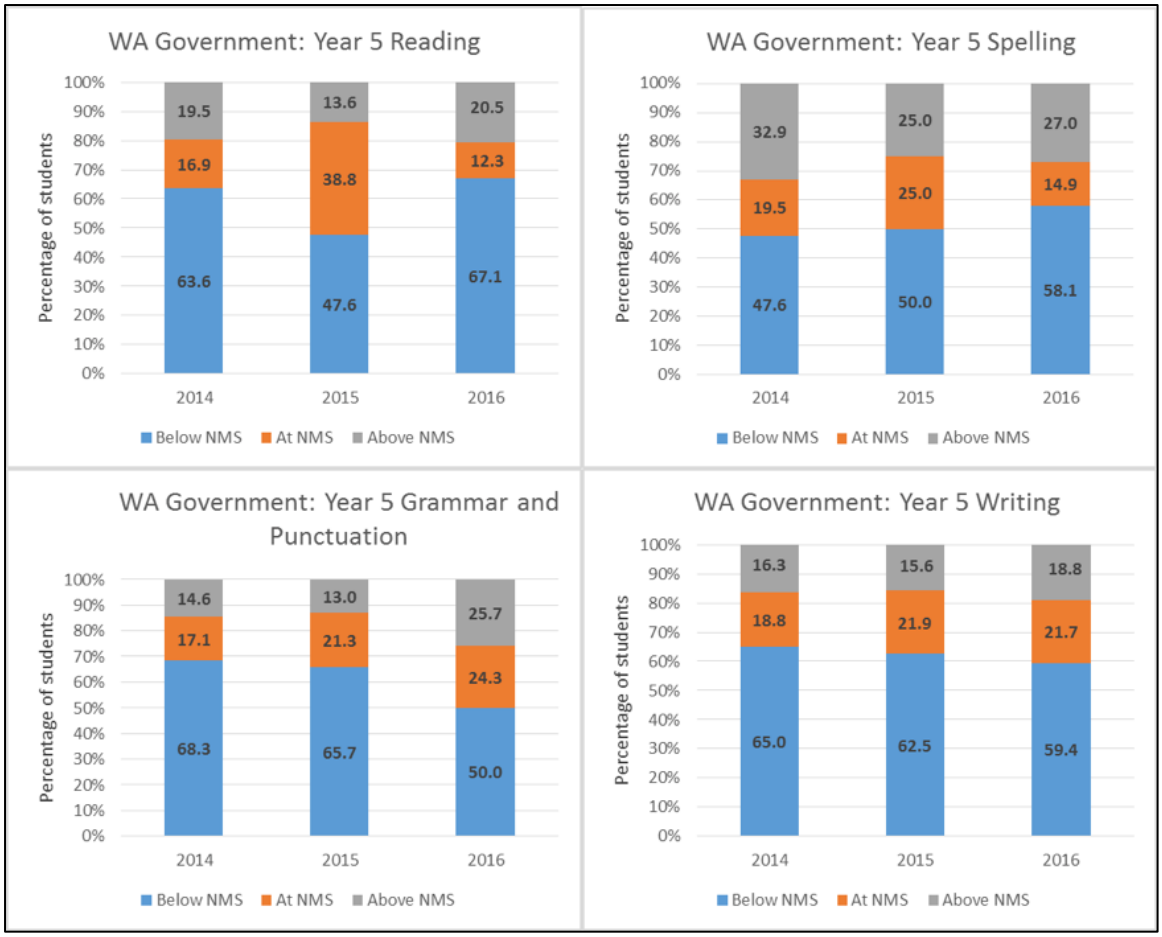


Figure 32. **Percentage of Students (Year 5) Below NMS for WA Government schools.**

NAPLAN: INDIVIDUAL SCHOOL ANALYSIS

At the individual school level, NAPLAN analysis focused on the mean (average) change in scores between Year 3 (2014) and Year 5 (2016) for the same cohort (unmatched), as well as confidence intervals, as this analysis facilitates a pre-post implementation comparison. The results were compared with State average gain for each assessment for WA. This was undertaken for all NAPLAN assessments and is presented in *Figure 33*. A majority of government program schools in WA demonstrated positive mean change in NAPLAN scores with some exhibiting scores above the WA state average, on Spelling and Writing, depending on the relevant NAPLAN assessment. These results reflect expected outcomes for NAPLAN after 2 years of program implementation and for changes between Year 3 and Year 5.

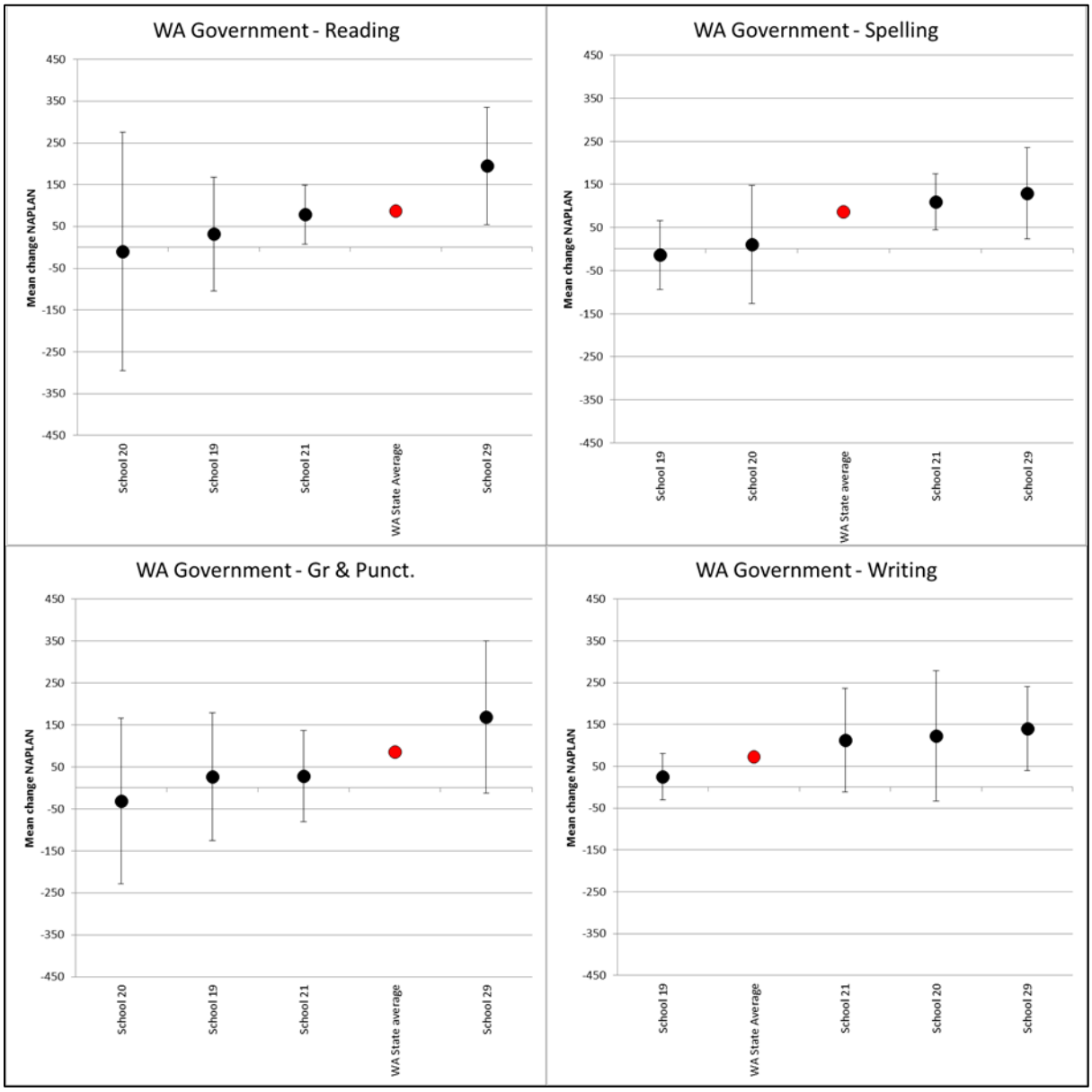


Figure 33. **WA Schools NAPLAN Change Year 3 (2014) to Year 5 (2016) by School.**

An average effects size2 for all WA Government Schools (n = 5) was also calculated based on the difference between on the mean (average) change in scores between Year 3 (2014) and Year 5 (2016). Schools demonstrated a small to large effect size for change over two years; NAPLAN Reading Hedge's g = 0.62;

NAPLAN Writing Hedge's g = 0.64; NAPLAN Spelling Hedge's g = 0.65; NAPLAN Grammar and Punctuation Hedge's g = 0.34.

NAPLAN: PARTICIPATION RATES

NAPLAN participation rates for WA Government schools were well below the State average for all years (2014 pre-program, 2015 and 2016). Over the three-year timeframe, participation rates for all assessments declined. Reasons for both the decline over time and for rates well below State average are not immediately apparent and require further investigation.

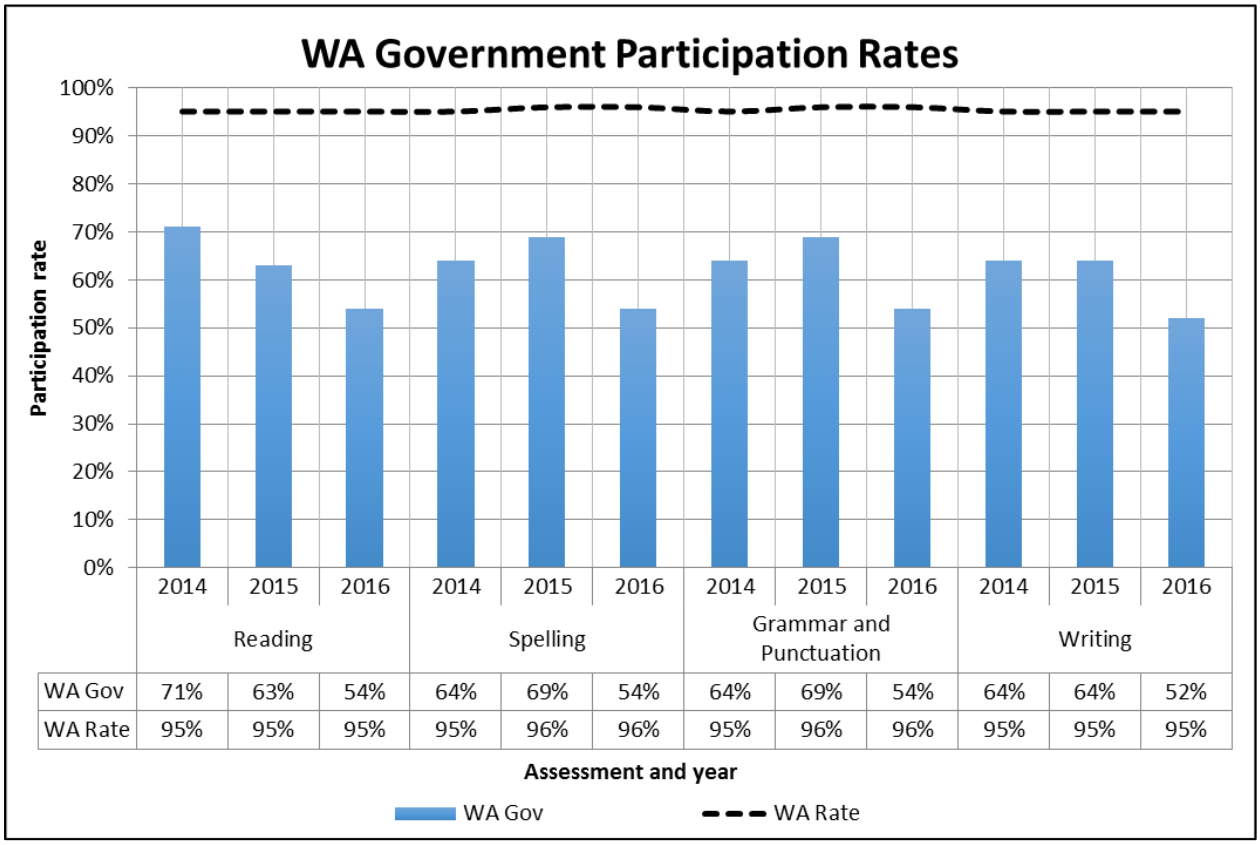


Figure 34. **NAPLAN Participation Rates by Assessment Type and Year for WA Government Schools.**

IMPACT: OTHER DATA SOURCES

TEACHER SURVEY

The teacher survey in WA Government schools had a response rate of n=4. Teachers who responded had at least 10 years of teaching experience. At the time of the survey, three teachers taught DI and one teaching EDI. The evaluation team intends to continue interviewing teachers and attending school visits and observations to gather a larger number of perspectives and experiences of school staff in the program.

Teachers were asked to provide their perceptions of the impact of FLFRPSP on students in their school. As mentioned, these questions considered factors such as literacy skills, engagement, wellbeing, and whether students like the program. *Figure 35* shows the results of these questions for WA Government schools. As demonstrated in *Figure 35*, the majority of teachers felt that the program had made an overall difference to students and improved their literacy skills. Fifty percent of teacher respondents also thought that students had engaged with the program. Responses were more mixed and largely uncertain when considering whether

the program had improved wellbeing and attendance and whether the students liked the program. For example, whilst one teacher felt as though "*the program doesn't suit students who are not at the direct level of the program",* another stated "*I enjoy seeing the education of students moving forward and that they can make good progress quickly".*

However, when asked if they felt there were students *not* improving on the program, all teachers responded in the affirmative. Explanatory factors offered for this lack of improvement included attendance issues (n=2) and that the program was too hard and boring for some students (n=1).

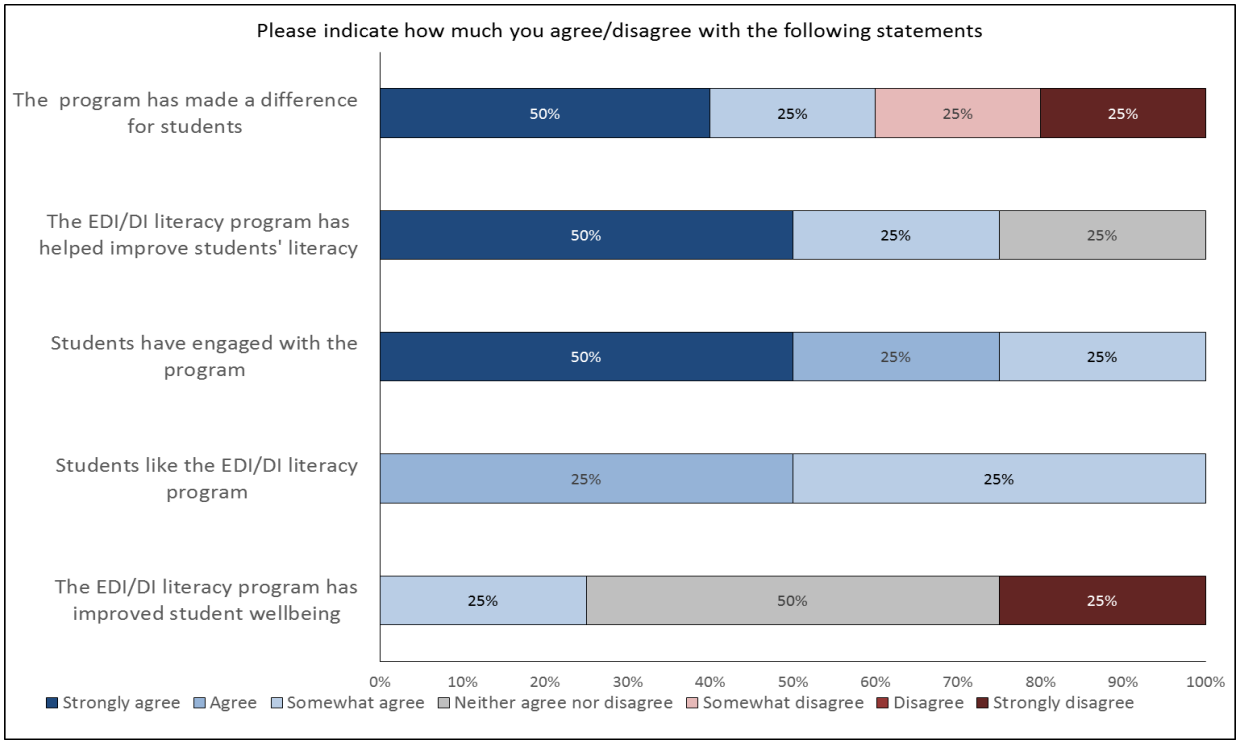


Figure 35. **Teacher Perception of the Impact of the Program on Students (n=4).**

PRINCIPAL INTERVIEWS

One interview with a Principal was conducted as part of the preparation of this report. The evaluation team intends to continue interviewing Principals and attending school visits and observations to gather a larger number of perspectives and experiences of school staff in the program.

The Principal interviewed was overwhelmingly positive about the program. In terms of impact on students, the Principal did acknowledge that the first term of implementation was very difficult for students, but that since then, they were getting exponentially better at English skills in an environment where English was considered their third language. The Principal also commented that the children had progressed from extrinsic motivation via reward to more intrinsic motivation, and that they were "enthusiastic about their own progress."

DOCUMENT ANALYSIS

Recurring themes throughout the WA school documents focused on results of DI/EDI and included literacy progress, student engagement, and attendance. Implementation information from these schools and outcome information from all WA Government schools is detailed below.

Literacy progress. Schools noted positive, though nonspecific, student progress in reading, spelling, and/or writing skills. In 2016, one remote school found that DI "had a particular impact on students who had low attendance but attended at regular intervals. These students were able to continue to improve their reading and language skills with only a small amount of targeted practice". Another school noted in 2015 that DI "has seen Reading, Writing, and Spelling levels rise comprehensively."

Student engagement. DI/EDI was seen as the cause of increased student engagement in several schools. In some cases, the predictable and consistent structure of DI was also especially beneficial. In 2016, one school noted that students "are now motivated purely by their own success". An EDI coach found that EDI allowed teachers to "deliver powerful and meaningful lessons in a way that maximises the learning opportunity for Indigenous students", resulting in increased student engagement. Some students also benefitted from the consistent structure of the DI/EDI programs: one school found that students taught with EDI were familiar with classroom routines when they moved from year to year, which was especially beneficial for Aboriginal students who travel in and out of the town frequently.

Attendance. Overall, the implementation of DI/EDI has seen an increase in attendance. Low attendance, however, is also a negative factor affecting the effectiveness of DI. A WA Inquiry into educational opportunities for Aboriginal and Torres Strait Islander students noted that, by March 2015, the schools that had implemented DI had already seen a dramatic improvement in classroom performance and attendance (MacTiernan, 2016). In 2016, one school found that DI "had a particular impact on students who had low attendance but attended at regular intervals. These students were able to continue to improve their reading and language skills with only a small amount of targeted practice". In Semester 2, 2016, students with more than 70% school attendance achieved 76 out of the 80 recommended lessons, and students with 40-70% school attendance finished 50-65 lessons. "Low attendance continues to hamper the reading levels of students, but DI allows students to have more targeted lessons and to catch up quickly." This school has had two-thirds of students with 80% attendance or above since DI was implemented(Taylor, 2015).

TEACHERS

Teacher experience data was available for two WA government schools, and indicated teachers to be relatively experienced, with an average of 10.56 years' experience. Teacher turnover data was only provided for 3 schools in the WA Government group. The results are shown in Table 11. Across the three schools there were variable turnover rates. One school experienced very low teacher turnover rates in 2016, but greater rates in 2015 where turnover was considerably high.

Table 11: ***Teacher and Teaching Assistant Turnover 2015 and 2016 for WA Government Schools (n=4)***

	2015		2016	
School ID	Teacher Turnover	Teaching Assistant Turnover	Teacher Turnover	Teaching Assistant Turnover
School 19	-	-	185%	100%
School 20	129%	-	0%	44%
School 21	83%	86%	114%	0%

IMPACT

For WA Government schools, teacher effectiveness as measured by the FLFRPSP contains three dimensions already described in the NT Government section. *Figure 36* shows the development of teacher effectiveness as measured by the programs over Terms 1 and 3 for 2015 and Terms 2 and 3 for 2016.

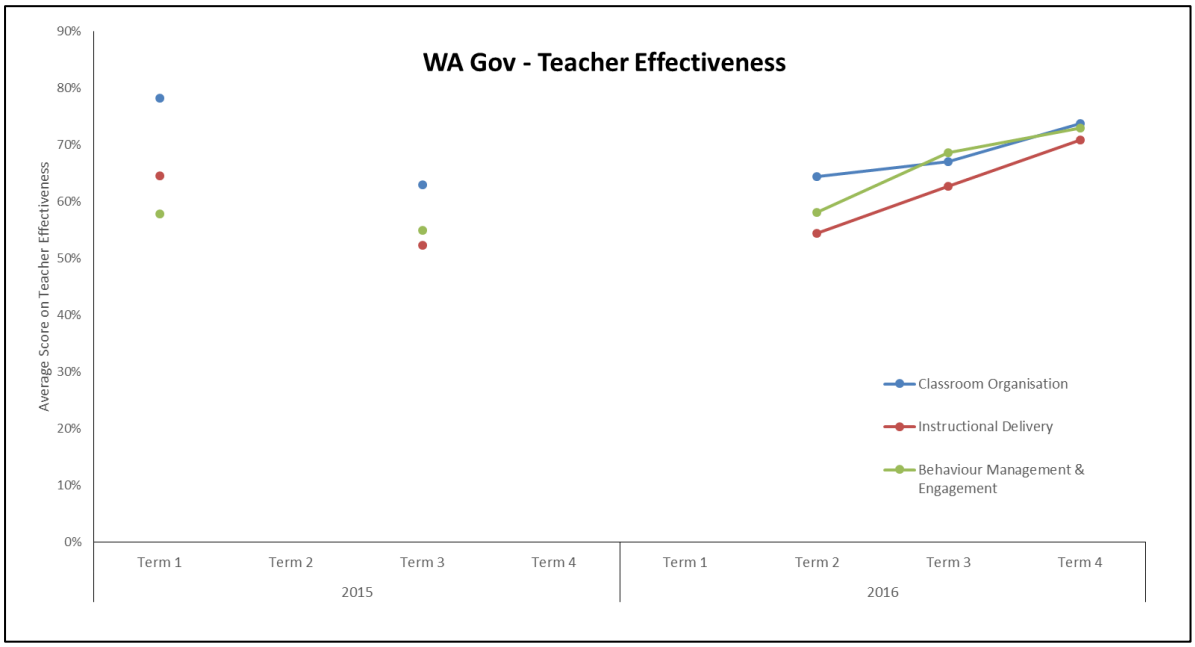


Figure 36. **Teacher Effectiveness Measures from DI/EDI Program Data.**

Teachers were also asked to reflect on how they believed the program had impacted their teaching practices, their knowledge of the program, and teaching literacy more generally. *Figure 37* and *Figure 38* show the results of these questions. While all teachers displayed high levels of confidence in teaching the program, one teacher reported a low level of confidence in relation to teaching literacy more generally. However, many teachers did not feel as though the program benefited their teaching practices, with a majority strongly disagreeing with the statement “the program has improved my ability to teach literacy”.

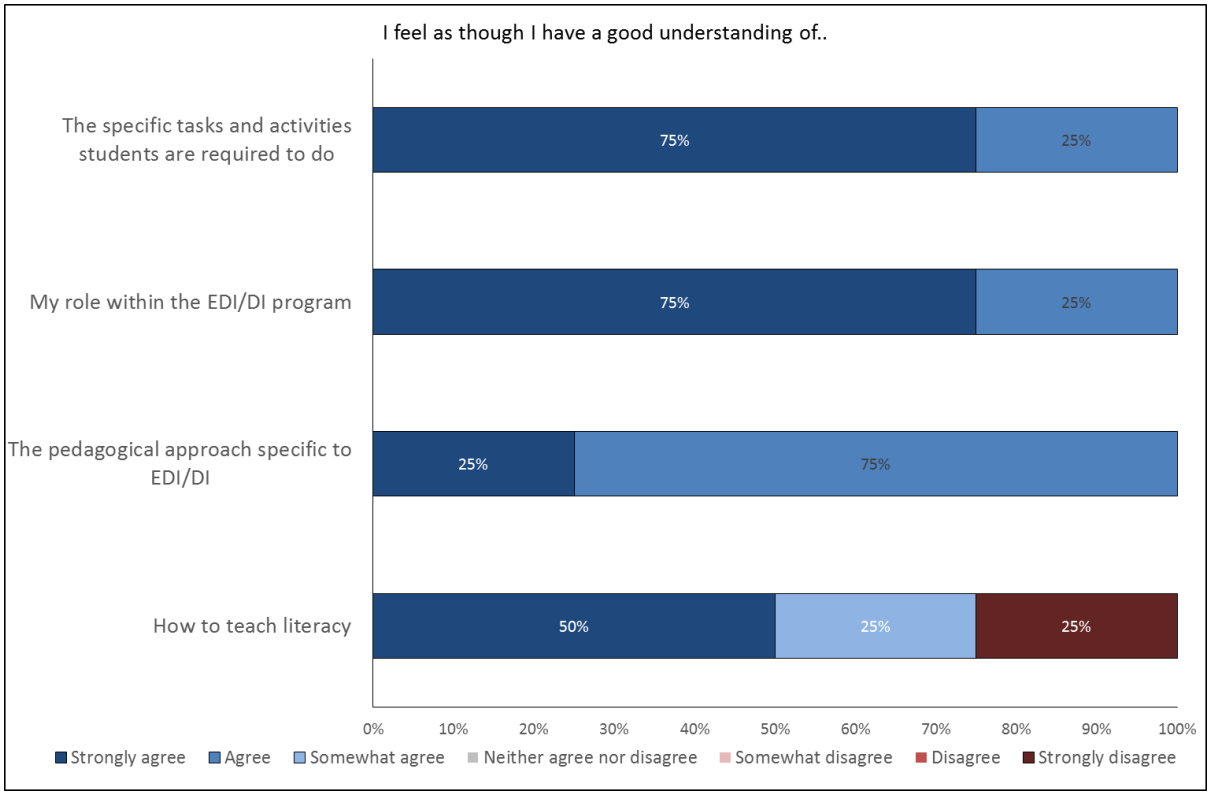


Figure 37. **Teacher Perception of the Program's Impact of their Role and Work as Teachers (n=4).**

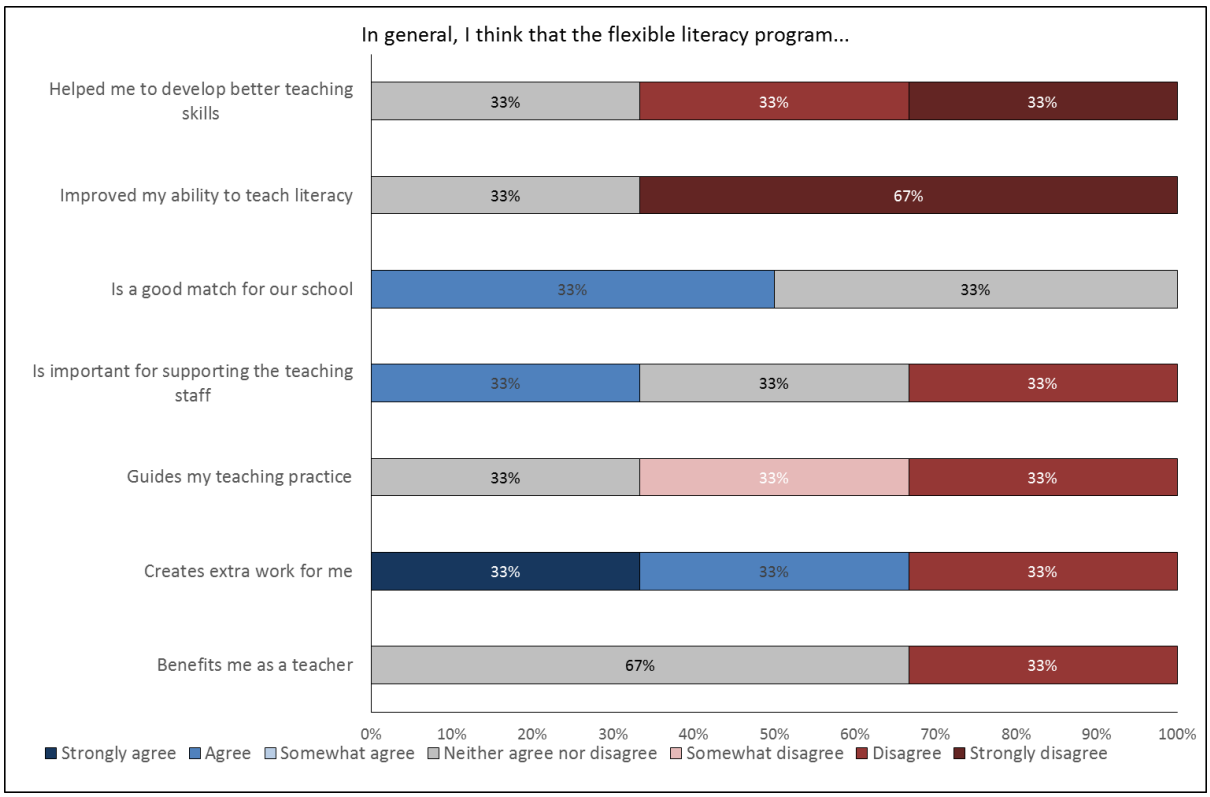
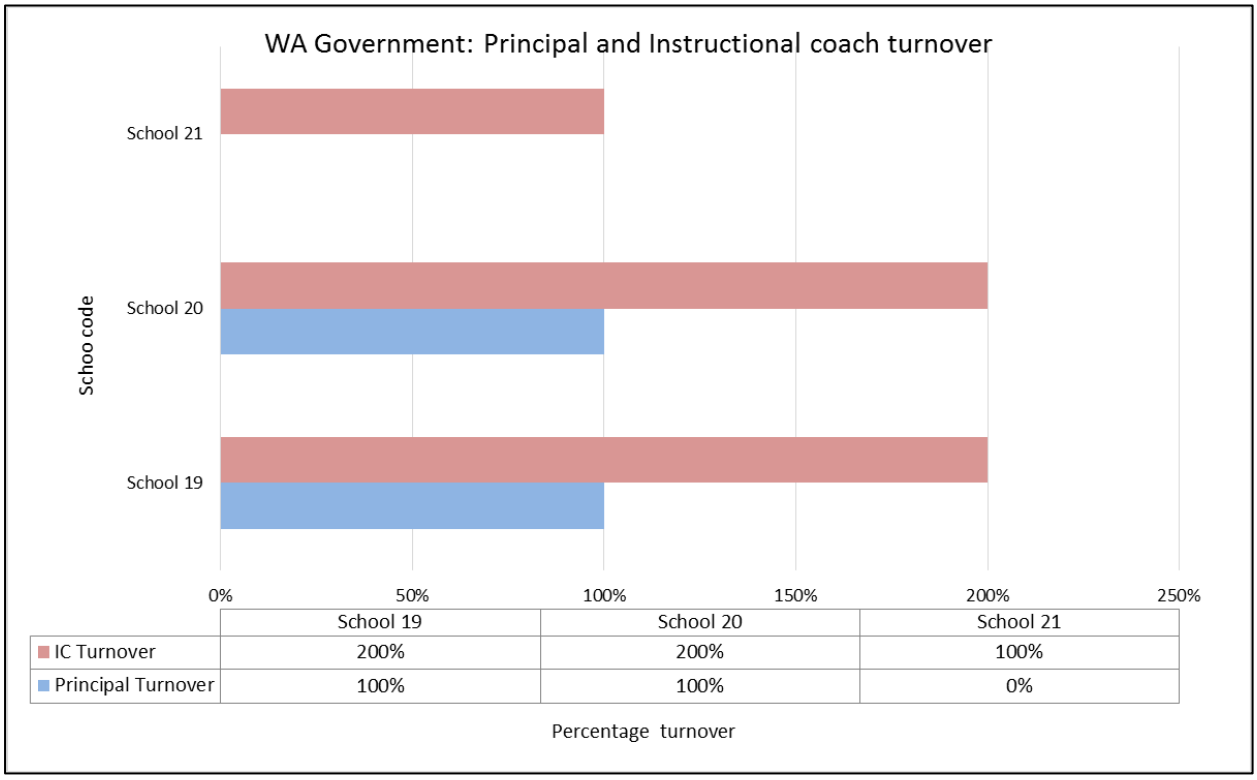


Figure 38. **Teacher Perception of the Programs Impact on their Pedagogy (n=3).**

SCHOOL LEVEL DATA

Relevant demographics in addition to several indicators of program impact and implementation have been considered at the school level. Demographic data includes publicly available information such as that available on the MySchool website, as well as provided information such as turnover in leadership staff. As for NT Government schools, program data includes measures of school fidelity and program support provided by GGSA through training and observations. Finally, perceptions of impact from teachers and principals are also included where relevant to the school as a whole.

Figure 39 shows turnover rates for principals and instructional coaches in DI school in WA between 2015 and 2017 only. Instructional coaches are school-based staff who provide on-site coaching support to teaching personnel. Note that in the data table '0' refers to no data whereas '0%' reflects no turnover. Turnover of these leadership roles with respect to the program is high across most schools with the majority replacing a leadership role at least once per year. This has implications for the program in terms of continuity, advocacy, and implementation.

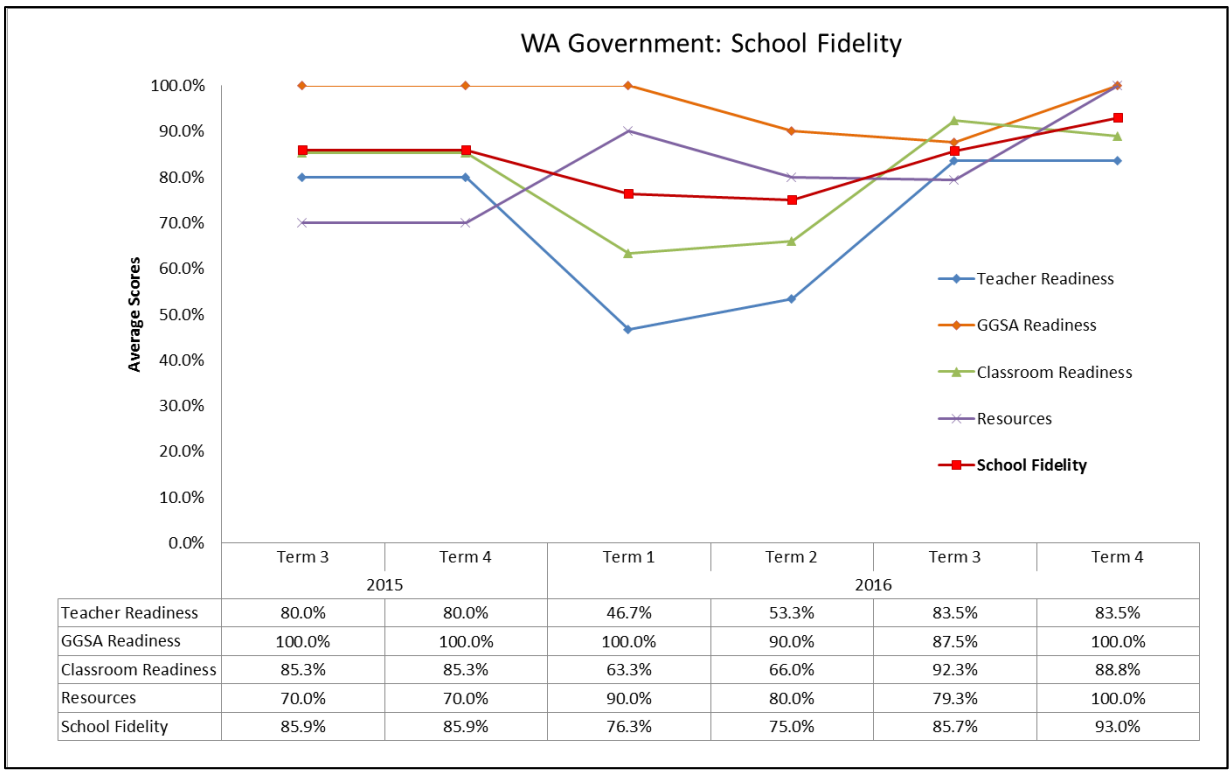


	School 19	School 20	School 21
IC Turnover	200%	200%	100%
Principal Turnover	100%	100%	0%

***Figure 39.* Principal and Instructional Coach Turnover 2015 to 2017 for WA Government Schools (DI only).**

IMPLEMENTATION

At the school level, understanding of implementation is measured primarily with program data. School fidelity comprises four factors: teacher readiness, GGSA readiness, classroom readiness, and resources that are then combined into an overall score. Table 8 describes each of these factors in more detail and *Figure 40* shows scores for all factors and overall across 2015 and 2016. Similar to NT Government schools, WA Government schools show some variability across all four factors over time. GGSA readiness was consistently high and indicates that support and monitoring from GGSA is well implemented. Teacher readiness scored lowest of the four factors, which again may reflect teacher turnover rates, which have a flow on effect to other school-based factors such as classroom readiness and resource readiness.



	Term 3	Term 4	Term 1	Term 2	Term 3	Term 4
	2015		2016			
Teacher Readiness	80.0%	80.0%	46.7%	53.3%	83.5%	83.5%
GGSA Readiness	100.0%	100.0%	100.0%	90.0%	87.5%	100.0%
Classroom Readiness	85.3%	85.3%	63.3%	66.0%	92.3%	88.8%
Resources	70.0%	70.0%	90.0%	80.0%	79.3%	100.0%
School Fidelity	85.9%	85.9%	76.3%	75.0%	85.7%	93.0%

Figure 40. **School Fidelity Measures for WA Program Schools 2015 and 2016.**

IMPACT

Several questions in the survey of teachers addressed the potential impact of the program at a school level. These questions focused on the drivers of implementation and aspects of the program that were easy or difficult to implement.

WA FLFRPSP teaching staff identified several positive program drivers including the support provided from GGSA and NIFDI as well as school staff more generally. FLFRPSP drivers that did not support the program included poor student attendance, poor support staff attendance and some of the large paper work requirements of the program. Respondents also mentioned that the program was not easy to implement and could be boring and monotonous.

The principal interview reiterated the support from NIFDI and GGSA which helped with implementation, particularly in the first 18 months. They also cited teacher buy-in as a key factor to success. Finally, the principal also alluded to the need to continue the program to realise its benefits. They commented,

> ***"What we need is to stick to it because for us it's working and we don't want it to come to an end after three years because we don't want to give it up. For us it is proving to be a winner. In 10 years' time we are going to see the biggest benefits of this program for this community and for these students."***

PROGRAM SUPPORT

Program support refers to implementation and monitoring assistance provided by GGSA, in-school coaches, and program developers support (NIFDI, DataWorks).

Figure 41 shows the average number of observations and average teacher training for NT program schools in 2015 (Terms 3, 4) and 2016. DI program training, and support training for teachers, teaching assistants, and school leadership variously. Observations are the average number of observations per school, and consist of two-minute observations, five-minute observations, and extended observations. These observations were conducted by an implementation manager, principal, instruction coach, teacher coach, or teaching principal. As to be expected, the WA Government data shows there to be a spike in the number of observations and training sessions earlier in the year. Overall, school staff participated in a high number of training and observations.

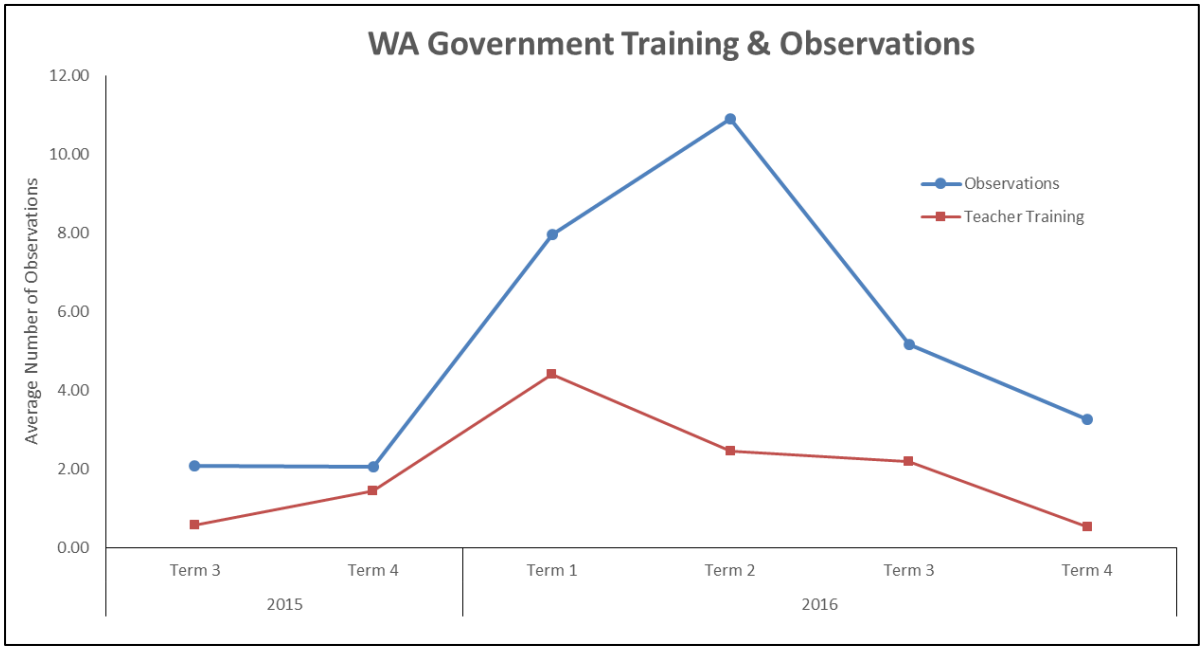


Figure 41. **Average Number of Training Sessions Attended and Coaching Observations by Term for WA Government Schools.**

Teaching staff were also asked in the survey their perceptions of their readiness to teach using either DI or EDI following training. *Figure 42* displays the results of this question for teaching staff in WA Government schools. The majority of staff agreed that they were ready to implement the program, had enough resources, and understood the potential benefits of the program following training. They also felt that they were supported to implement the program.

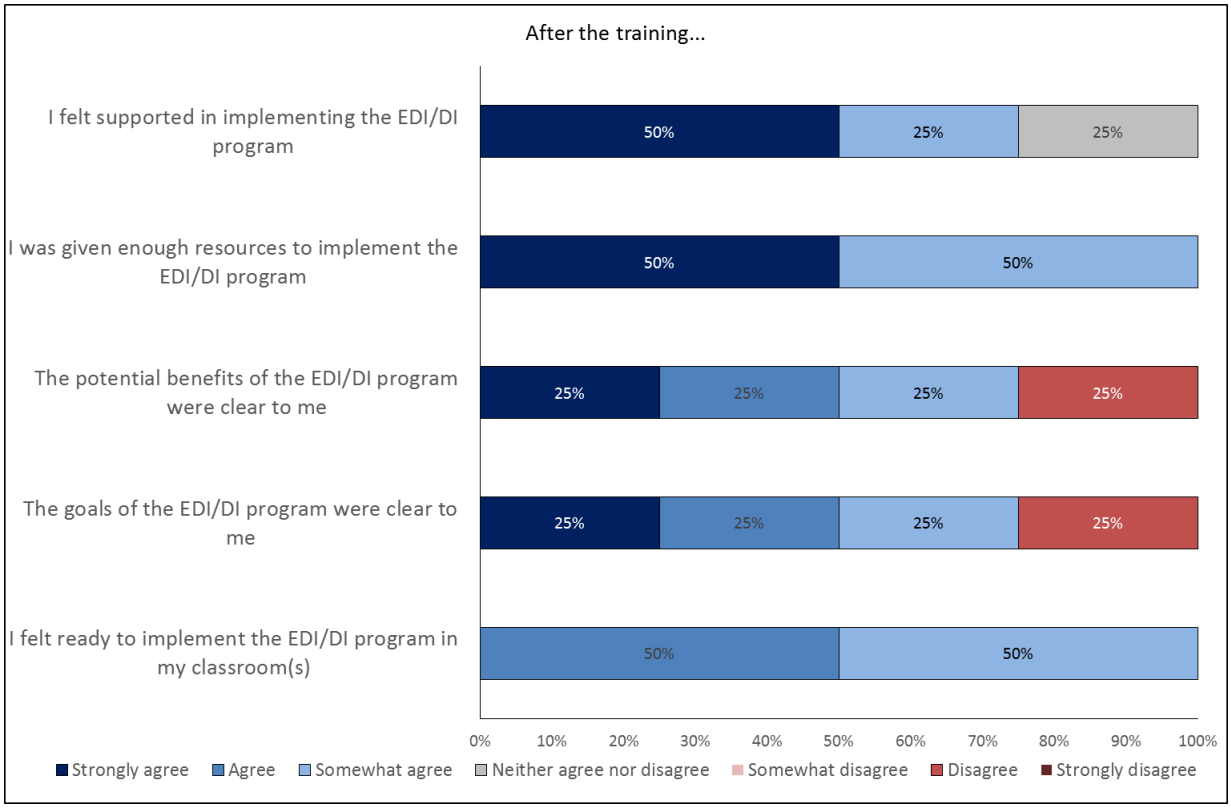


Figure 42. **Teacher Perception of Training (n=3).**

OVERALL SUMMARY OF COLLECTIVE RESULTS FOR WA GOVERNMENT SCHOOLS

In summary, there are variable results from WA Government program schools at the student, teacher, and school level.

For students in WA Government schools there are currently mixed results of early impact, suggesting that additional data and time is required to fully understand the impact of the program. More complete literacy data at the jurisdictional level will assist in better understanding the program effects. The NAPLAN analysis does not indicate any current impacts from the program, as would be expected. Nonetheless, there are pockets of promise: steady gains from year 3 to year 5, reduction in the percentage of students below the NMS in several domains, and some schools achieving gains greater than the State average. Participation rates are low across the board and not increasing.

Teacher survey data and Principal feedback on student impact suggests that schools perceive that the program is having a positive effect on student literacy. Key drivers that arose from most respondents related to the importance of regular student and teaching assistant attendance in contributing to the program's overall impact.

At the teacher and school level, turnover of teachers and teaching assistants is a significant problem and one that encroaches on implementation and fidelity. Nonetheless, teachers report being well prepared to teach the program and understand the pedagogical approach following initial training. However, they do not necessarily consider that working with the program has influenced their ability to teach more generally suggesting that their learnt skills in DI/EDI are a discrete set of instructional activities. Program data measuring overall school fidelity against four domains indicate that schools are receiving excellent levels of support from GGSA whereas teacher readiness, a measure of staff being trained and other activities requires further refinements and improvement.

Identified drivers[10]:

- Support from GGSA, NIFDI (+)
- Lack of regular student attendance (-)
- Lack of regular teacher assistant attendance (-)
- Teacher and leadership turnover (-)

HEAT MAP OF WA GOVERNMENT PROGRESS IN FLFRPSP

Figure 43 illustrates for each data source the extent to which schools in the jurisdiction are operating below, at, or above expected levels for this stage of the evaluation. This is based on review of all the available evidence and an evaluative judgment by the evaluation team. It provides a visual summary of the status of the jurisdiction and will inform the next phase of the evaluation. It also provides a basis on which progress can be monitored and re-assessed. Overall, it indicates that WA Government FLFRPSP schools are meeting expectations for this phase of the evaluation and has varied results.

	Implementation	Impact		
	Program data	Proximal primary data	Distal primary data	Other primary data
Students	Lesson progress and mastery assessments	PM Benchmark (more data required)	NAPLAN	Teacher and Principal report (more data required)
Teachers	Training and observation (GGSA provided)	Teacher effectiveness program measures	N/A	Teacher and Principal report
Schools	School fidelity	N/A	N/A	Teacher and Principal report

Figure 43. **Heat Map for WA Program Schools.**

Rating criteria

Above expectations for this phase of the evaluation	Meets expectations for this phase of the evaluation (mixed results – some aspects require attention while others are good.	Below expectations for this phase of the evaluation (requires substantial improvement)	Not enough information available	Not applicable N/A

[10] Enablers are identified with a positive sign, while barriers are identified with a negative sign.

WESTERN AUSTRALIA CATHOLIC EDUCATION OFFICE

There are 163 Catholic schools in WA across four Catholic Dioceses. Seven WA Catholic schools are part of the FRLPS program with 5 of these schools delivering EDI in the primary years. The remaining 2 schools, Christ the King and Warlawurru commenced with EDI and transferred to DI in Term 4 2015 and Term 3 2016 respectively.

WA CEO schools undertake literacy assessments using EYLND which is administered by University of Western Australia (UWA) in conjunction with the WA Catholic Education Office. The assessments are designed for Years 1-3 and cover a range of literacy skills including Letter Identification, Concepts about Print, the Duncan Word Test, Writing Vocabulary, Dictation, Burt Word Reading Test and Running Records of progress.

DEMOGRAPHIC PROFILE

Table 12 summarises relevant demographic information regarding FLFRPSP schools in WA Catholic Education and compares them with the same data for all schools in the program. As can be seen, WA Catholic Education schools have a higher mean ICSEA value than the mean for all program schools, a lower percentage of Indigenous students, and lower % of students with a LBOTE. Attendance rates for schools in this cohort were higher compared with attendance for the all program schools.

All schools are geographically classified as Very Remote apart from one which is Remote.

Table 12: ***Mean Values for Key Demographic Variables for WA Catholic Education Schools with All Program Schools Data as a Comparison***

	Teachers	Staff	ICSEA	Total students	Girls	Boys	Indig %	LBOTE %
WA Cath	9.86	8.29	752.57	101.14	51.29	52.86	78.00	75.00
PROG	9.94	5.81	666.32	100.71	49.19	52.06	92.00	83.00

Table 13: ***Attendance Rates for WA Catholic Education Program Schools***

	2014			2015			2016		
	Days Present	Days Absent	Attendance Rate	Days Present	Days Absent	Attendance Rate	Days Present	Days Absent	Attendance rate
WA Cath	120	32	79%	116	34	78%	117	27	81%
Overall	111	44	72%	110	44	73%	108	44	71%

STUDENT LEVEL DATA

IMPLEMENTATION

EXPLICIT DIRECT INSTRUCTION SCHOOLS

EDI schools measure student mastery in Weeks 5 and 10 of each term. Percentage scores have been used in this analysis because total scores can differ depending on year level. *Figure 44* graphically demonstrates average student mastery scores for students in EDI Catholic Education schools over Terms 3 and 4 2015 and

for all four terms in 2016. Despite an initial dip in average mastery score in Term 4 2015, scores have remained consistent in 2016.

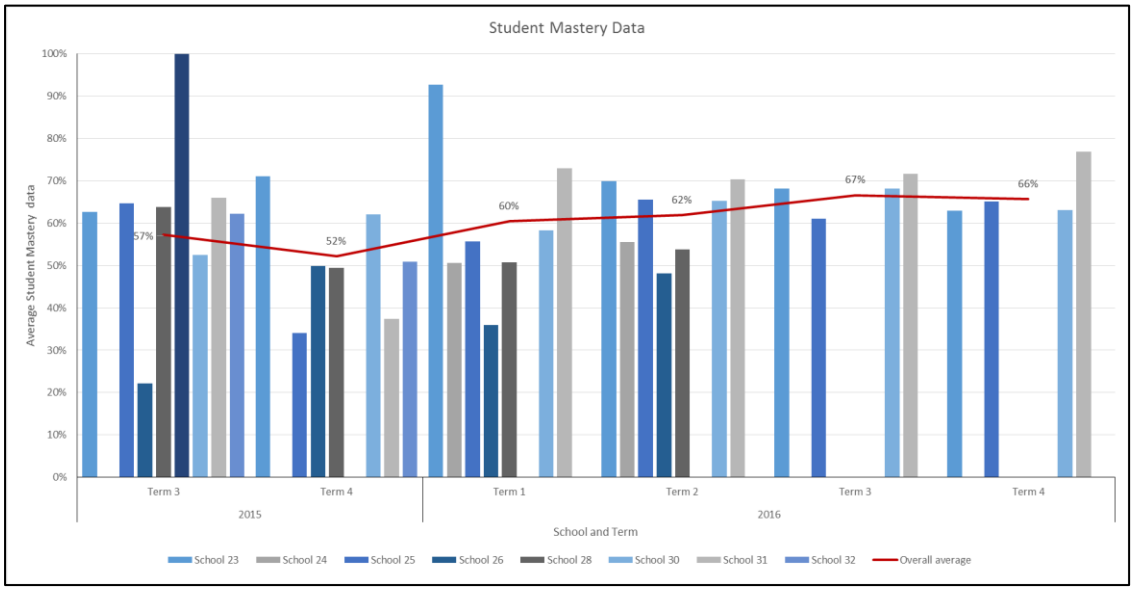


Figure 44. **Student Mastery Scores by School and by Term for WA Catholic Education EDI Schools.**

** The overall score for all schools is shown as the red line.

DIRECT INSTRUCTION SCHOOLS

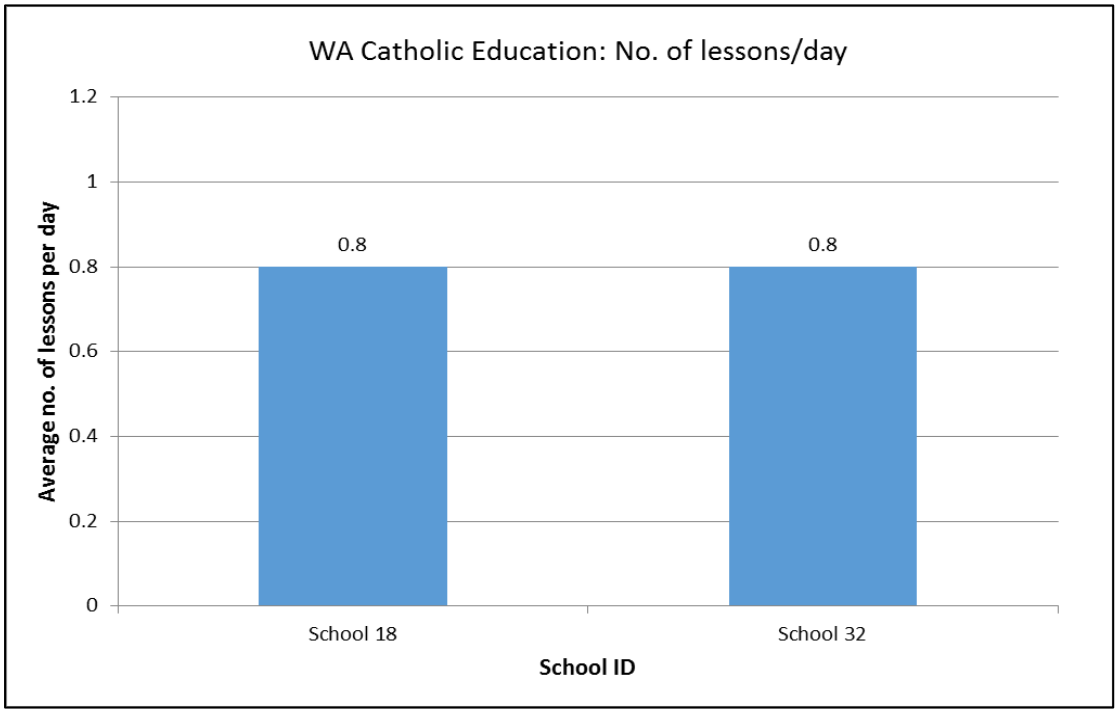


Figure 45. **Average Number of Lessons Delivered Per Day for WA Catholic Education DI Schools.**

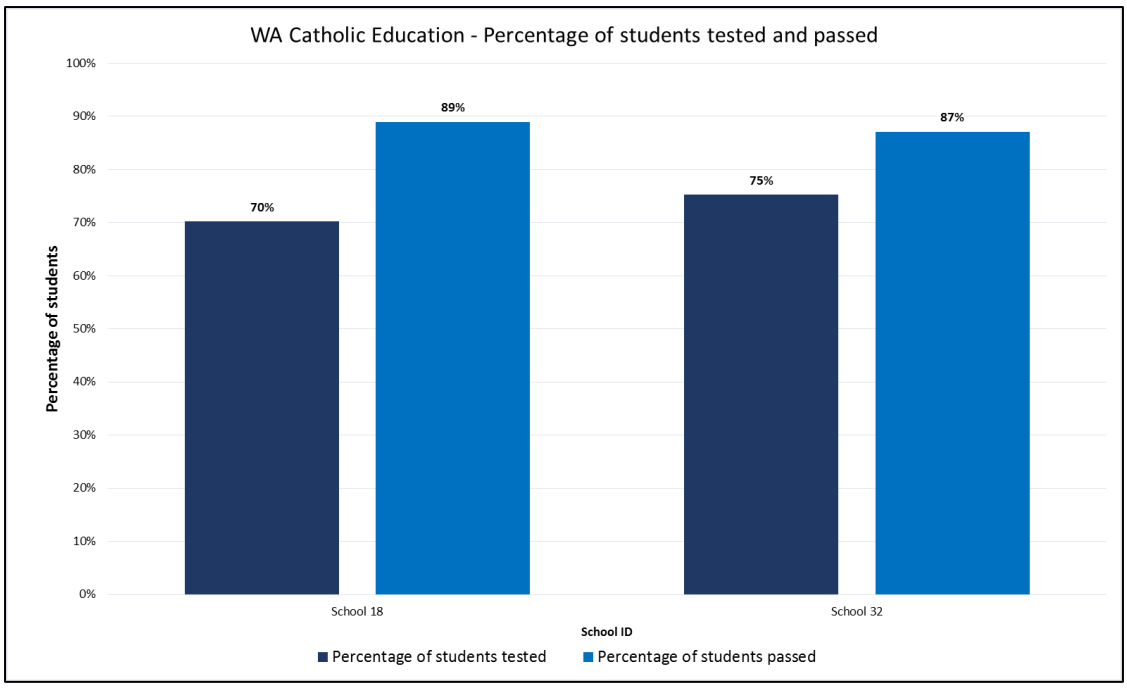


Figure 46. **Percentage of Students Tested and Percentage of Students Passes on DI Assessments for WA Catholic Education DI Schools.**

The DI program is timetabled to be delivered for 2.5 hours per day. This allows for a benchmark lesson progress rate across all groups of 1.0 lesson per day. Program data tracks the lessons taught so that an average lesson per day can be calculated at the school level. *Figure 45* shows this information and indicates the average lesson delivery per day was 0.8 lessons in each school. This data represents the number of lessons that schools complete each day, on average, and not the hours spent on instruction. It is possible that, for a range of reasons, teachers are unable to complete the full lesson within the time available. For example, students who have been absent and returned to school may need to be re-assessed to determine their lesson level. Hence, this will take time away from normal lesson teaching. While this data shows that schools are not reaching full fidelity in terms of lesson completion, it doesn't explain why, with further insight offered by the teacher survey.

DI program data for students provides indicators of student progress against mastery assessments. This information has been collated at the individual school level and is presented in *Figure 24*. This graph indicates that both schools are testing at least 70% of students and at least 88% are passing.

DOCUMENT ANALYSIS

Implementation. Implementation, as it was referred to in the Catholic school documents, was fairly straightforward: schools shared with parents details of the program, what it was, and when it was going ahead. Two schools switched from EDI to DI partway through the program: One school began 2015 with EDI but changed to DI in Term 4 of 2015. By the end of the term, DI had been successfully implemented and the program was "proving to be better suited to the learning needs of the students". Similarly, another school implemented EDI in their first year but re-trained teaching staff in January 2016 for implementation of DI in that school year; the school reported to parents that they "were impressed with the implementation of EDI and congratulated the staff on their dedication and commitment to teaching".

IMPACT: STANDARDISED ASSESSMENTS

The following section presents the results of standardised assessments that are considered indicators of impact on student outcomes.

JURISDICTIONAL LITERACY DATA - EYLND

The EYLND program is run as a partnership between the UWA and the Catholic Education Office of WA. It assesses student literacy progress from year one to year three. The assessment consists of a number of subtests: Burt Word Reading Test, Concepts About Print, Dictation, Duncan Word Test, Letter Identification, Running Record, and Writing Vocabulary, as well as a number of numeracy assessments which are not covered in this evaluation.

To explore literacy growth over time between program and control schools, student EYLND data from all participating WA Catholic Education schools and a sample of control schools was collected and matched. This process focused on matching students who were in year 1 in 2014 (pre- intervention), year 2 in 2015 and year 3 in 2016 (post-intervention). However, it should be noted that due to the EYLND testing timeframe, data from 2014 to 2015 reflects the 2014 year, while data from 2015 to 2016 reflects the 2015 year[11].

While there were 147 students in year 1 across all 13 schools in 2014, of these students only 71 students had matching data in 2015 and 2016, this data was also limited to 11 schools[12]. As not all schools were represented in the final sample an additional school group comparison was conducted to identify potential school level differences between program and control schools that may confound the analysis. A series of one-way ANOVAs were performed in conjunction with sensitivity testing (Welch's ANOVA and/or Kruskal–Wallis H test) to examine potential differences. The results show no significant differences between schools (see Table 14).

Table 14: *EYLND Schools Comparison Means and Standard Deviations*

	Control	Program	F	Sig. (2-tailed)
Fulltime Teachers	12.53 (10.73)	8.42 (2.06)	0.88	0.502
Fulltime Non-Teaching Staff	9.43 (3.50)	7.25 (2.99)	1.11	0.323
ICSEA	797.75 (253.26)	716 (36.67)	0.52	0.493
Fulltime Student Enrolment Numbers	174.75 (243.39)	71.5 (36.67)	1.11	0.46
Percentage of Indigenous Students	0.71 (0.48)	0.89 (0.22)	0.69	0.431
Percentage LOTE Students	0.65 (0.46)	0.84 (0.36)	0.53	0.487

ANALYSIS 1: CONTROL AND PROGRAM DIFFERENCES ON OVERALL EYLND RESULTS

Additionally, prior to performing analyses the data was inspected for missing data. Some missing data was present in the data set, thus a Little's Missing Completely at Random (MCAR) Test was conducted, which indicated that data was missing at random $\chi^2(669) = 22.62$, $p = .100$. As a result, an Expectation Maximisation

[11] EYLND is require for 2017 before an analysis.

[12] One school did not collect EYLND data in 2015 and 2016 and another smaller school did not have any matching data across all years.

was conducted to impute missing values[13].For the purpose of this analysis all EYLND subtests were averaged to create a composite score.

To explore change in composite EYLND results overtime, a two-way repeated measures ANOVA was conducted. A two-way repeated measures ANOVA is often used in studies where a dependent variable (in this case student literacy scores) has been measured over two or more time points, or when subjects have undergone two or more conditions (in this case program and control schools) (i.e., the two factors are "time" and "conditions"). The primary purpose of a two-way repeated measures ANOVA is to understand if there is an interaction between these two factors on the dependent variable; that is, whether literacy score change over time, and whether these changes are the same (or different) between the program and control schools.

A two-way repeated measures ANOVA was conducted to determine if student literacy scores on EYLND changed over time and whether there was any difference between the program and control schools. Preliminary assumption testing Analysis of the Studentized residuals showed that there was non-normality, as assessed by the by Kolmogorov-Smirnov test of normality. However, as skewness and kurtosis were within ± 1, the violation of the underlying assumption of normality was not deemed problematic, particularly given the robustness of ANOVA to this assumption breach. No outliers were present, as assessed by no Studentized residuals greater than ± 3 standard deviations. There was sphericity for the interaction term, as assessed by Mauchly's test of sphericity ($p > .05$); thus, Huynh-Feldt will be interpreted for the interaction.

Overall, literacy levels increased significantly in both program and control schools (main effect of time: $F(1.76, 121.45) = 307.15$, $p < .001$, $\eta_p^2 = .82$). However, the interaction between program type and literacy was not significant, $F(1.76, 121.45) = 2.91$, $p = .065$. As seen in *Figure 47*, program schools and control schools performed equally in the 2014 year (pre-intervention); however, program schools achieved higher scores in 2015 year (post-intervention) than the control schools. As no difference in 2014 is expected, a repeated measures two-way ANOVA focusing solely on the 2015 school year was conducted. The results of this analysis showed a significant difference between control schools and program school for the 2015 year $F(1, 69) = 7.66$, $p < .05$.

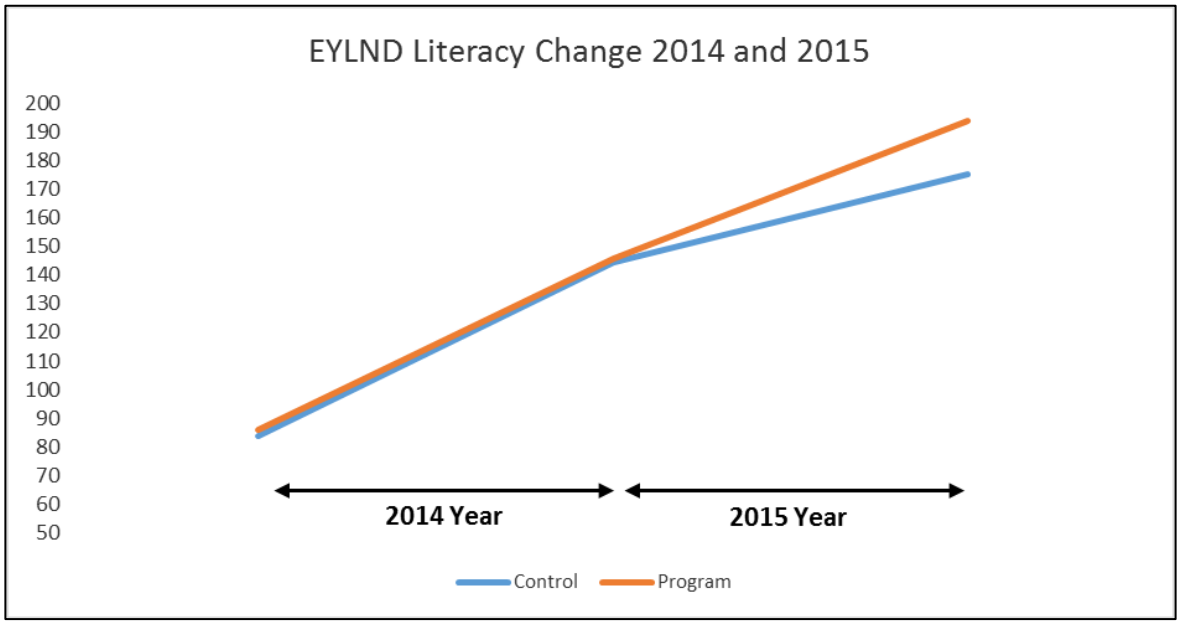


Figure 47. **Change in EYLND Literacy Scores for Program and Control Schools for 2014 and 2015.**

[13] A description of these procedures can be found in the Analysis Procedures Utilised section on page 30.

EFFECT SIZE CALCULATIONS

As there was no difference between the program and control schools in 2014, which is expected in a pre-intervention year, the magnitude of differences in 2015 only will be examined. For control schools the effect size change for 2015 was Cohen's *d* = 0.38, which is between a small and medium sized effect. While the effect size for program schools 2015 was Cohen's *d* = 0.64, which is between a medium and large sized effect.

ANALYSIS 2: CONTROL AND PROGRAM DIFFERENCES ON INDIVIDUAL EYLND SUBTESTS (2015)

A second series of one-way ANOVAs were performed to examine potential differences between program school and control school on the EYLND subscales. As seen in *Figure 48*, program and control schools showed similar scores on most subscales; however, there was a significant difference on Dictation.

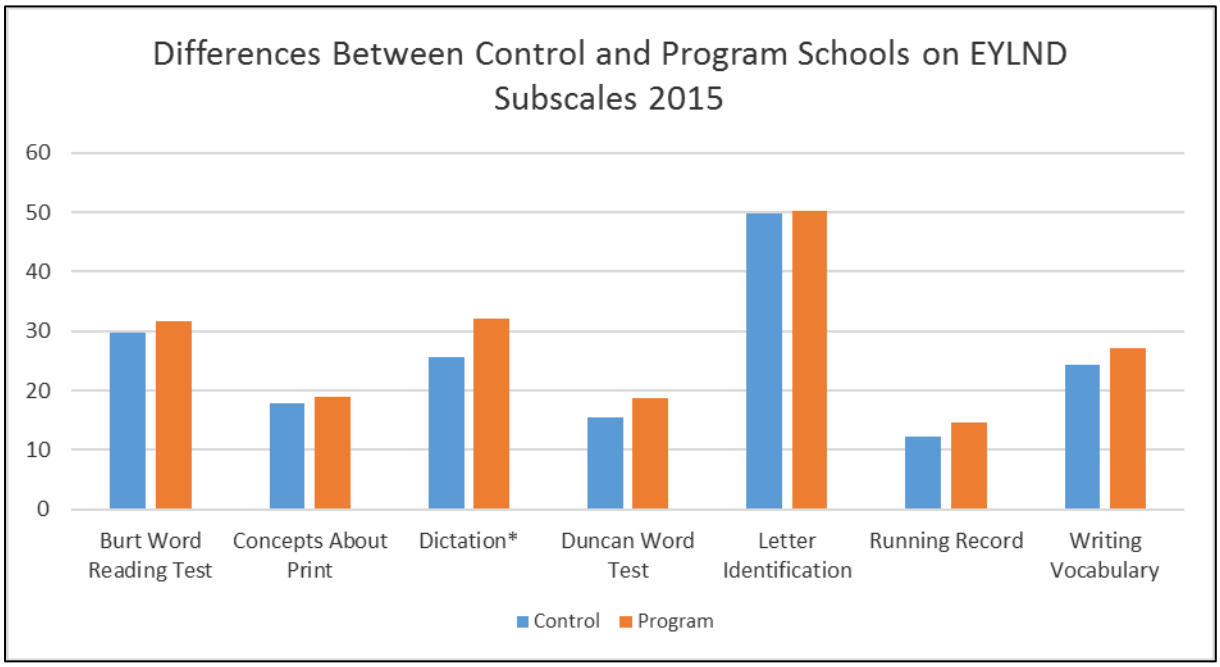


Figure 48. **Difference Between Control and Program on EYLND Subscales 2015.**

Note. *Significant difference detected, $p < .05$

NATIONAL ASSESSMENT PROGRAM FOR LITERACY AND NUMERACY (NAPLAN)

NAPLAN assessment scores formed part of four separate analyses:

- Cohort tracking compared with control schools between 2014 (pre-FLFRPSP implementation) and 2016 (post-FLFRPSP implementation)
- Change in each year level NAPLAN average scores over time (e.g. Year 3 NAPLAN for years 2014, 2015 and 2016)
- Percentage of students below National Minimum Standard (NMS) over years 2014, 2015 and 2016 for Grades 3 and 5
- Average change in NAPLAN scores between 2014 (pre-FLFRPSP implementation) and 2016 (post-FLFRPSP implementation)
- NAPLAN participation rates by year and grade level.

Each of these analyses is now presented for WA Catholic Education schools.

NAPLAN: PROGRAM SCHOOLS AND CONTROL SCHOOLS

To explore change in NAPLAN over time, both within and between the program and control schools, a series of independent *t*-tests were conducted comparing NAPLAN data from year 3 in 2014, and year 5 in 2016, for the NAPLAN subscales of Reading, Writing, Spelling, and Grammar and Punctuation. For this analysis change variables were created which reflect the change in NAPLAN subscale scores between 2014 and 2016 (Change Variable = 2016 – 2014).

After this analysis was conducted, and to better display results, mean scale scores for each subscale were displayed for 2014 and 2016 compared to the National Average and Very Remote National Average; these graphs can be found in *Figure 49*.

Preliminary assumption testing did detect some assumption violations, with the presence of minor outlier, and skewness/kurtosis between greater then ±1; however, there was homogeneity of variances for the NAPLAN scores for control and program schools, as assessed by Levene's test for equality of variances. Due to these assumption breaches, sensitivity testing is also conducted using Mann Whitney U tests. Further, as multiple analysis were conducted a Bonferroni correction was applied to adjust for Type 1 error (detecting a difference when none is present), significance values have been set to less than 0.01.

Overall, no differences were detected in overall growth on NAPLAN between control and program schools within the WA Catholic Schools cohort between 2014 and 2016 on NAPLAN reading $t(141) = -0.53$, $p = .597$; NAPLAN writing $t(139) = -1.04$, $p = .30$; NAPLAN spelling $t(140) = -1.40$, $p = .02$; or NAPLAN Grammar and Punctuation $t(140) = -1.55$, $p = .12$. These results suggest that both the control and program schools progressed at a comparable rate for students moving from year 3 in 2014 to year 5 in 2016.

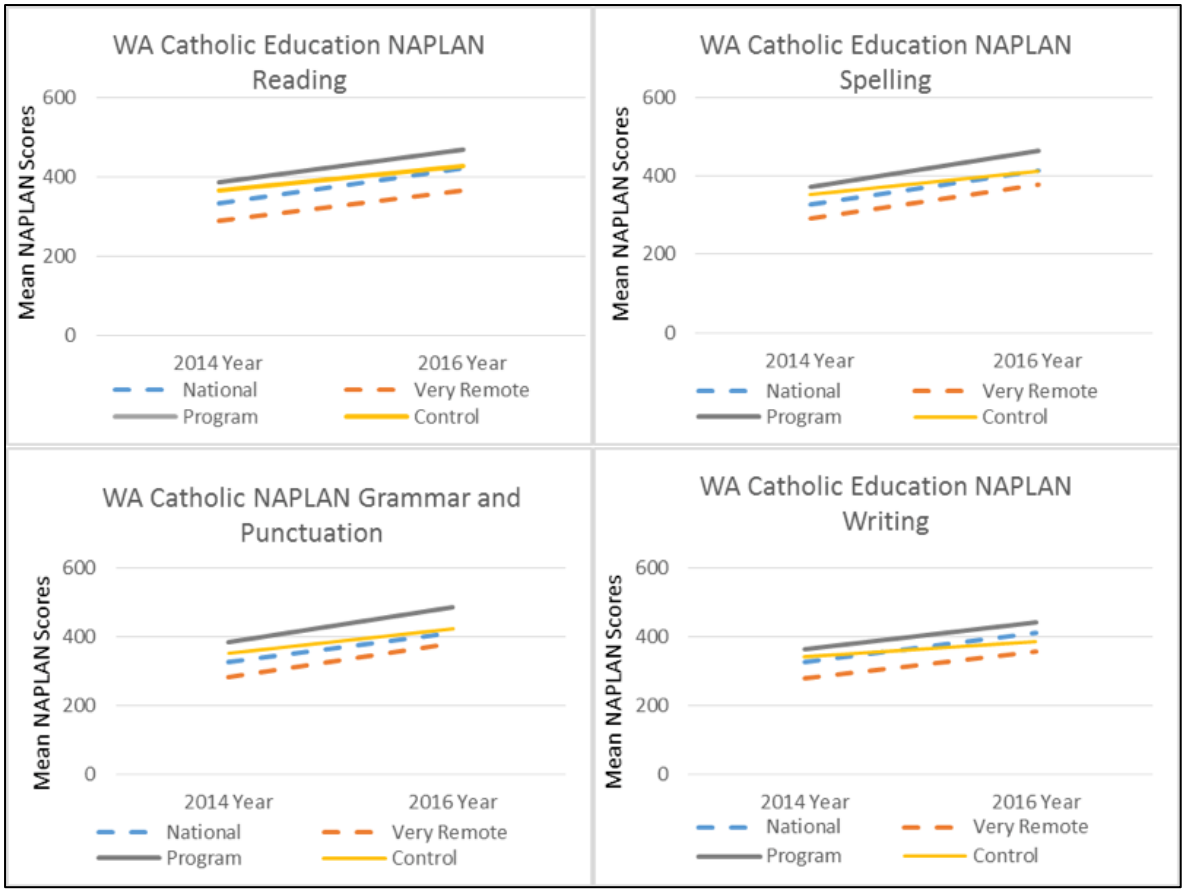


Figure 49. **Change in Mean NAPLAN Scores for WA Catholic Education Program Schools Against Control, National Average and Very Remote National Average.**

NAPLAN: VARIABILITY

To explore changes in student NAPLAN scores over time, mean scores for each NAPLAN domain were created for Year 3 and 5 for each school for 2014, 2015, and 2016. A One-way ANOVA was conducted to compare scores over this period. Preliminary assumption testing detected outliers, as assessed by inspection of a boxplot, and breaches the assumption of normality, as assessed by Shapiro-Wilk's test ($p > .05$), as well as values greater then ±2. As such, sensitivity testing was also conducted using Kruskal-Wallis H tests.

The results indicate no significant difference[14] over 2014, 2015 and 2016, for any NAPALN domains in the WA Catholic Education Schools cohort; reading $F(2, 39) = 0.12$, $p = .89$; writing $F(2, 39) = 2.00$, $p = .15$; spelling $F(2, 39) = 1.12$, $p = .34$; or grammar and punctuation $F(2, 39) = 0.45$, $p = .64$. Despite there being no significant differences, there is a trend for growth on all NAPLAN domains for year 3, and Writing and Spelling for year 5.

[14] Confirmed by Kruskal-Wallis H tests.

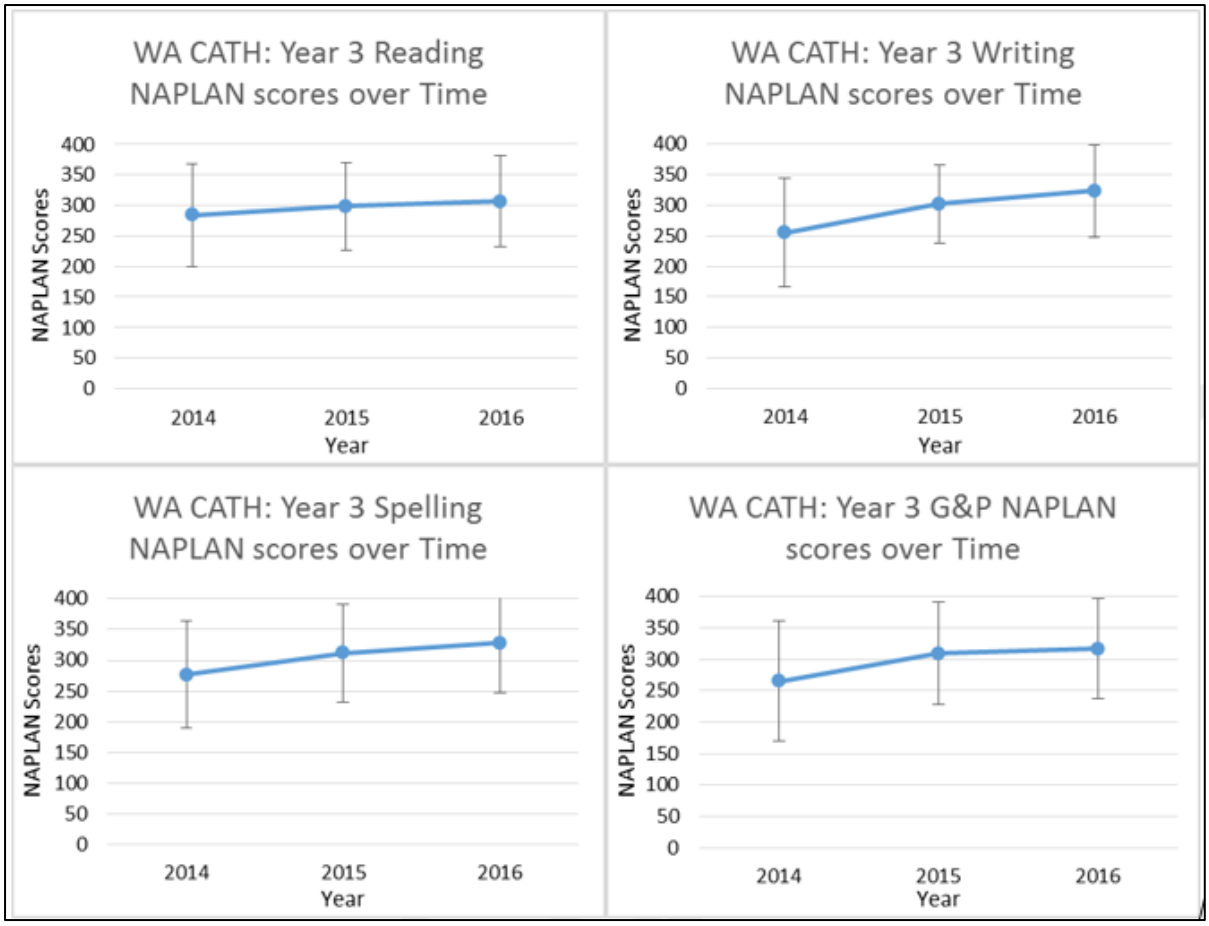


Figure 50. **WA Catholic Education Mean NAPLAN Scores 2014 to 2016 for Year 3.**

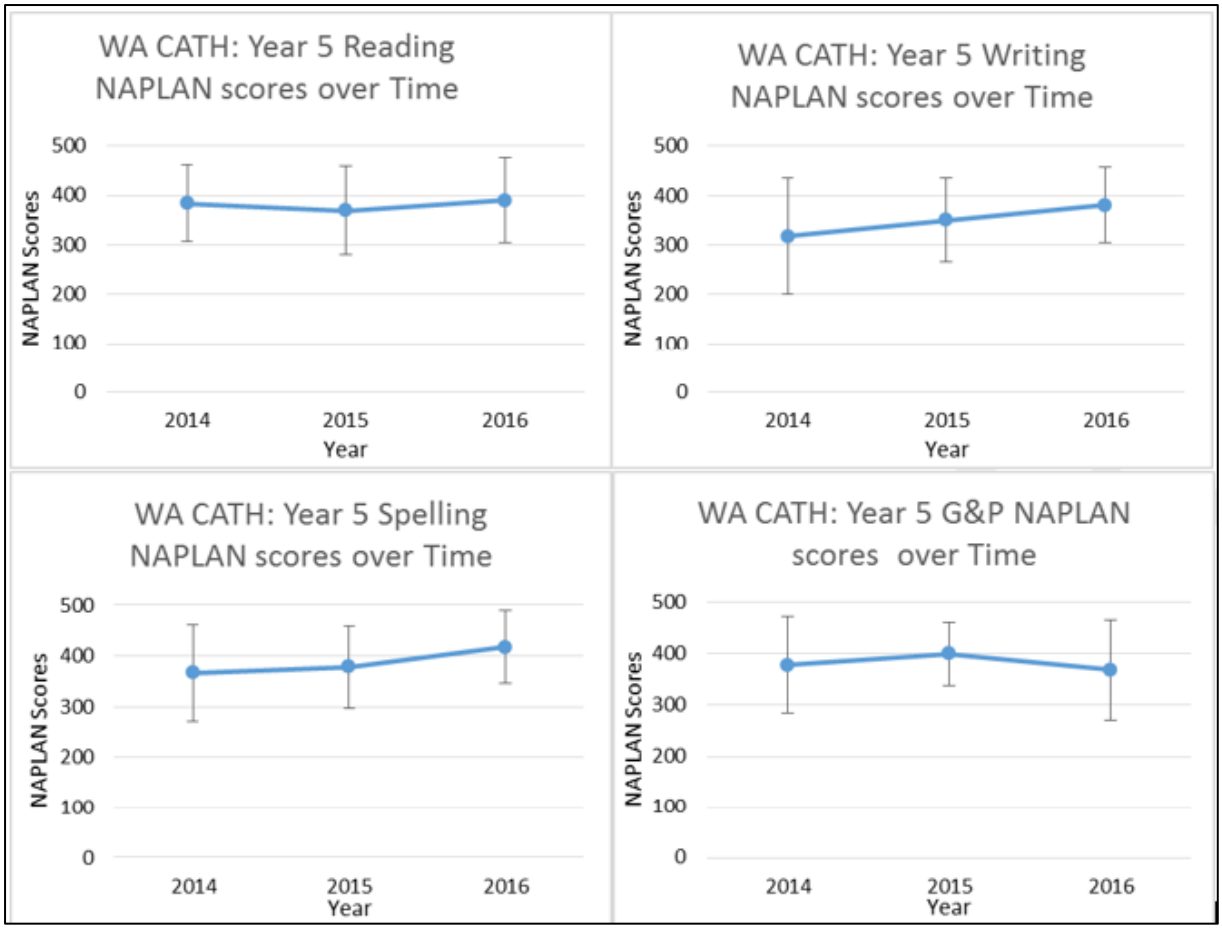


Figure 51. **WA Catholic Education Mean NAPLAN Scores 2014 to 2016 for Year 5.**

NAPLAN: PERCENTAGE OF STUDENTS BELOW NATIONAL MINIMUM STANDARD

Analysis of the change in the percentage of students below the national minimum standard (NMS) for years 2014, 2015 and 2016 and for Grades 3 and 5 are presented in *Figure 52* and *Figure 53*. Overall, there was a decrease in the percentage of students below the National Minimum Standard for students in year 3 across all fours subscales, with values lower in 2016 compared to 2014. This trend is not as apparent in year 5 across 2014, 2015, and 2016, with slight decreases on Reading, Writing, and Spelling, and a more pronounced change on Grammar and Punctuation. However, it should be noted that percentages of student below the National Minimum Standard were relatively lower in 2014 (pre-program), but have decreased further.

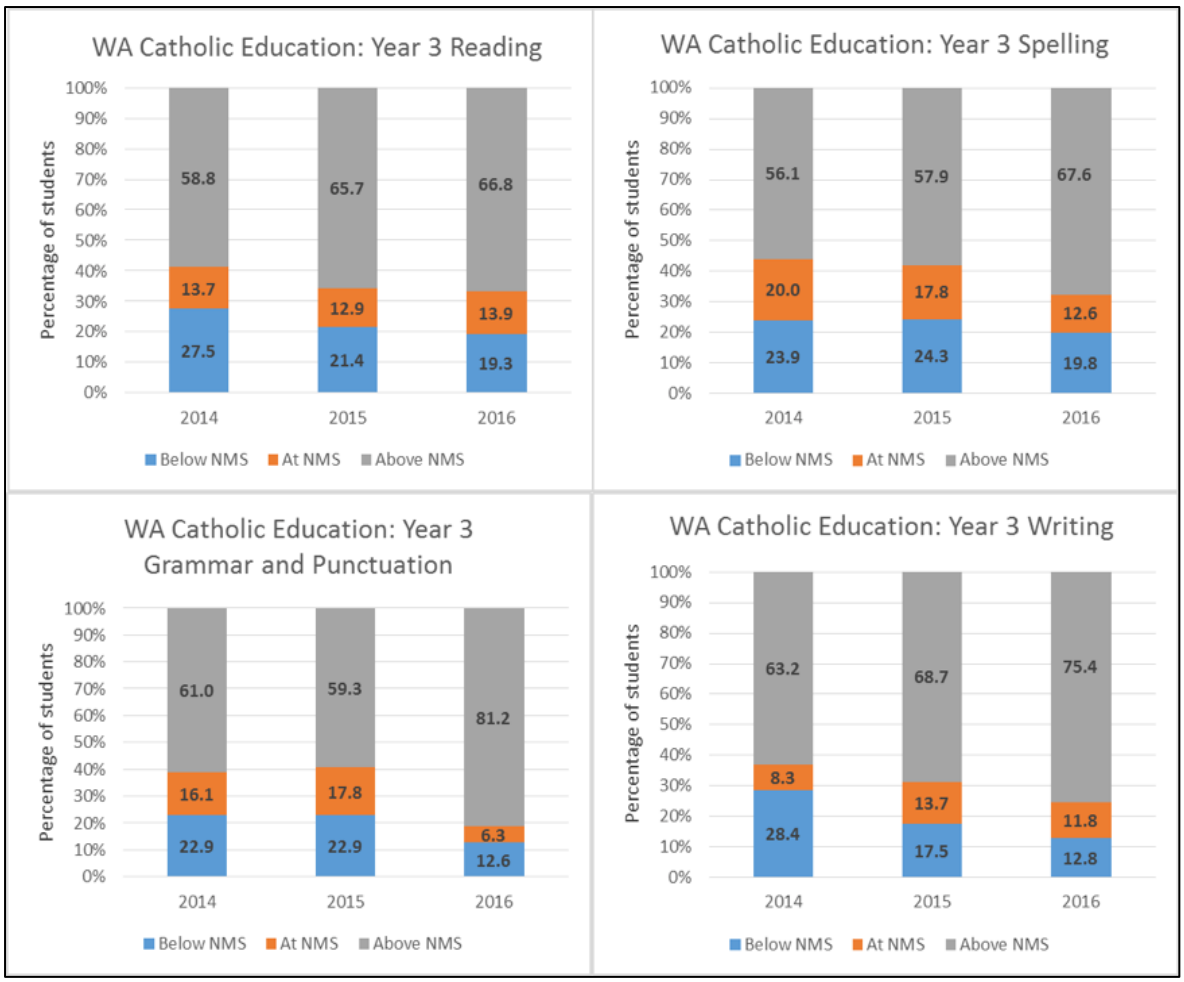


Figure 52. **Percentage of Students (year 3) Below NMS for WA Catholic Education Schools.**

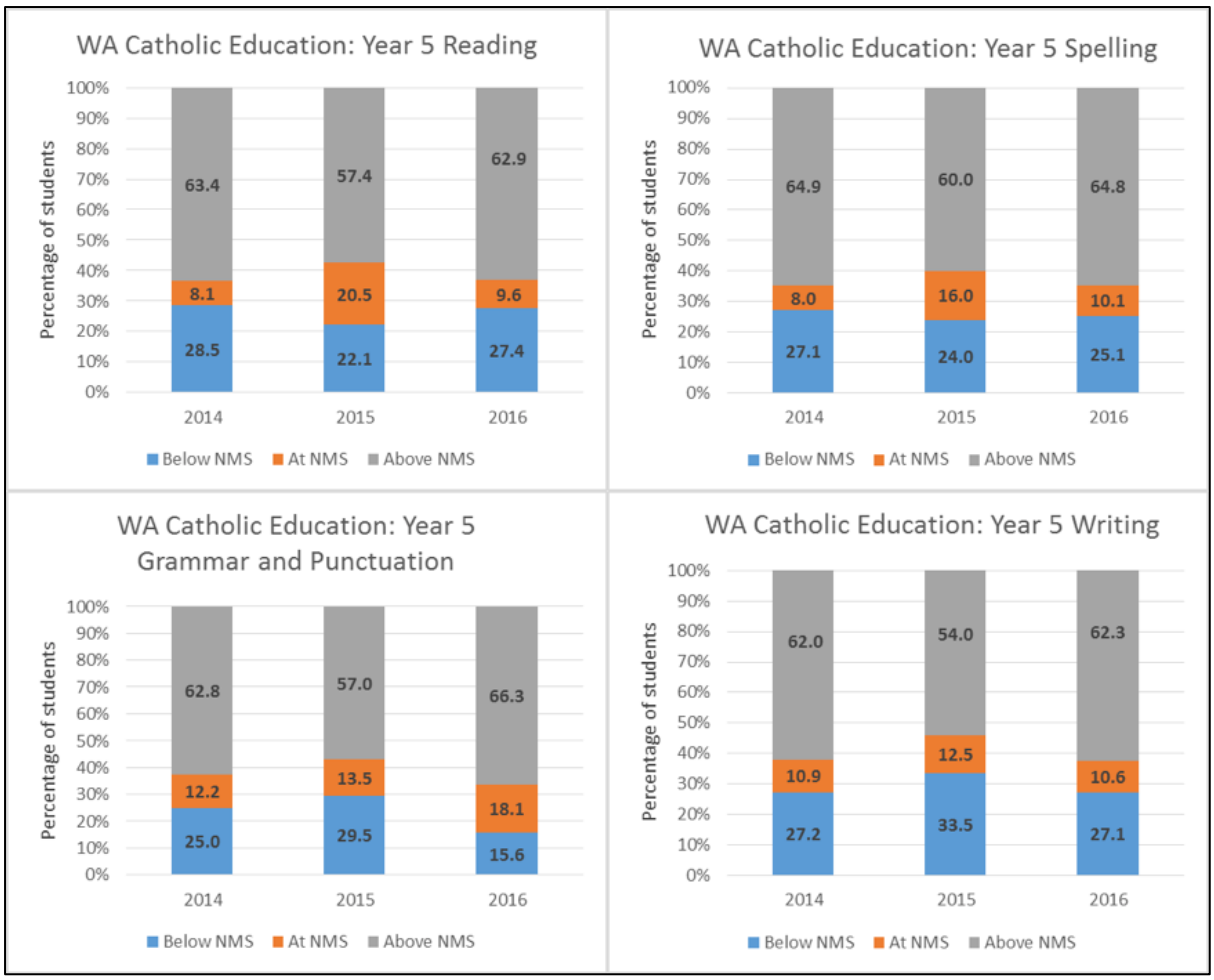


Figure 53. **Percentage of Students (year 5) Below NMS for WA Catholic Education Schools.**

NAPLAN: INDIVIDUAL SCHOOL ANALYSIS

At the individual school level, NAPLAN analysis focused on the mean (average) change in scores between Year 3 (2014) and Year 5 (2016) for the same cohort (unmatched), as well as confidence intervals as this analysis facilitates a pre-post implementation comparison. The results were compared with the state average gain for each assessment for WA. This was undertaken for all NAPLAN assessments and is presented in *Figure 54*. As shown in the graphs, Catholic Education schools in the program implementing either DI or EDI demonstrated larger mean gains in Spelling, and Grammar and Punctuation. However, it should be noted that the confidence intervals of these measures are much wider than those for the State average figures. In Reading, three schools were below State average.

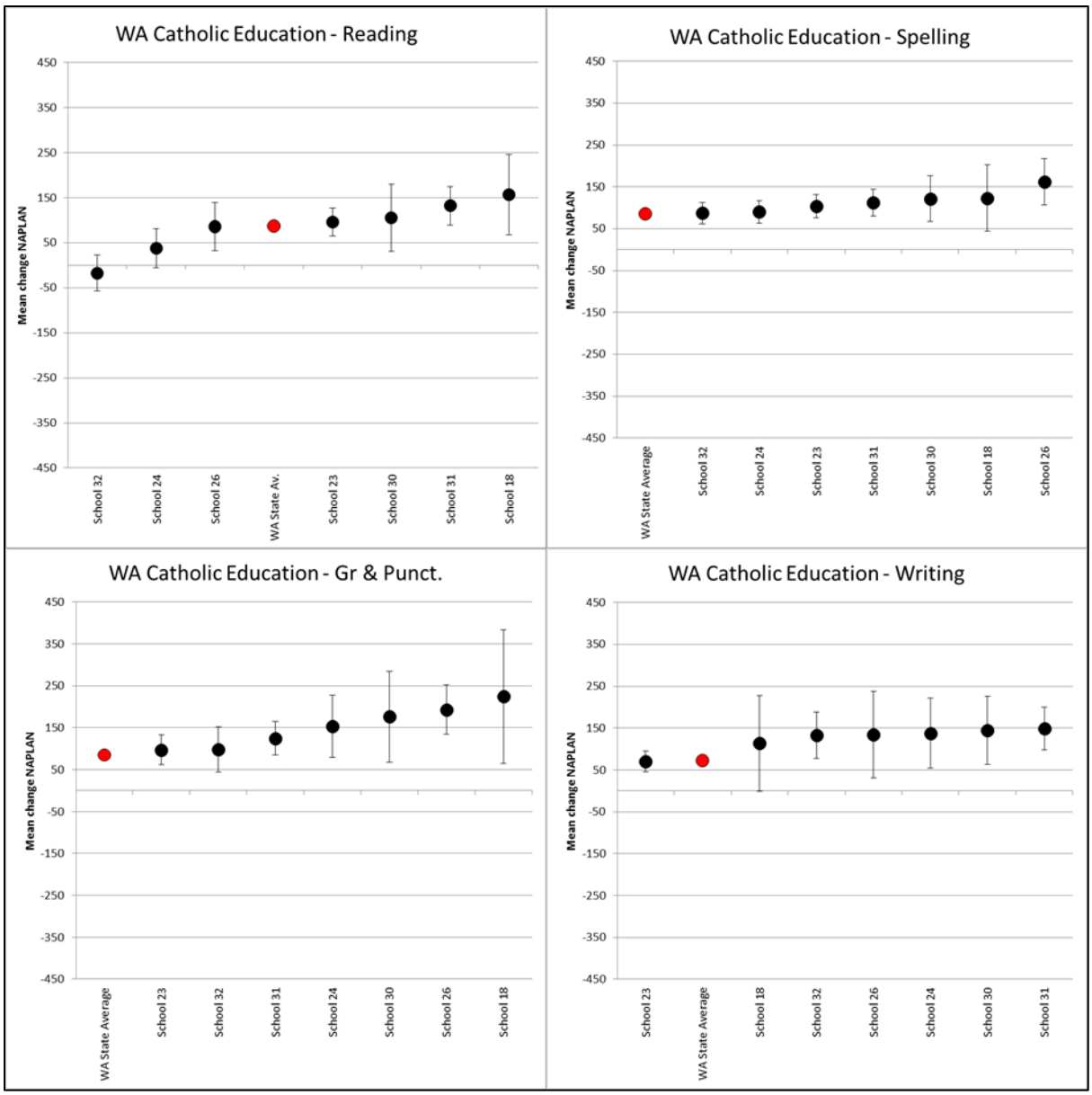


Figure 54. **WA Catholic Education Schools Mean NAPLAN Change Year 3 (2014) to Year 5 (2016).**

An average effects size3 for all WA Catholic Education Schools (n = 7) was also calculated based on the difference between on the mean (average) change in scores between Year 3 (2014) and Year 5 (2016). Schools demonstrated a medium to large effect size for change over two years; NAPLAN Reading Hedge's g = 1.20; NAPLAN Writing Hedge's g = 1.53; NAPLAN Spelling Hedge's g = 2.33; NAPLAN Grammar and Punctuation Hedge's g = 2.00.

NAPLAN: PARTICIPATION RATES

Despite NAPLAN participation rates falling well below the WA overall rate in 2014, 2015 and 2016, WA Catholic Education Office schools demonstrated increases in participation rates for each assessment type across the three years. Between 2014 and 2016 rates increase by between 8 and 10%.

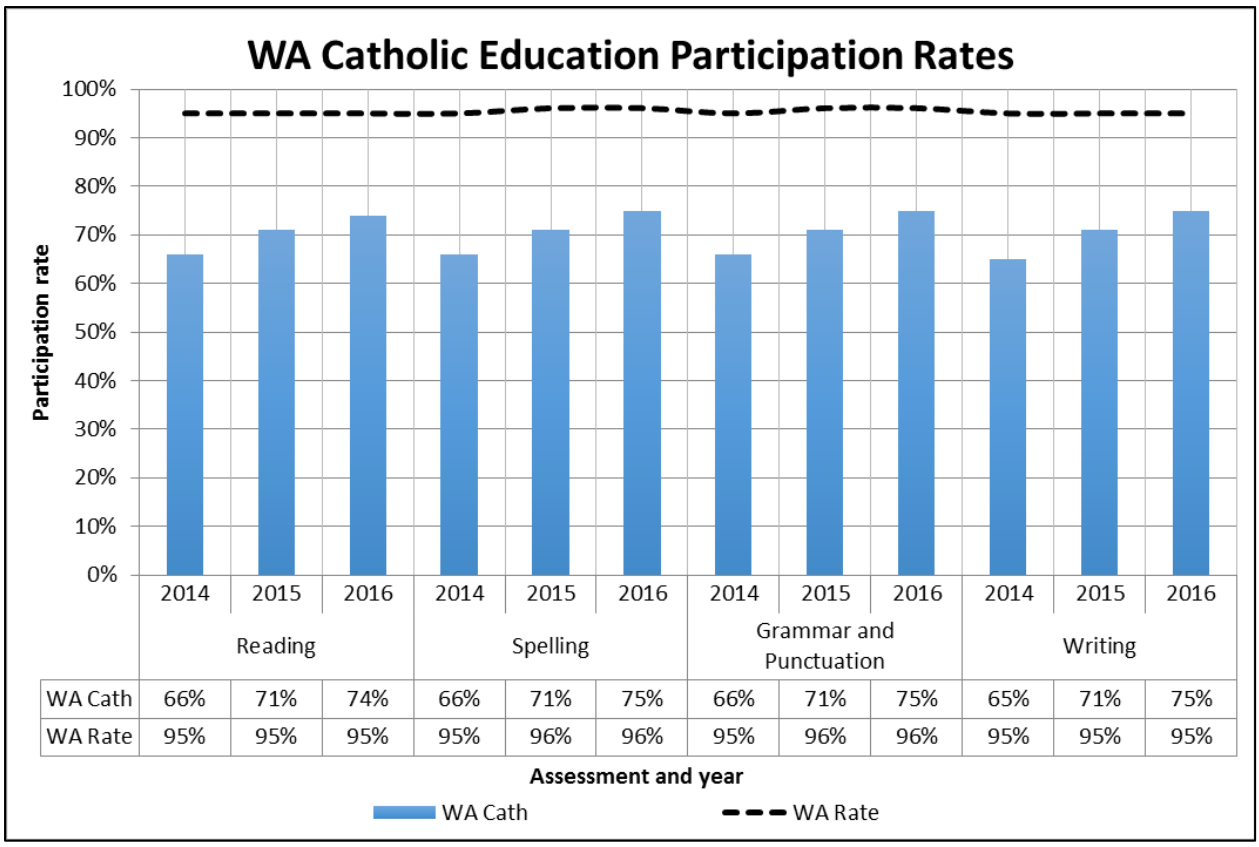


Figure 55. **NAPLAN Participation Rates by Assessment Type and Year for WA Catholic Education Office Schools.**

IMPACT: OTHER SOURCES

TEACHER SURVEY

There were 9 respondents from staff in 3 of the 7 Catholic Education schools involved in the program (please note that not all respondents answered all survey questions). Two of the three schools were implementing EDI and one DI. Those that responded were teachers (n=6), literacy support (n=1), Principal (n=1) and a teacher coach (n=1). There was a range of experience; however, all respondents apart from one had been working in education for at least 5-10 years with several having 15+ years' experience.

Sixty percent of respondents agreed or strongly agreed that the program had improved student literacy and made a difference for students. Furthermore, 80% of staff that responded felt that students had engaged with the program. One teacher commented,

> ***"I taught the current year 6 class in years 4 and 5 and their growth in literacy with NAPLAN and other testing was outstanding."***

Another observed,

> ***"According to our data, our literacy levels have improved in all areas across the school."***

In terms of other outcomes for students, perceptions of impact were more varied across the respondent group. One teacher suggested that students were more engaged and that their behaviour had improved.

All teachers indicated that there were small groups of children that did not seem to be improving or were improving at a slower rate. Reasons for this slower rate of progress included attendance, poor working memory/concentration, and lack of engagement.

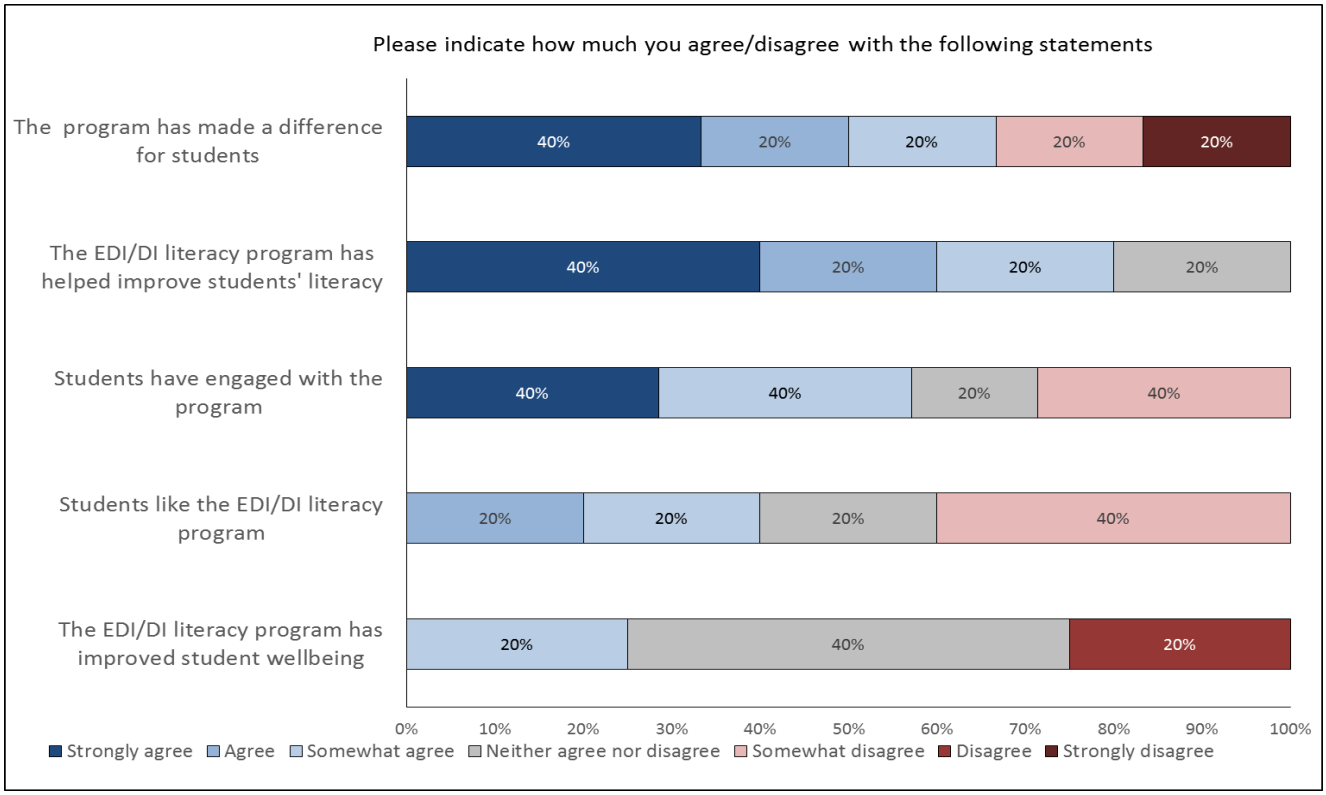


Figure 56. **Teacher Perception of The Impact of The Program on Students (n=8).**

PRINCIPAL INTERVIEWS

Principals from two Catholic Education program schools were interviewed as part of the evaluation. Their perceptions of the impact of the program on students were similar, both citing that the program had improved literacy skills across the year levels as well as in other areas such as behaviour. One of the Principals commented about the need to take a longer-term view of the impact on student outcomes. They remarked,

> ***"… our expectation is that in the next 2-3 years we are going to see a big jump in, well in everything. Unfortunately everyone looks at NAPLAN to see how well you are going - but while we have made steady improvement in that area, I think the real jump in that is going to happen in the next year and the year after. "***

DOCUMENT ANALYSIS

Recurring themes throughout the WA school documents focused on results of DI/EDI and included literacy progress, student engagement, and attendance. The Catholic Education schools in WA also tended to address their experiences with the implementation process of DI as well as its outcomes. Information regarding student outcomes information is detailed below.

Literacy progress. Schools noted positive, though nonspecific, student progress in reading, spelling, and/or writing skills. One school's Annual Report noted early success in literacy scores with EDI, citing PIPS and EYLND results for their Year 1s. Another school reported that in 2016, Year 3s achieved greater scores in Grammar and Punctuation than in Reading, possibly due to DI's emphasis on oral language. Their Prep and

Year 1 students constantly moved forward, making progress in their DI content. This school also noted that high-achieving Year 2s and 3s were not held back by lower-achieving students, due to successful testing and streaming conducted by the DI manager.

Student engagement. DI/EDI was seen as the cause of increased student engagement in several schools. In some cases, the predictable and consistent structure of DI was also especially beneficial. Several newsletters published by one school included quotes from students about the program: "It's a better and more interactive way of learning" (Year 6); "I enjoy doing it, because we write and share answers" (Year 6). Prep teachers at this school also noted that their students "have shown lots of enthusiasm during their EDI lessons. They are particularly enjoying learning their High Frequency Words. They are stars!"

Attendance. Overall, the implementation of DI/EDI has seen an increase in attendance. Low attendance, however, is also a negative factor affecting the effectiveness of DI. Holden (2016) notes that "in the first four months of [one school's] DI implementation, from 2 February to 2 May, attendance increased significantly by between twenty and thirty percent for all classes" (p. 12).

Managing trauma. It was noted that DI helped with managing the effects of trauma among students. One school "highlighted in their interviews that the routine, predictability and order of DI is having positive benefits in managing the trauma that is unfortunately experienced by many students" (Holden, 2016, p. 12), and the structure and routine of the DI program enables students with special needs, for example students with autism, to become more engaged with the content and with the classroom. The Literacy Key Leader of another school also noted the benefits of DI to this effect: "It's the structure [of DI]. The children know what is coming next. They love the predictability and routine. They know what's happening every day. They know how to answer. All these things are very important for children who are experiencing trauma" (Holden, 2016, p. 30).

Other outcomes. There were few negative outcomes mentioned in the literature. Some schools had to manage the appropriateness of the program for students. Two Catholic schools both changed from EDI to DI during the project based on suitability to their students. Additionally, one of these noted that although they had observed positive outcomes in their students, the 2.5 hours of literacy required by the program were a challenge.

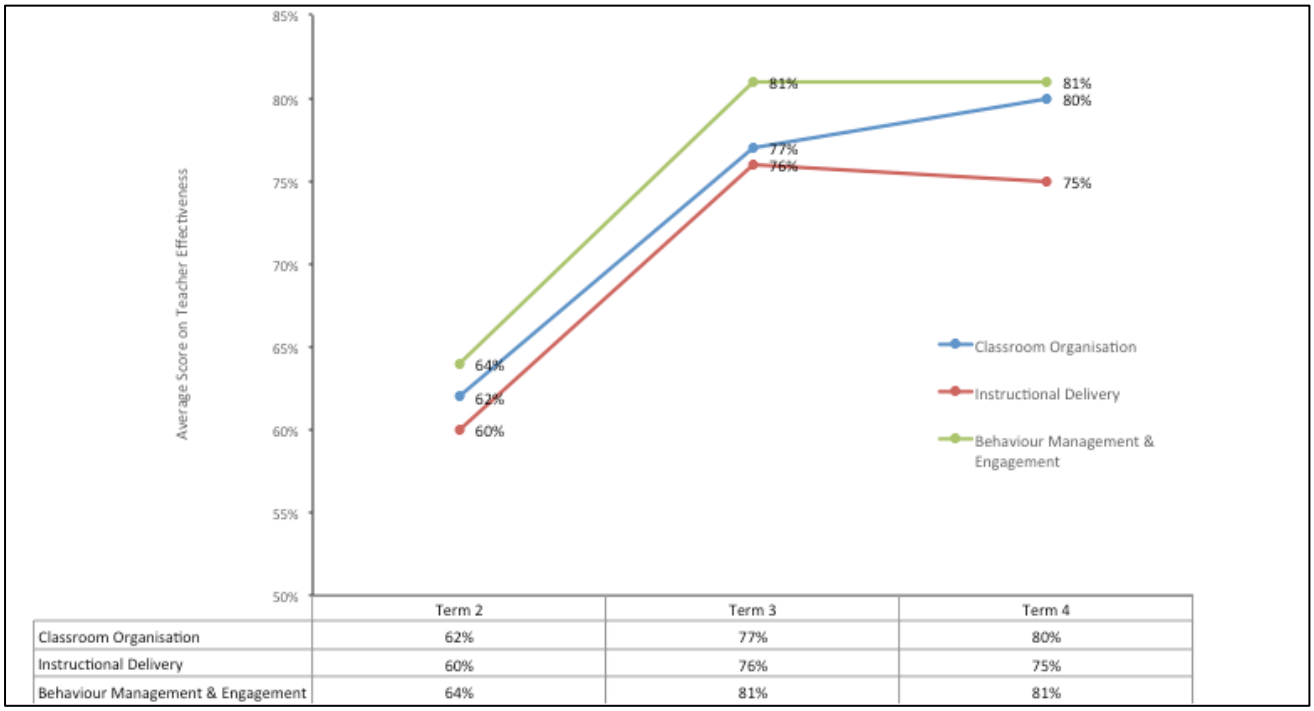
TEACHERS

Teachers in Catholic Education program schools have on average 11.5 years of experience teaching. Teacher turnover was only provided for two schools for the 2016 school year and is shown in Table 15 below. Relative to other jurisdictions, turnover is lower in the two schools for which data was received for both teacher and teaching assistant turnover. It should be noted that data for all Catholic Education program schools was not provided so a full understanding of teacher turnover in these schools is not possible.

Table 15: ***Teacher and Teaching Assistant Turnover 2016 for Catholic Education Schools (n=2)***

	2016	
School ID	Teacher Turnover	Teaching Assistant Turnover
School 18	19%	57%
School 26	20%	_

Both the EDI and DI programs provide a measure of teacher effectiveness across the same three dimensions: classroom organisation, instructional delivery, and behaviour management. *Figure 57* shows the teacher effectiveness scores for all WA Catholic Education schools for 2016 only from Term 2. As shown, teacher effectiveness scores showed a steep increase between terms 2 and 3, then plateaued.



	Term 2	Term 3	Term 4
Classroom Organisation	62%	77%	80%
Instructional Delivery	60%	76%	75%
Behaviour Management & Engagement	64%	81%	81%

Figure 57. **Teacher Effectiveness Measures for All WA Catholic Education Schools.**

Teachers were also asked to rate both their confidence in and their perception of the impact the program had on their pedagogy and teacher skills using a 7-point Likert scale from Strongly Agree to Strongly Disagree. In terms of confidence regarding their role in the program, the pedagogical approach, the specific tasks and activities, and their ability to teach literacy, respondents were overwhelmingly positive in their responses (see *Figure 58*).

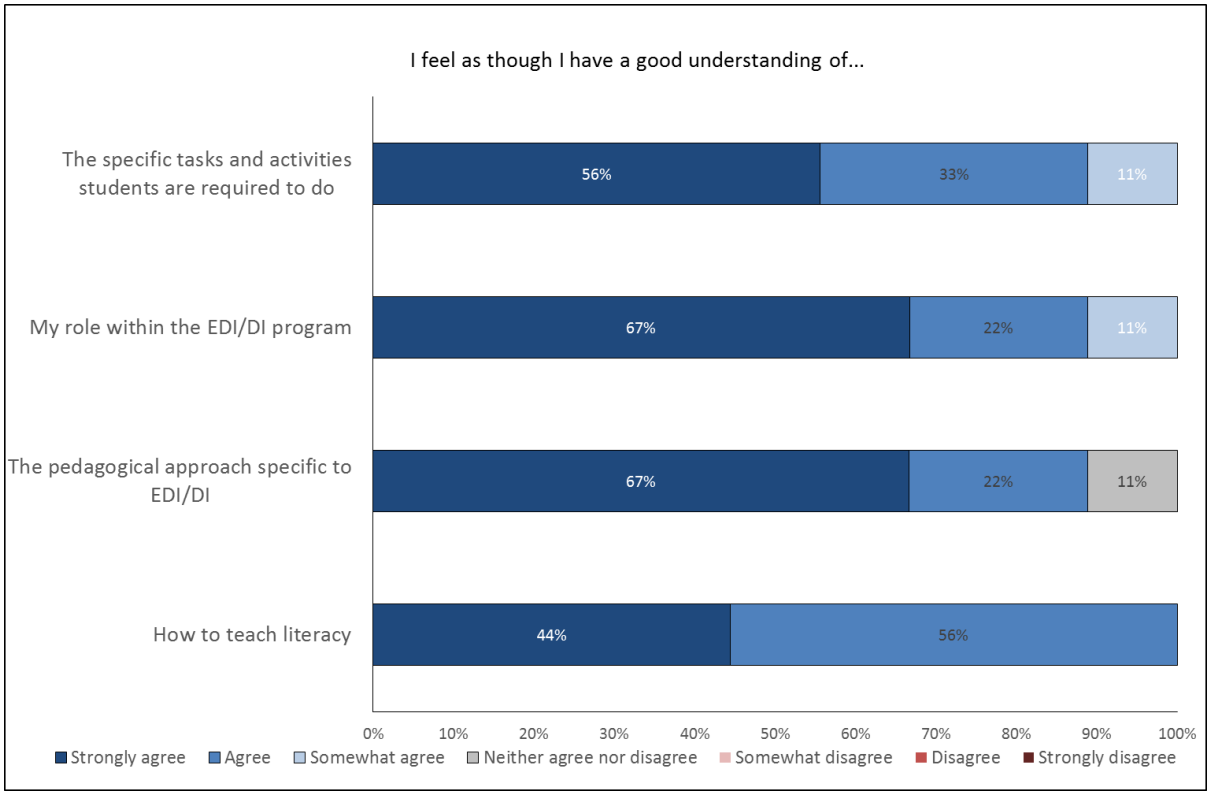


Figure 58. **Teacher Perception of Their Understanding of The Program (n=9).**

There were more mixed responses with regard to the impact of the program on their teaching skills and the fit the program had within their school. The majority of respondents did not agree that the program gave them better teaching skills or improved their ability to teach literacy. Nonetheless, 50% of respondents did feel that the program was a good match for their school and was important for supporting the teaching staff. Furthermore, these views did not indicate overall that the teachers would not work with the program. In fact all respondents said they would continue with the program if given the opportunity, mainly because they could see the growth in student skills and changes in other behaviours.

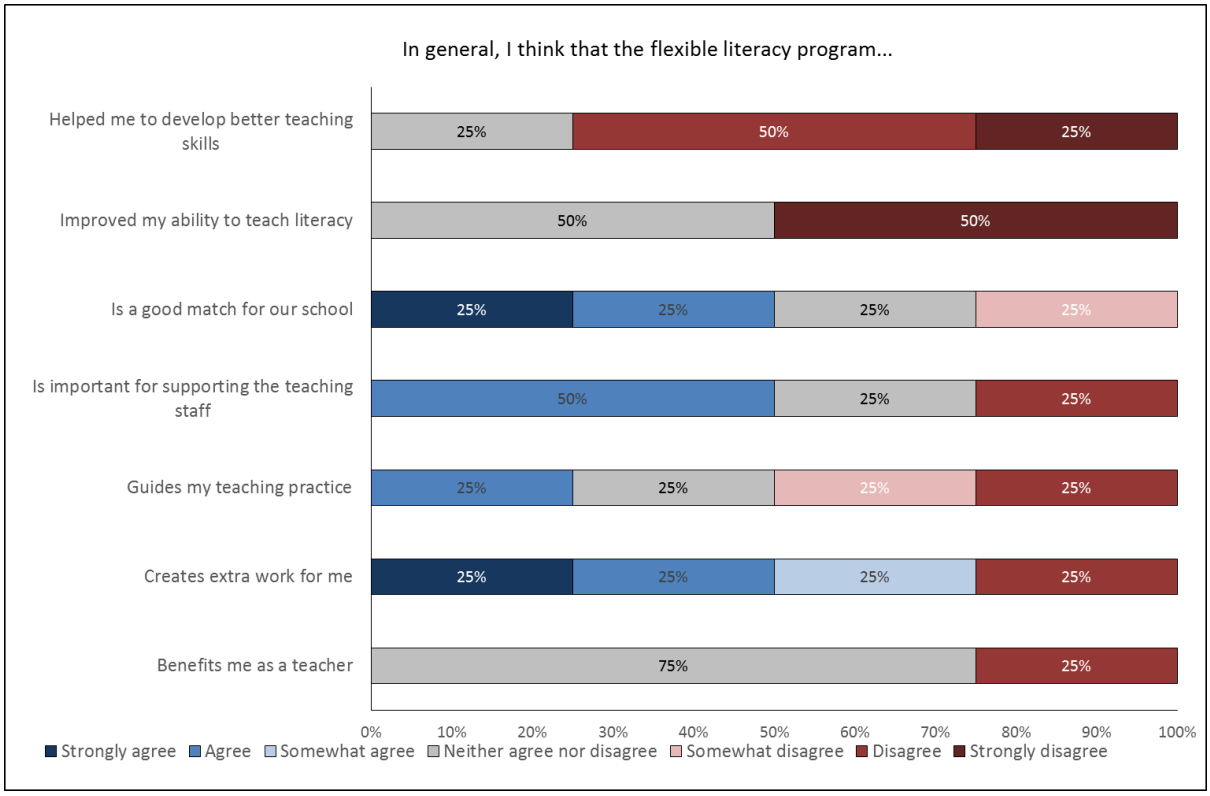


Figure 59. **Teacher Perception of the Program's Impact of their Role and Work as Teachers (n=9).**

SCHOOL LEVEL DATA

Relevant demographics, in addition to several indicators of program impact and implementation, have been considered at the school level. Demographic data includes publicly available information such as that available on the MySchool website (myschool.edu.au) as well as provided information such as turnover in leadership staff. As for NT and WA Government schools, program data includes measures of school fidelity and program support provided by GGSA through training and observations. Finally, perceptions of impact from teachers and principals are also included where relevant to the school as a whole.

Figure 60 shows turnover rates for principals and instructional coaches for DI schools in WA Catholic Education between 2015 and 2017 only. Instructional coaches are school-based staff who provide on-site coaching support to teaching personnel. Note that in the data table '0' refers to no data whereas '0%' reflects no turnover. Turnover of these leadership roles with respect to the program is high across most schools with the majority replacing a leadership role at least once per year. This has implications for the program in terms of continuity, advocacy, and implementation.

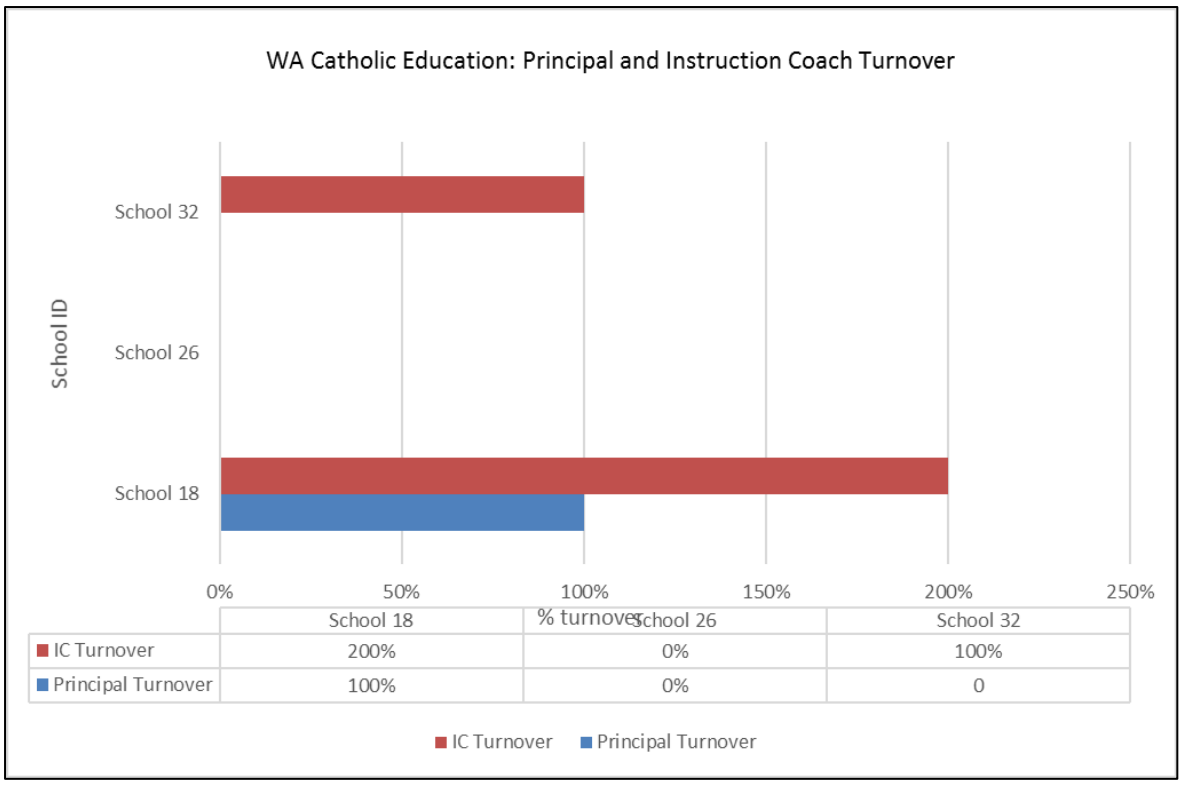
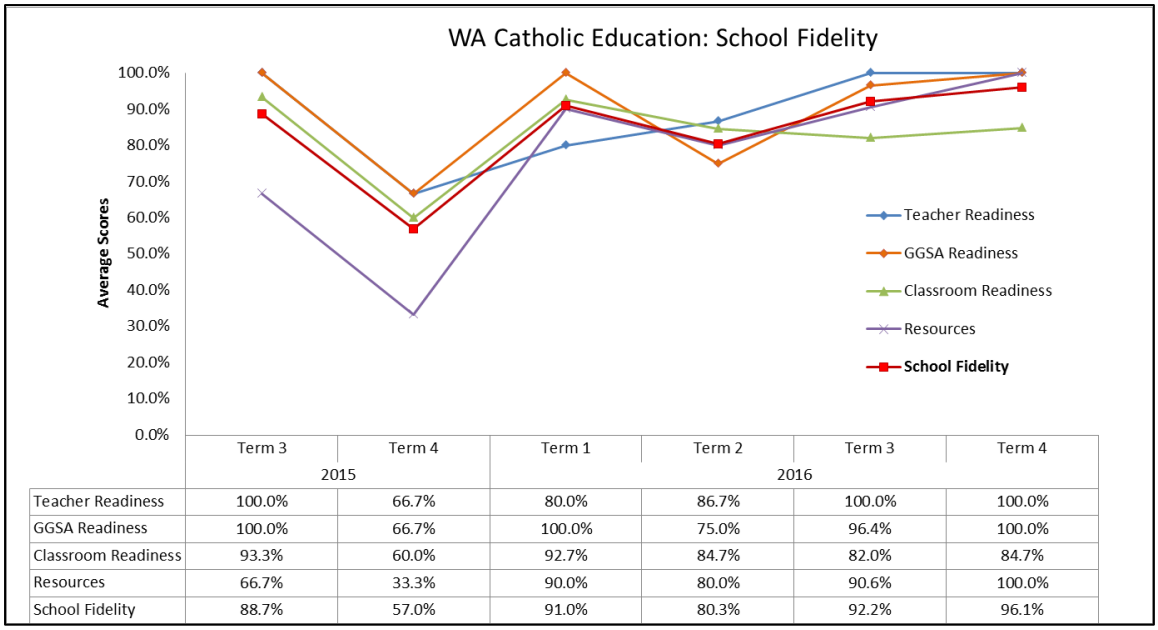


Figure 60. **Principal and Instructional Coach Turnover 2015 To 2017 for WA Catholic Education Program Schools (DI only).**

IMPLEMENTATION

At the school level, understanding of implementation is measured primarily with program data. School fidelity comprises four factors: teacher readiness, GGSA readiness, classroom readiness, and resources that are then combined into an overall score. Table 8 describes each of these factors in more detail and *Figure 61* shows scores for all factors and overall across 2015 and 2016. Similar to NT Government schools, WA Government schools show some variability across all four factors over time. GGSA readiness was consistently high and indicates that support and monitoring from GGSA is well implemented. Teacher readiness scored lowest of the four factors, which again may reflect teacher turnover rates which have a flow-on effect to other school-based factors such as classroom readiness and resource readiness.



	Term 3 2015	Term 4 2015	Term 1 2016	Term 2 2016	Term 3 2016	Term 4 2016
Teacher Readiness	100.0%	66.7%	80.0%	86.7%	100.0%	100.0%
GGSA Readiness	100.0%	66.7%	100.0%	75.0%	96.4%	100.0%
Classroom Readiness	93.3%	60.0%	92.7%	84.7%	82.0%	84.7%
Resources	66.7%	33.3%	90.0%	80.0%	90.6%	100.0%
School Fidelity	88.7%	57.0%	91.0%	80.3%	92.2%	96.1%

Figure 61. **School Fidelity Scores for WA Catholic Education Schools in FLFRPSP for 2015 and 2016.**

IMPACT

Several questions in the survey of teachers addressed the potential impact of the program at a school level. These questions focused on the drivers of implementation and aspects of the program that were easy or difficult to implement.

WA Catholic Education teaching staff identified several positive program drivers, including the support provided from GGSA and NIFDI or DataWorks, initial training, and strong leadership support and commitment to the program. They commented that student attendance could be a problem, and that schools did not necessarily have the staff and resources to catch up students who have been absent on community business. One teacher commented that group variability can be a challenge to teach and that timetabling of other activities such as camps or professional development days can make it difficult to cover the program content. All teachers replied that they would continue with the program if given the opportunity.

The two Principal interviews shared mixed views about the impact of the program on the school and community. Both felt that there were positive impacts on student outcomes in literacy and other areas. However, one principal felt that aspects of implementation were quite difficult and that fit to context was an important factor for the program developers to consider in terms of materials and other resources.

The second principal commented,

> ***"It's become embedded in what we do here, and it's just the way that we teach English. So when I employ people- I say to them, look- this is an EDI school, we teach Explicit Direct instruction – and if you want to come and teach at this school- this is what you must do- there's no compromise there- this is what you have to do if you come here this is what you must do."***

PROGRAM SUPPORT

Program support refers to implementation and monitoring assistance provided by GGSA, in-school coaches, or program developers (NIFDI, DataWorks).

Figure 62 shows the average number of observations and average teacher training for NT program schools in 2015 (Terms 3, 4) and 2016. Training data is the average number of training sessions per school, and includes behaviour training, DI program training, and support training for teachers, teaching assistants, and school leadership variously. Observations are the average number of observations per school, and consist of two-minute observations, five-minute observations, and extended observations. These observations were conducted by an implementation manager, principal, instruction coach, teacher coach, or teaching principal. As to be expected, the WA Catholic data shows there to be a spike in the number of observations and training sessions earlier in the year. Overall, school staff participated in a high number of training and observations.

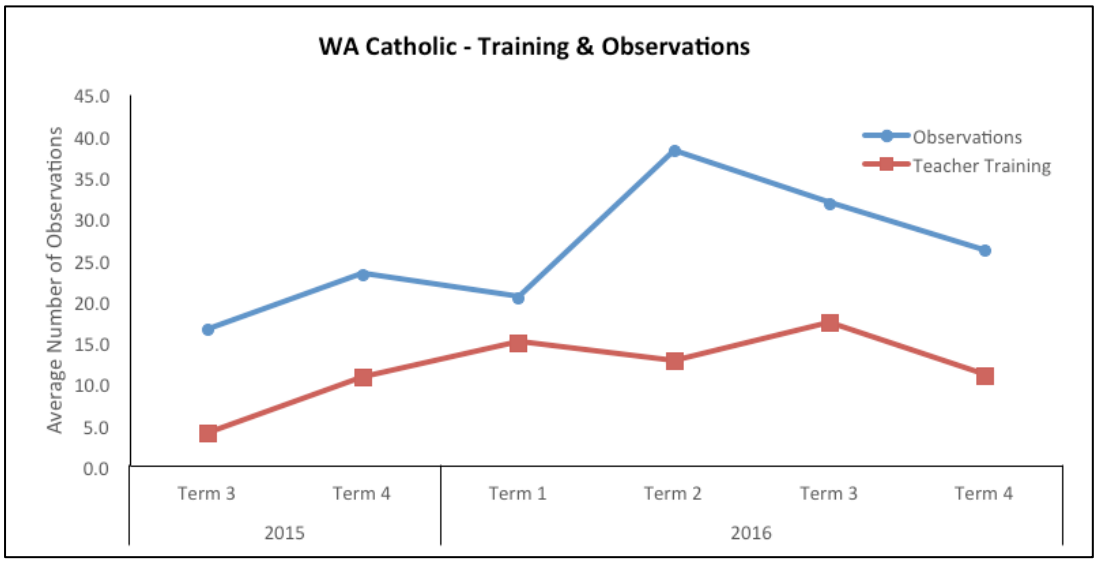


Figure 62. **Average Number of Training Sessions Attended and Coaching Observations by Term for WA Catholic Education Schools.**

Teaching staff were also asked in the survey their perceptions of their readiness to teach DI or EDI following training. *Figure 63* displays the results of this question for teaching staff in WA Government schools. The majority of staff agreed that they were ready to implement the program, had enough resources and understood the potential benefits of the program following training. They also felt that they were supported to implement the program.

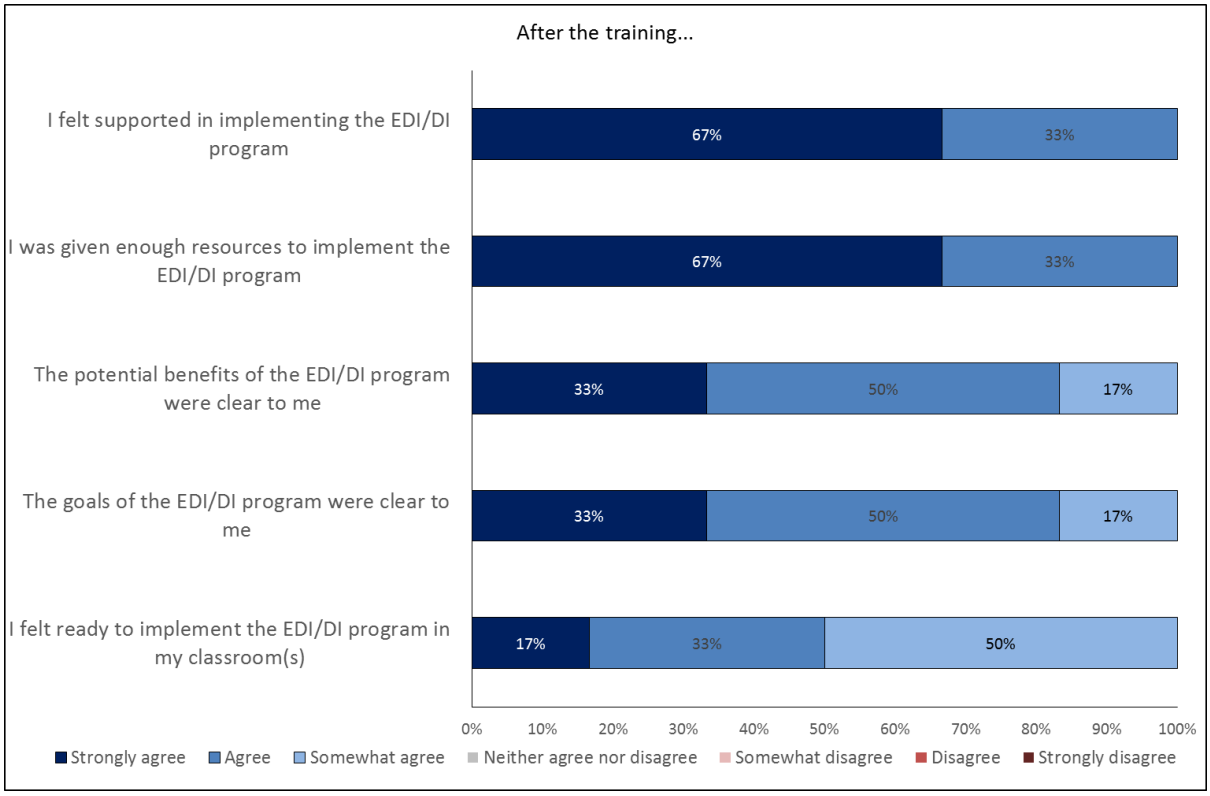


Figure 63. **Teacher Perception of the Training (n=6).**

SUMMARY

In summary, there are more consistent results from WA Catholic Education than compared to the other two jurisdictions.

For students in WA Catholic Education schools jurisdictional literacy data, EYLND shows early signs of positive impacts from the program. Program schools showed greater gains than control schools (measured as effect size), and 2016 EYLND levels were significantly higher than control schools. The NAPLAN analysis shows that Catholic Education program schools were higher than control schools in mean NAPLAN scores, and gains with the Spelling domain demonstrating a significant difference. The percentage of students below NMS reduced in all domains for Year 3 and in Writing and Grammar and Punctuation for Year 5. Mean gain in NAPLAN scores were higher than the State average for all schools except one across Spelling, Grammar and Punctuation, and Writing domains. Four schools saw higher mean gains than the State average for Reading. NAPLAN participation rates were generally low but increased over the three years.

Teacher survey data and Principal feedback on student impact suggested that teachers perceive the program as having a positive effect on student literacy and that they are engaged with the program. Attendance, on-task behaviours, and working memory were cited as reasons for some students not improving, but they were considered in the minority.

At the teacher and school level, turnover is lower than NT and WA. Interestingly, WA Catholic Education demonstrated higher teacher readiness scores within the school fidelity measure, which may be influenced by smaller turnover rates. Teacher effectiveness measures showed a large initial increase and then stabilisation at a high level. Teachers also reported being well prepared to teach the program and understand the pedagogical approach following initial training. Program data measuring overall school fidelity against four

domains indicates that schools have consistently achieved excellent scores across all four domains and overall in 2016.

Identified drivers[15]:

- Lack of regular student attendance (-)
- Teacher and leadership turnover (+)

HEAT MAP OF WA CATHOLIC EDUCATION PROGRESS IN FLFRPSP

Figure 64 illustrates for each data source the extent to which schools in the jurisdiction are operating below, at, or above expected levels for this stage of the evaluation. This is based on review of all the available evidence and an evaluative judgment by the evaluation team. It provides a visual summary of the status of the jurisdiction and will inform the next phase of the evaluation. It also provides a basis on which progress can be monitored and re-assessed. Overall, it indicates that WA Catholic Education program schools are exceeding expectations for this phase of the evaluation in several areas.

	Implementation	Impact		
	Program data	Proximal primary data	Distal primary data	Other primary data
Students	Lesson progress and mastery assessments	School/jurisdiction literacy measures	NAPLAN	Teacher and Principal report
Teachers	Training and observation (GGSA provided)	Teacher effectiveness program measures	N/A	Teacher and Principal report
Schools	School fidelity	N/A	N/A	Teacher and Principal report

Figure 64. **Heat Map for WA Catholic Education Schools.**

Rating criteria

Above expectations for this phase of the evaluation	Meets expectations for this phase of the evaluation (mixed results – some aspects require attention while others are good.	Below expectations for this phase of the evaluation (requires substantial improvement)	Not enough information available	Not applicable N/A

[15] Enablers are identified with a positive sign, while barriers are identified with a negative sign.

NAPLAN ANALYSIS OF ALL PROGRAM SCHOOLS

NATIONAL NAPLAN

To explore NAPLAN change for all program schools, data from all schools were matched between 2014 and 2016 for both program and control schools. This process focused on students who sat NAPLAN in year 3 in 2014 (pre- intervention), and NAPLAN in year 5 in 2016 (post-intervention); this was a pool of 774 students.

Data was examined for anomalies[16] using the 2014 and 2016 NAPLAN Equivalence Tables. During this screening eight NAPLAN scale scores of zero were identified; as this value is not a valid scale score they were removed and coded as missing. Anomalies in attendance data were also found, with two students having NAPLAN data but an attendance rate of zero. These attendance scores were removed and coded as missing. Further, eight students had changed schools from a program to a control schools or vice versa, and were not appropriate for the analysis and therefore removed.

Some missing data was present in the data set; thus, a Little's Missing Completely at Random (MCAR) Test was conducted. This indicated that data was not missing at random $\chi^2(394) = 1036.86$, $p < .001$.The, data was investigated for patterns. Examination revealed 46 (5.9%) of students do not have NAPLAN scores in 2014 and 2016, 297 (38.3%) have data at only one time point (either 2014 or 2016), and 431 (55.6%) students have data at two time points. There was also a substantial portion of students (28.3%, *n* = 219) who did not have attendance data.

As 44.2% of student had data missing at either one or both time points an analysis was conducted to determine if a relationship existed between student test scores and attendance rates. Students were assigned to a group based on the nature of their reported NAPLAN data. The groups were *No Data* (student with no reported data), *One-Time Point* (students with data reported at only one-time point), *Two-Time Points* (student how had data at two-time points) and a One-Way ANOVA was conducted to determine if attendances rates varied between these groups.

Assumption testing for ANOVA was conducted, while attendance data normally distributed (skewness and kurtosis between ±1), with not outliers, the assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances ($p = .008$). As a result *Welch's F* was interpreted with Games-Howell used for post hoc testing.

A one-way Welch ANOVA was conducted to determine if attendance rates differed by NAPLAN group. The analysis revealed a statistically significant difference on attendance rates between groups *Welch's F*(2,442) = 45.80 $p < .001$, $\eta^2 = 0.17$. A Games-Howell post hoc test revealed student attendance was significantly higher for students with data at *Two Time-Points* (128.32 ± 35.15 $p < .001$) when compared to the *One Time-Point* (98.56 ± 33.96) and *No Data* group (75.41 ± 33.84). There was also a statistically significant difference between the *One Time Point* or *No Data Points* groups ($p < .001$). These results suggest that attendance likely played a role in non-participation in NAPLAN testing as students with two time points of data have significantly higher attendance rates then students with one time point or no data (as can be seen in *Figure 65* [17]). The effect size was rather small, with only 17% of the variance in missing score explained by overall attendance.

[16] Scale score that are not possible according to the NAPLAN Equivalence Tables.

[17] Sample numbers in graph differ due to missing attendance data.

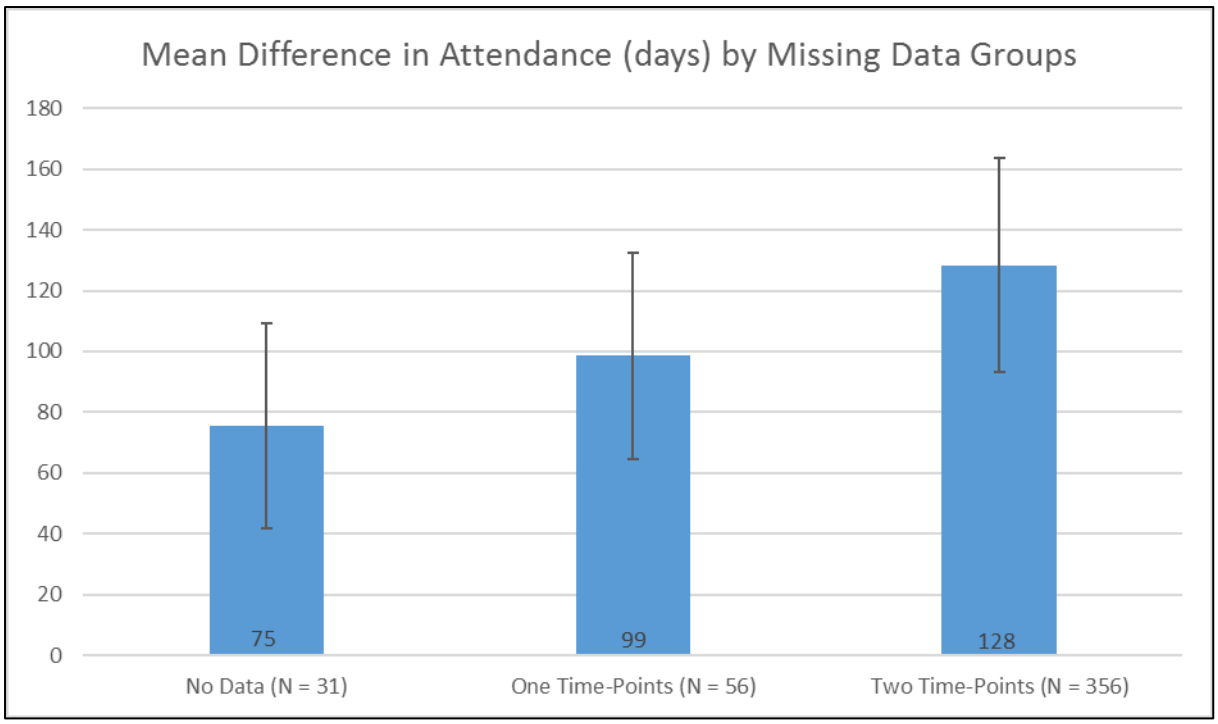


Figure 65. **Mean Difference in Attendance (Days) by Missing Data Groups.**

Due to the level of missing data (greater than 5%), Multiple Imputation is the most suitable method to manage missing data, and ensures the retention of the highest number of students for analysis. However, as the no data group does not have sufficient data to impute replacement values reliably, they have been removed from the analysis, leaving a final sample size of 728. The remaining missing values will be imputed based on attendance and other NAPLAN scores.

The multiple imputation was conducted[18] and, while a two-way mixed ANOVA was an ideal analysis for this scenario, preliminary assumption test uncovered a violation of the homogeneity of variances assumption and no correction is available. As a result, this method was not used and alternative testing was conducted.

An Independent *t*-test was conducted comparing NAPLAN Change scores between control and program schools. Preliminary assumption testing did not detect any assumption violation. There were no outliers in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. NAPLAN change data was normally distributed (skewness and kurtosis between ±1), and there was homogeneity of variances for the NAPLAN change scores for control and program schools, as assessed by Levene's test for equality of variances (p = .627). Thus, no alternations are required. Further, as multiple analysis were conducted a bonferroni correction was applied to adjust for Type 1 error (detecting a difference when none is present), significance values have been set to less than 0.01. When results from the original analysis align with the results of the pooled multiple imputation results, original results are reported.

READING CHANGE

NAPLAN scores for 372 students from control schools and 351 students from programs schools were included in the analysis. An independent-samples *t*-test was run to determine if there were differences in mean Reading

[18] Multiple imputation model uses a Mersenne Twister random number generator with a Fixed Set Starting Point 2000000 based on 20 imputations. Constraints were set based on the 2014 and 2016 NAPLAN Equivalence Tables.

Change scores between control and program schools. The results indicate no statistically significant difference in mean Reading Change scores between control and program schools, $t(396) = 1.50$, $p = .13$[19].

Table 16: ***NAPLAN Reading Means, Standard Deviation and Effect Size for Control and Program Schools 2014 and 2016***

	Year 3 2014		Year 5 2016		
	Mean	SD	Mean	SD	Cohen's d
National	332.90	94.00	422.10	89.60	**0.97**
Control	257.13	124.54	348.92	128.63	**0.73**
Program	278.17	140.17	378.67	144.23	**0.71**

WRITING CHANGE

NAPLAN scores 203 students from control schools and 196 students from programs schools were included in the analysis. An independent-samples *t*-test was run to determine if there were differences in mean Writing Change scores between control and program schools. The results indicate no statistically significant difference in mean Writing Change scores between control and program schools, $t(397) = 2.14$, $p = .03$[20].

Table 17: ***NAPLAN Writing Means, Standard Deviation and Effect Size for Control and Program Schools 2014 and 2016***

	Year 3 2014		Year 5 2016		
	Mean	SD	Mean	SD	Cohen's d
National	327.90	95.10	412.40	82.90	**0.95**
Control	246.93	110.98	351.76	105.74	**0.97**
Program	268.93	127.11	370.29	120.20	**0.82**

SPELLING CHANGE

NAPLAN scores for 210 students from control schools and 200 students from program schools were included in the analysis. An independent-samples *t*-test was run to determine if there were differences in mean Writing Change scores between control and program schools. The results indicate no statistically significant difference in mean Writing Change scores between control and program schools, $t(402) = -1.40$, $p = .16$[21].

Table 18: ***NAPLAN Spelling Means, Standard Deviation and Effect Size for Control and Program Schools 2014 and 2016***

	Year 3 2014		Year 5 2016		
	Mean	SD	Mean	SD	Cohen's d
National	327.90	95.10	412.40	82.90	**0.95**
Control	287.72	90.59	377.12	90.86	**0.99**
Program	287.86	105.63	397.73	105.67	**1.04**

[19] Original results align with pooled multiple imputation results.

[20] Original results align with pooled multiple imputation results.

[21] Original results align with pooled multiple imputation results.

GRAMMAR AND PUNCTUATION CHANGE

NAPLAN scores for 210 students from control schools and 200 students from programs schools were included in the analysis. An independent-samples *t*-test was run to determine if there were differences in mean Writing Change scores between control and program schools. The results indicate no statistically significant difference in mean Writing Change scores between control and program schools, $t(397) = -0.12$, $p = .91$[22].

Table 19: *NAPLAN Grammar and Punctuation Means, Standard Deviation and Effect Size for Control and Program Schools 2014 and 2016*

	Year 3 2014		Year 5 2016		
	Mean	SD	Mean	SD	Cohen's d
National	327.90	95.10	412.40	82.90	**0.95**
Control	237.12	145.85	361.82	135.25	**0.89**
Program	248.33	174.99	393.71	154.86	**0.88**

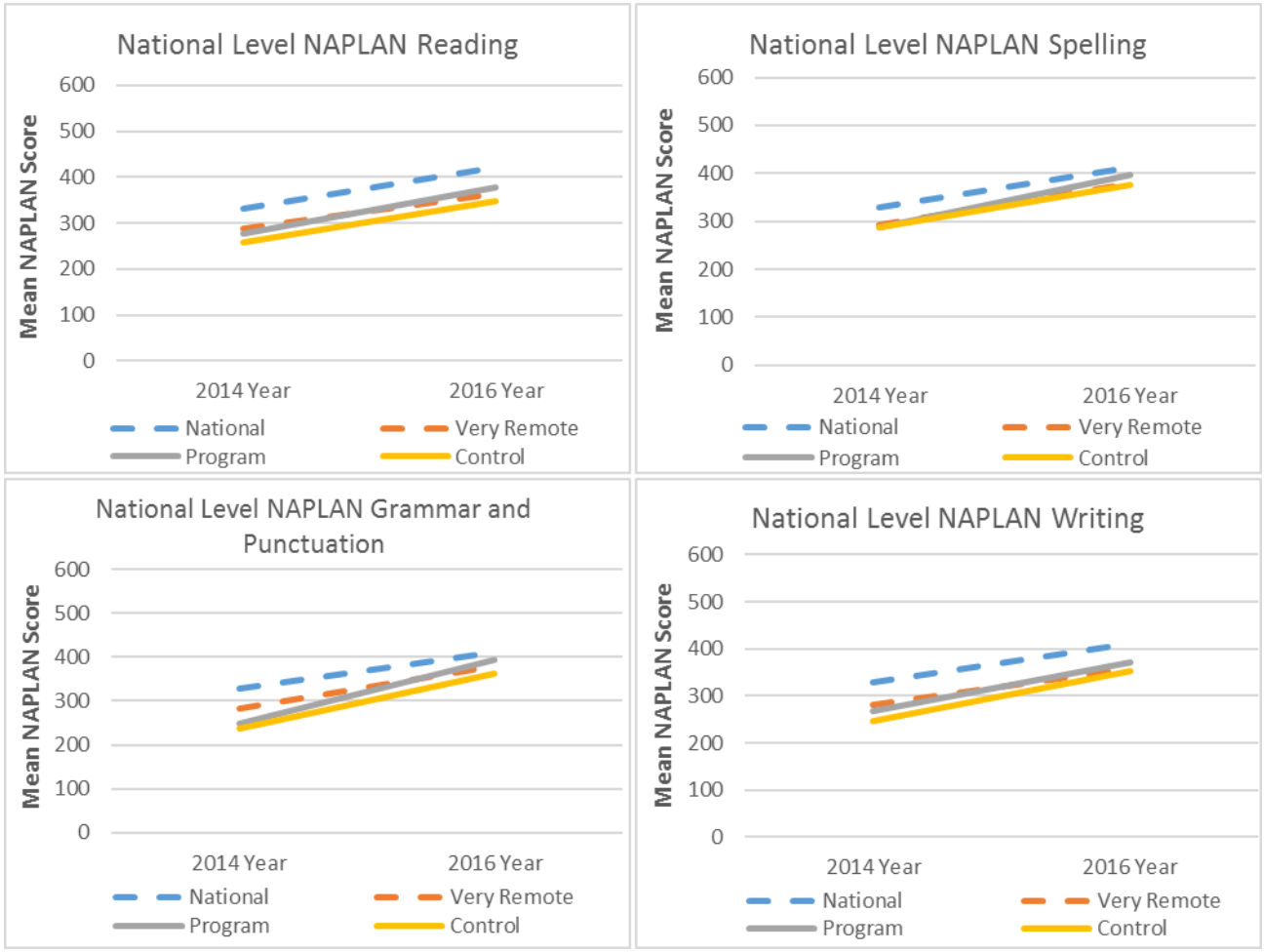


***Figure 66.* Change in Mean NAPLAN Scores for all Program Schools Against Control, National Average and Very Remote National Average.**

[22] Original results align with pooled multiple imputation results.

CONCLUSIONS

The following section outlines the collective view of the evaluation based on triangulation of all the available evidence and with reference to relevant theory and research. It describes an overall conclusion regarding this phase of the evaluation, presents a case study of impact, and identifies the drivers that influence implementation. Finally, it discusses the lifecycle of implementation and outlines a set of recommendations for the next phase of the program based on the findings in this report.

AN OVERVIEW

The FLFRPSP appears to be on track to influence the learning lives of students in remote areas of northern and WA. There is little doubt that the program is having an impact on the literacy levels in participating program schools. This is evident, even where teacher buy-in or leadership support may be wanting. For example, a very small school in WA indicated that, while there were many challenges with implementing the program in their community, conceded that there had been improvement in literacy for students. Another principal admitted to being skeptical at first but has now completely 'bought in' because of the changes seen in students' learning. Likewise, the majority of teachers responding to the survey indicated that student literacy had improved through the program even if they considered that their teaching practice had not improved in the process.

Where available, evidence suggests that schools engaged in the program are demonstrating increases in literacy levels across standardized jurisdictional literacy measures. While there is caution in suggesting that the program in this early phase of implementation is impacting on NAPLAN scores, the current trajectory for future impact is a positive one. This has been further reiterated by a number of schools through the interviews with principals and teacher survey.

> ***"Unfortunately everyone looks at NAPLAN to see how well you are going, but while we have made steady improvement in that area, I think the real jump in that is going to happen in the next year and the year after."*** **(*Principal*)**
>
> ***"…three years is not long enough to gauge whether it is successful. Having worked in remote Aboriginal school for some time it was wonderful to see students' progress as fast as they have done with DI." (teacher)***

While the implementation of the program is not without challenge, it is important to note that clusters of success are being achieved in the face of considerable challenge of context that is beyond the control of the program. In considering the merits and worth of the program, levels of implementation, teacher, and student engagement must be factored in to the valuing process.

Education in remote areas of Australia is both complex and complicated. Student attendance, teacher turnover, leadership turnover, and resourcing are key factors that affect not only the day to day running of schools, but also any initiative being implemented. Almost every teacher and principal who responded to the survey or interview raised the issue of poor student attendance. Most principals also discussed high teacher turnover as a major challenge. The FLFRPSP is subject to all these contextual factors.

Naturally, however, questions arise around program goodness of fit for remote areas and whether the program can deliver impact given the complicated environments within which it is being delivered. After considering the available evidence, it appears that the program is taking hold and establishing a foundation for literacy change. This is despite the fact that both teachers and principals regularly described the program as rigid and inflexible. Nonetheless, its greatest weakness could also be its greatest strength. It is the rigidity and structure of the program that allows students with irregular attendance to re-enter the program, and more

importantly re-enter the learning. In addition, it appears to ameliorate high teacher turnover. In this respect, the nature of the program acts as a protective factor for the system wide issues of student attendance and teacher turnover.

> ***"…the ones that don't turn up every day and pop in and out, they know what's going on, at least there is routine," (Principal)***

> ***"as we get many new graduates in our schools and the turnover rate of teachers is high, I feel it's success is based on the fact that we are all teaching from the same book. This continuity is absolutely essential if we are to have any success with these children learning to read." (Teacher)***

THEORY OF CHANGE

Direct instruction pedagogies have strong evidence for their theory of change. Despite their many opponents, Fisher, Frey, and Hattie (2016) note that with an effect size of O.59, "direct instruction programs offer a pedagogical pathway that provides students with the modelling, scaffolding, and practice that they require when learning new skills" (pg. 53). The life course model of the program is built upon the premise that direct instruction is a viable pedagogical approach to literacy enhancement for students who are struggling to achieve.

The evaluation team undertook initial work with GGSA in mapping the theory of change or logic model for the program (see *Figure 70*). All logic models are considered working documents to facilitate monitoring and evaluation and should be updated to reflect the results of data analysis. Once further literacy data has been made available to analyse, the logic model will be updated to reflect progress towards outcomes. This will be re-presented in the closing report. There are outcomes in the existing logic model that can be attributed to the program and other outcomes that have now become more apparent as outside the program's sphere of influence such as student attendance and behaviour. Subsequent models will reflect this deeper understanding of the program and its context.

Other than the theory of change, time as a contextual factor must also be considered when evaluating the FLFRPSP. Significant and sustained change can be generational, and in an educational sense, generational may mean the passage of an entire cohort through the program levels (6 years). While the initial rate of success may appear to be somewhat faster than expected, widespread changes in in literacy levels that lead to improvements in student achievement as measured by NAPLAN are considered a much longer-term prospect. Despite evidence in the NAPLAN analysis of positive trends over the past 2 years (other than WA government), very few reached statistical significance and cannot and should not be attributed to the program at this point. Sustainably improving literacy achievement in a complex education sector takes time.

Examples of social policy reform suggest that it can take a considerable period of good implementation, program design, research, and evaluation before change is evident in the desired outcome. Sadly, successful interventions may have been excluded when, in reality, poor implementation practices and processes may have explained why intended outcomes were not achieved. A recent review of several studies of educational implementation (Young & Lewis, 2015) indicated education reform and interventions take time and go through many stages; thus, the benchmarks and expectations about the time and extent of implementation are not always clear.

For the program, the expected progression of change is over a 5-year period as illustrated in *Figure 67* below.

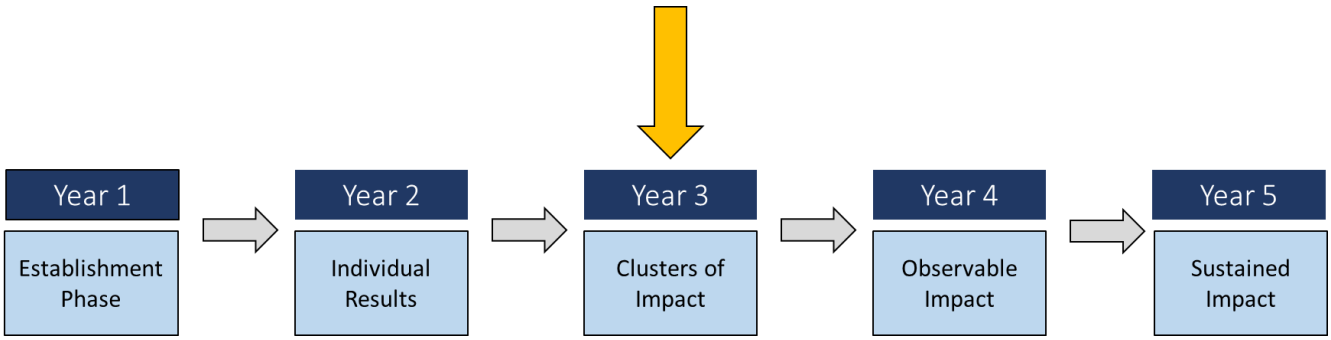


Figure 67. **Progress of Outcomes Over a 5-Year Period.**

While it is not possible to see this theory of change at an overall national level for program schools, there are clear cohorts of students and schools where the predicted causal path is becoming observable. The evaluation has uncovered, through its analysis of the available data, clusters of success in the WA Catholic Education system.

CASE STUDY CLUSTER OF IMPACT – WA CATHOLIC EDUCATION SCHOOLS

The triangulated evidence from WA suggests that change is occurring in literacy. The heat map for WA Catholic Education) schools (see *Figure 64* captures this visually and is reproduced here for reference.

While there are pockets of success in NAPLAN (that some teacher and principals accredit to the program), we suggest that it is not yet appropriate to attribute the relationship between NAPLAN and literacy changes whether positive of negative. Nonetheless, a focus on WA catholic schools does provide a view of the clusters of impact in literacy levels that can be attributed to the program. As demonstrated, the EYLND data shows a significant difference between the control and program schools.

Similarly, effect sizes for change in literacy scores over time at school level were medium to high (see *Figure 68*), and the analysis revealed a .38 effect size difference between program and control. In educational terms, this is equivalent to almost a year of learning (Hattie, 2008).

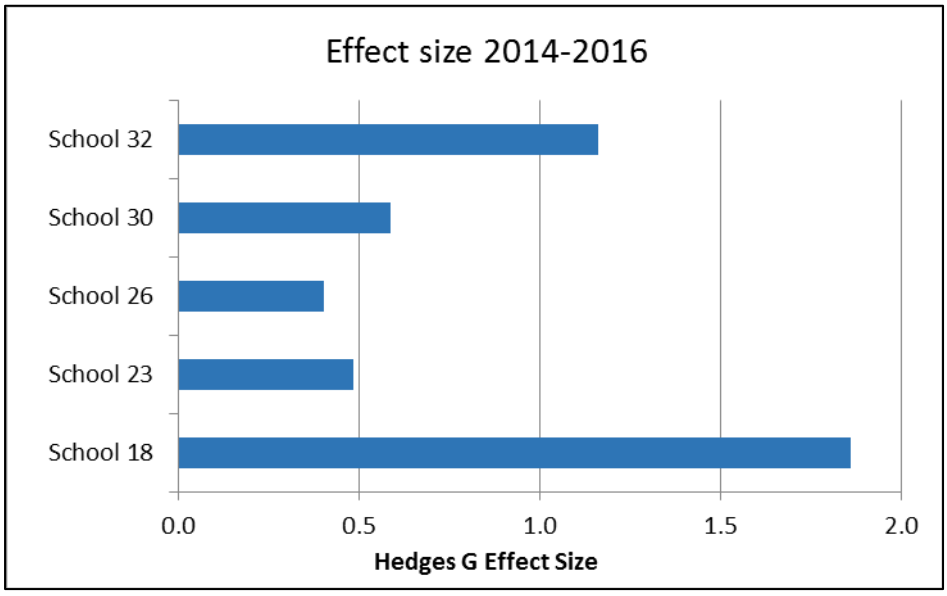


Figure 68. **Effect sizes for EYLND for 2014-16 in Program Schools.**

To better understand the drivers for this progress in literacy for WA Catholic Education schools, relationships between program, contextual, and achievement factors were analysed both statistically and conceptually.

Table 20: ***Correlation Matrix of Program, Contextual and Achievement Factors***

Correlation Matrix					
	Staff student numbers	**Level of need**	**Fidelity**	**NAPLAN Change**	**EYLND**
Staff student numbers	1	-.843*	0.312	0.101	-0.478
Level of need[23]		1	-.011	.186	.326
Fidelity			1	.124	.921*
NAPLAN Change					.188
EYLND					1

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

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

As shown in Table 20, achievement in jurisdictional literacy measures (EYLND) was significantly correlated with program fidelity. This fidelity variable captures both school fidelity and teacher effectiveness. It is a measure of the degree of implementation by school and teachers, and is a percentage; higher scores reflect higher implementation.

Conceptually, WA Catholic education had strong results in both teacher effectiveness and school fidelity, which supports the statistical relationship seen here. The next step then is to understand what it is about this context that promotes fidelity.

WA Catholic Education schools are characterized by teachers with high levels of experience and moderate teacher turnover. Feedback from the staff indicated they felt overall that the program was a good fit for their school; therefore, teacher buy-in was evident. There was a clear relationship between witnessing initial results and teacher commitment to the program.

> ***"It is a very different approach to teaching, and some teachers struggled with that initially, but now that they have been doing it for a while- and they can see the results it's been great." (Principal, Catholic Education FLFRPSP program)***

Furthermore, most teachers agreed that the program supported their knowledge in terms of teaching literacy and strengthened their teaching skills so that there was professional reward for them in teaching the program.

It is also interesting to note that there was a general perception that the program assisted with student engagement; however, low attendance was still seen as a hindrance. As has been described, student attendance in educational activities is a national priority.

Other perceptions of the program from teachers and principals saw a number of factors emerge as potentially important. These were:

- A readiness by the school and senior leadership to engage in change
- Community support and encouragement
- High levels of training
- Fidelity of implementation from program developers, teachers, and school leadership
- Open mind-set to some form of direct instruction

[23] Level of need is a composite score based on a schools ICSEA score, and total number of Indigenous and LBOTE students.

- Engagement and of quality assessment and reporting on that assessment
- GGSA and NIFDI/DataWorks to remain open minded to program adaptations that help to fit context better.

While these factors are all important in their own right (and important to NT and WA Government as well), the potential difference for WA Catholic Education could be the constellation of these factors working together. This aligns with complex systems theory, which suggests that it is not the individual players (or factors) in the system, but the relationship between these players (or factors) that is important (Snyder, 2013). Now that data collection and measurement models are established, future evaluation work will focus on understanding the inter-relationships between factors in the system. This profile of WA Catholic Education schools provides a window into this aspect of the evaluation.

In summary, the focus on WA demonstrates the possibilities of national impact of a DI program, on literacy levels as measured by a standardized assessment. Furthermore, this case analysis has been able to determine that implementation of the program is a critical variable in determining outcomes. Additionally, it is clear that implementation is also affected by a number of variables. A statistical attempt was made to run a path analysis using the data from all schools. Unfortunately, the lack of literacy data from the NT meant the model would not emerge. The following conceptual model (see *Figure 69*) is proposed based on the triangulation of data. Further data collection in the next phase of the evaluation will allow for inference from a causal model to be made. The evaluation team is looking forward to working with all jurisdictions on collecting this data and involving them in better understanding the impact of FLFRPSP.

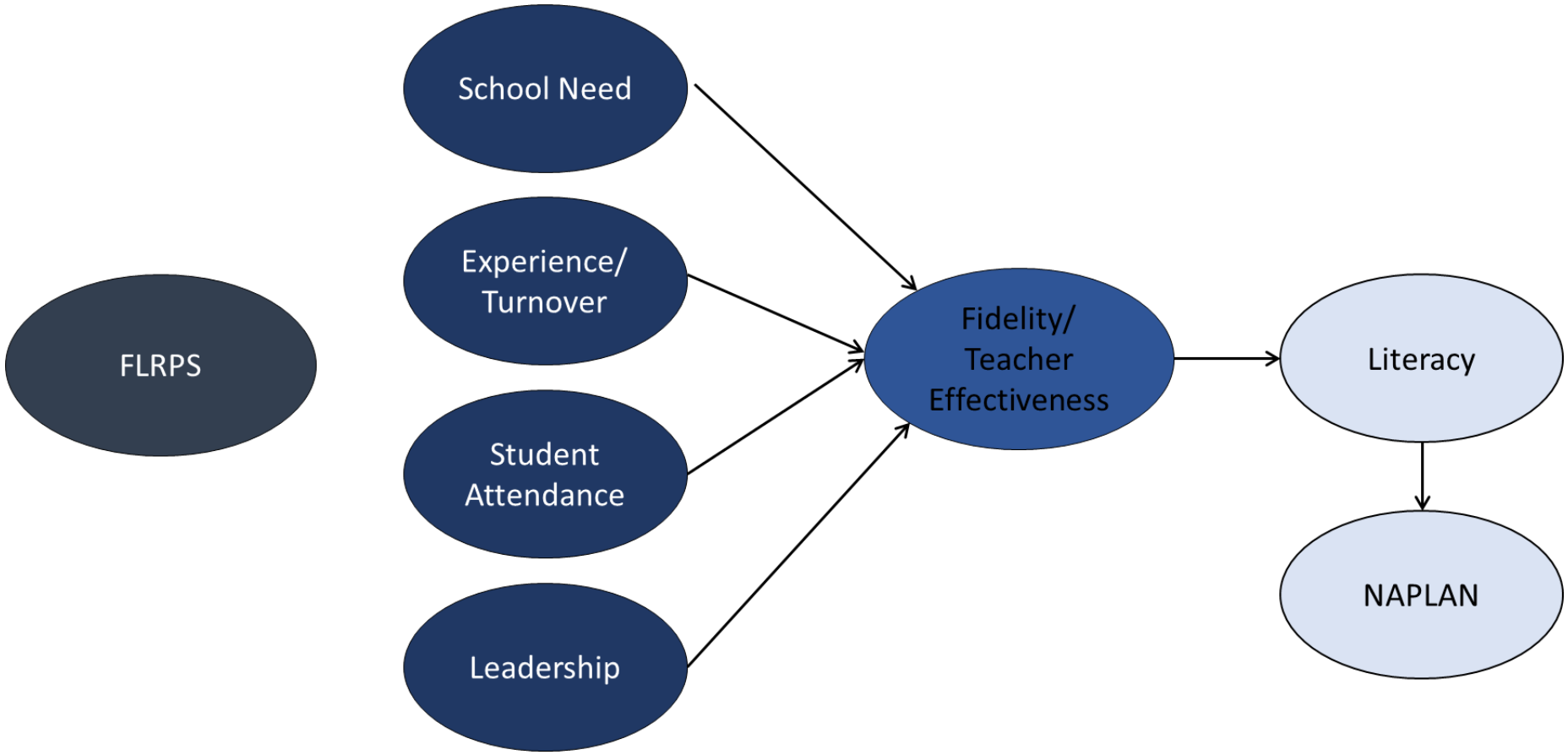


Figure 69. **A potential Path Analysis Designed to Determine a Causal Path Way.**

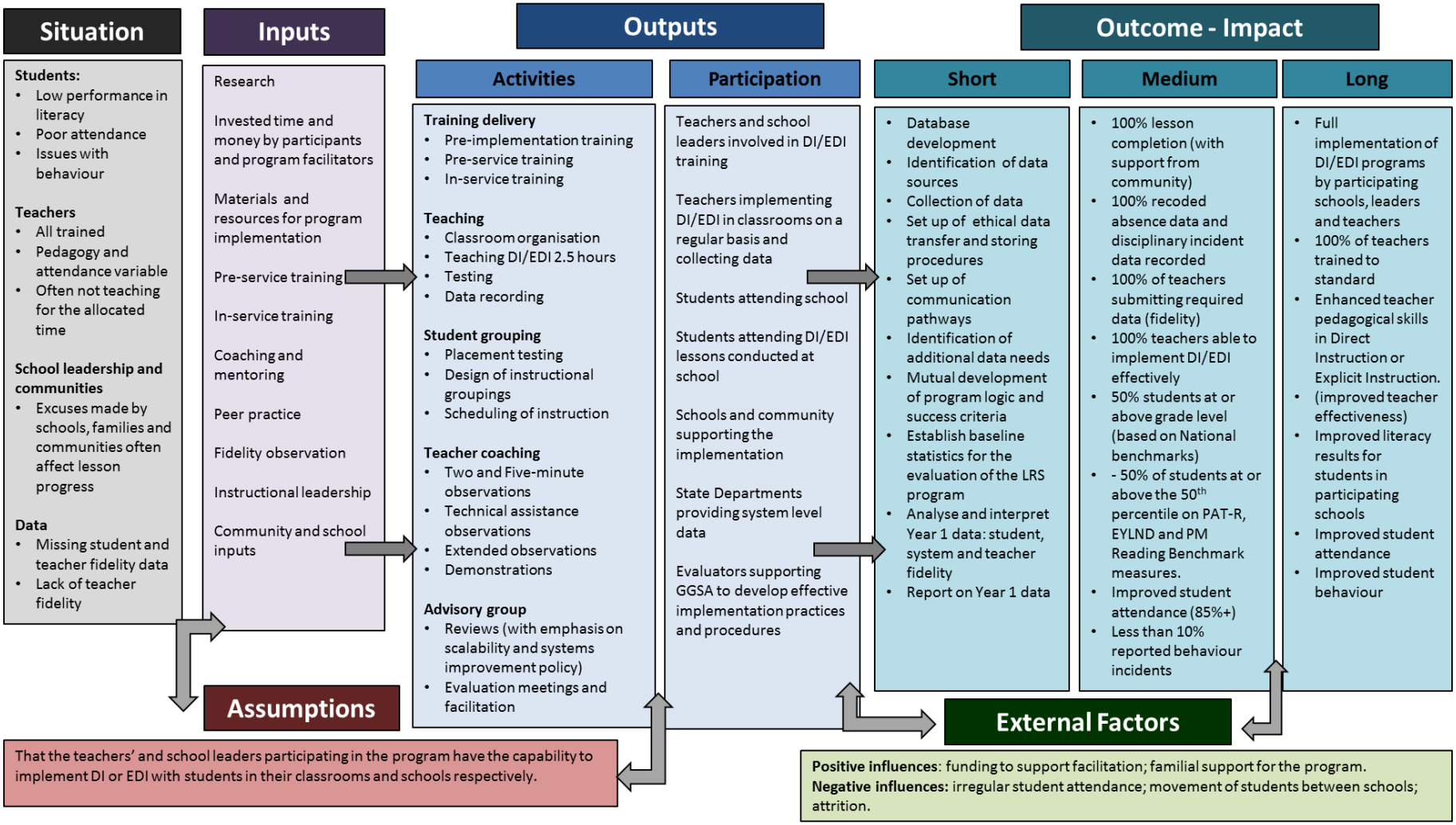


Figure 70. **Original logic Model Developed with GGSA in October 2015.**

IMPLEMENTATION – A CRITICAL FACTOR FOR IMPACT

As the case example of WA Catholic Education program schools demonstrated, implementation is a factor of significance in understanding the impact of the FLFRPSP. Implementation of educational programs is often plagued by complexity due to the diversity of communities and contexts within which education occurs in Australia. However, the influence of poor implementation should not be underestimated or ignored. As Barber, Rodriguez, and Artis (2016) suggested, 'poor implementation is like poison to a reform effort. It demoralizes the front line. It excites the opposition. It frays the coalition supporting it peeling off people who can use it as an excuse to defect while maintaining they still support the reform in theory' (p. 4).

"If you put the resources and time into it – it will work beautifully- if you don't it will fall apart." (Principal)

In analysing the FLFRPSP data, the evaluation team has taken into account that implementation is never linear and is strongly influenced by context. Hence, the expectation that measures will steadily increase over time is ill-founded and naïve. As was seen in some of the teacher effectiveness measures, initial increases were then met with quite significant decreases possibly due to the influence of turnover in teaching staff. Context is paramount when considering the notion of 'implementabilty' in diverse and remote school communities in Australia and for understanding the impact of the FLFRPSP. Another layer of complexity is added when one considers the process of simultaneously implementing multiple reforms within schools, which is often the case.

Honig and Hatch (2004) have noted the need for schools to 'craft coherence' between external demands (i.e., policy mandates regarding improvement) and their own goals and strategies. Honig (2006) describes the challenge as follows:

Those interested in improving the quality of education policy implementation should focus not simply on what's implementable and what works but rather investigate under what conditions if any, various education policies get implemented and work … implementability and success are the product of the interactions between policies, people, and places … the essential implementation question then becomes not simply 'what's implementable and works', but what is implementable and what works for whom, where, when, and why? (p. 2)

This quote highlights the pragmatic focus upon what is referred to as 'implementability', which can be understood as the 'readiness' or capacity to implement. No-one expects the program to work in every context. Equally, no-one expects the program to be implemented in the same way in every context, despite its consistent structure.

There are many examples through discussions with principals and surveying teachers where the program has had to be adapted to fit the context, but with positive results still realised. One prominent factor was the 2.5 hours of timetabled lessons. It was evident that schools are playing with this benchmark to suit their own context.

> ***"I only teach DI for 90 minutes every day this year. For the first two years, it was for the full two and a half hours. Now I have time to focus on important areas of Literacy."***

In any implementation, a tension will exist between fit to context and intended program structure. Successfully navigating this tension is likely to result in positive outcomes for all stakeholders. What can be conceded to context without losing program character and fidelity? A finding from this phase of the

evaluation is for GGSA to work with the schools and the program developers to understand how context can be accommodated.

DRIVERS FOR IMPLEMENTATION

As the depth and nature of implementation varies across groups, a diverse range of influences are uncovered which help to explain this variation. Variation in implementation and impact of programs is a significant issue in education but not unexpected given the complex and diverse nature of the sector. Utilising all the data collected, we can conceptually estimate the levels of variation in several identified drivers for FLFRPSP.

These drivers represent areas that relate to the program such as the degree and quality of implementation; teaching factors, such as teacher engagement, effectiveness, turnover, and teaching assistants; student factors, such as student attendance and behaviour; organisational systems, structures and environments, leadership, and adequate resourcing. An individual driver can have both positive and negative influence depending on the context.

***Figure 71*. Program Implementation Drivers.**

Figure 71 illustrates the positive and negative valence of influence of each program driver.

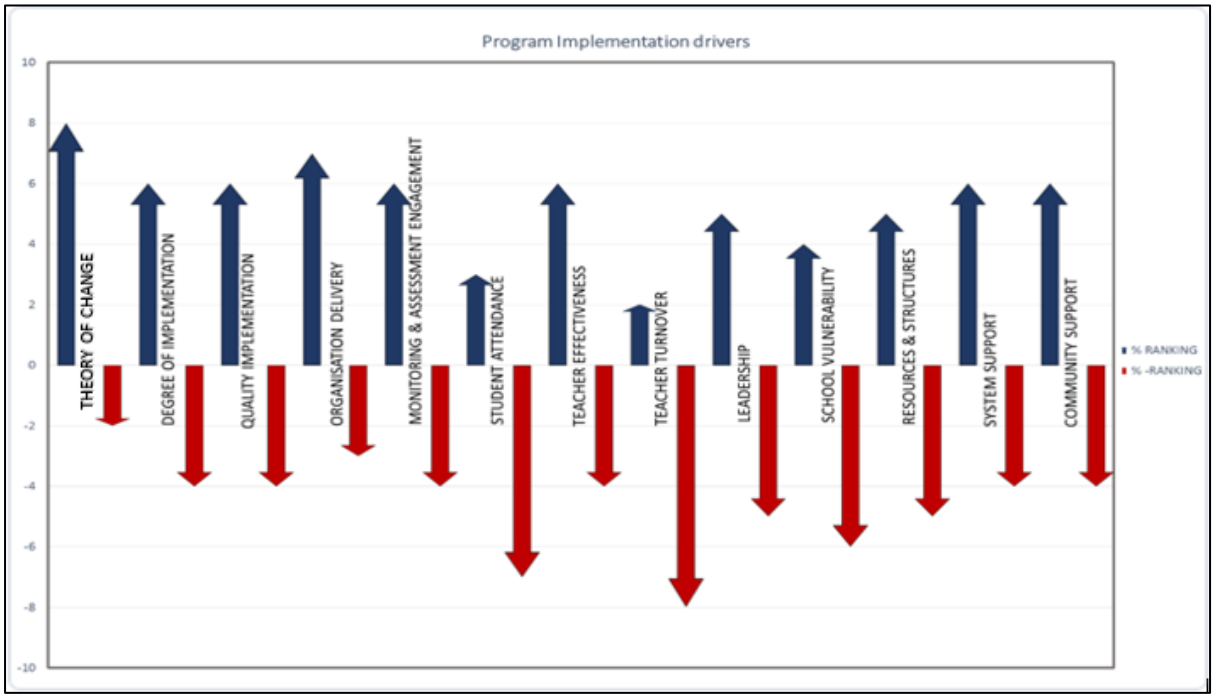


***Figure 71*. Program Implementation Drivers.**

CORRELATION WITH DRIVERS AND NAPLAN SCORES

Before discussing each driver in more detail, the evaluation conducted a correlation analysis with measures underlying these drivers and NAPLAN scores to understand whether and to what degree any of these variables related to one another. Table 21 displays these correlations and indicates those that are significant. As demonstrated, factors such as ICSEA, school size, number of TAs and teachers, and including LBOTE correlate with changes scores on assessments, and in some cases, the fidelity of the program.

Table 21: *Correlation Matrix of Myschool, NAPLAN and GGSA Program Data*

	Number of Teachers	Number of Non-teaching staff	School ICSEA Values	Number of Students	Percentage of Indigenous Students	MS: Percentage of LBOTE	Total Change - Composite Score	Reading Change	Spelling Change	GP Change	Writing Change	Average School Fidelity	Average Observation Per Teacher	Average Training Per Teacher	PD: Average Teacher Effectiveness	Teacher Turnover
MS: Number of Teachers	1.00	**.923****	**.436***	**.978****	**-.439***	**-.378***	0.02	0.20	-0.12	0.00	-0.27	-0.08	0.04	**-.489***	0.23	0.24
MS: No. of Non-teaching staff		1.00	**.529****	**.914****	**-.514****	**-.467****	0.11	0.32	-0.10	0.03	-0.10	-0.02	**-.465***	**-.454***	0.23	-0.02
MS: School ICSEA Values			1.00	**.532****	**-.821****	**-.662****	0.32	**.489***	0.14	0.13	0.09	-0.03	**-.585****	**-.482***	0.33	0.13
MS: Number of Students				1.00	**-.518****	**-.424***	0.11	0.27	-0.08	0.06	-0.22	-0.11	0.03	-0.40	0.10	**.402***
MS: Percentage of Indigenous Students					1.00	**.739****	-0.22	-0.30	-0.09	0.00	-0.12	-0.07	0.41	**.560****	-0.43	-0.12
MS: Percentage of LBOTE						1.00	0.06	-0.17	0.05	0.22	0.19	-0.02	0.36	**.591****	-0.26	-0.23
N: Total Change - Composite Score							1.00	**.781****	**.675****	**.820****	**.615****	-0.15	-0.32	0.02	0.08	0.07
N: Reading Change								1.00	**.391***	**.588****	0.33	-0.05	-0.40	-0.23	0.19	0.14
N: Spelling Change									1.00	**.403***	**.525****	-0.07	-0.34	-0.03	0.17	-0.37
N: GP Change										1.00	0.37	-0.39	-0.23	0.19	-0.07	0.09
N: Writing Change											1.00	0.19	-0.26	0.15	0.34	**-.603****
PD: Average School Fidelity												1.00	0.09	-0.09	**.485***	-0.22
PD: Average Observation Per Teacher													1.00	0.10	-0.33	0.44
PD: Average Training Per Teacher														1.00	-0.41	-0.17
PD: Average Teacher Effectiveness															1.00	-0.46
Teacher Turnover																1.00

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

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

FACTOR STRUCTURE

Using the correlations from Table 21, the underlying structure of the variables was explored through a factor analysis. Three main factors emerged. The first related to the number of teachers and students in smaller versus larger schools. The second relates to the change increases across the 4 NAPLAN tests. The third is a measure of Indigenous and LBOTE students. The important information is how the measure of School Fidelity of the GGSA program was related – The program had higher fidelity in schools with fewer Indigenous and LBOTE students; but was unrelated to the NAPLAN changes.

Table 22: *Variable Factor Structure*

	Factor		
	1	2	3
No teachers	**.938**	-.181	.191
No Non teachers	**.932**	-.049	.086
ICSEA	**.666**	.343	-.388
No students	**.959**	-.091	.149
No Indigenous	**-.636**	-.252	**.537**
Per LBOTE	**-.595**	.026	**.556**
Reading Change	.326	**.728**	.120
Spelling Change	-.128	**.766**	.015
GP Change	.069	**.717**	**.542**
Writing Change	-.276	**.763**	-.178
School Fidelity	-.145	-.073	**-.598**

Now that an understanding of relationships between these drivers and program impact has been explored and illuminated, each driver as they relate to different aspects of context will be discussed according to schools and community, GGSA, and Program Developers and evaluation.

SCHOOLS AND COMMUNITY

There are a number of elements from the school and community that influence implementation. These will be discussed under the following headings: Leadership, teachers, teaching assistants, students, and community.

LEADERSHIP

Barber et al. (2016) suggest that the presence or absence of the capacity to deliver supportive leadership is critical in the implementation of any educational intervention. Similarly, many researchers confirm that school leadership has a significant influence on student outcomes, second only to classroom teaching (Harris et al., 2013; Louis et al., 2010). Thus leadership needs to be seen as a critical determinant of outcomes and successful implementation of programs (Mourshed, Chijioke, & Barber, 2010).

Three key elements are noted:

- Support
- Structure
- Championing

Where school staff saw the implementation of the FLFRPSP as positive, this was often attributed to perceptions of good leadership and leadership support. Given problematic turnover rates in some pockets of

the program schools, attention to this aspect is critical. In addition, where leadership is not necessarily bought in to the program, understanding the reasons behind this and working closely with leadership on the reasons where possible may help further implementation in terms of dosage and quality.

> ***"The success at our school has been supported by the strong leadership and commitment to the program shown by school administration. It is such a clear expectation that this program will be implemented with fidelity and accuracy that all staff are on board and the program is benefited by this." (teacher)***

> ***"Great support from GGSA and in-school coach. Strong support from leadership; excellent training conducted from Dataworks." (teacher)***

TEACHERS

While the data from the EDI and DI programs indicates that teachers display a level of effectiveness, quality of implementation in classrooms remains unclear given the high level of turnover in teaching staff. Consequently, engagement of teachers in training to deliver the EDI and the DI program is severely compromised by teacher attrition. It is not clear if the program can deliver timely and sufficient professional development to ensure that students in the program are receiving consistent instruction from teachers who are proficient in delivering the program. In summary, the program is being delivered by a perpetually learning workforce of teachers.

> ***"GGSA give us all the support they possibly can, but they are not able to give us extra teachers, they are strict with our teachers- all of that is fine, but they cannot help with our teacher numbers."***

Kurt Engelmann (2017) suggests a teacher can master the process of direct instruction within a five-year period. Teachers within this program appear to have tenure of less than eight months. *Figure 72* illustrates the point that the probability of students having a teacher who has advanced training and experience in DI/EDI is negligible.

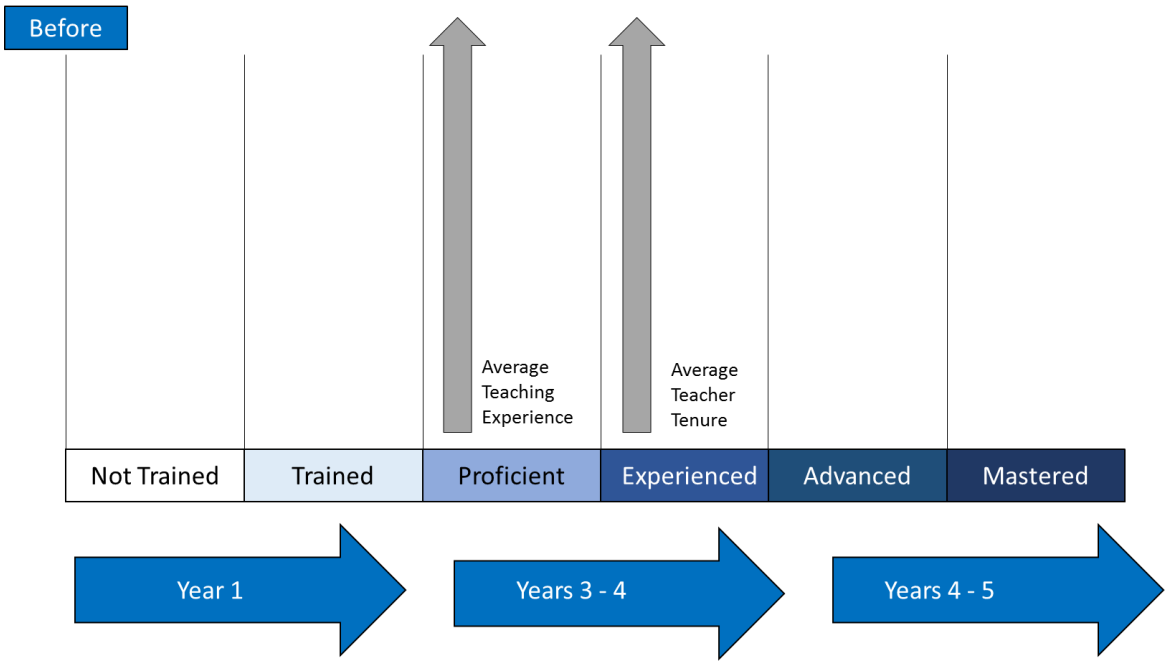


Figure 72. **Teacher Development Trajectory for DI with Average Teacher Tenure and Experience.**

The impact on the fidelity of the program as a consequence of teacher attrition and inconsistent experience levels with DI or EDI is immense and it is beyond the control the program. It is interesting to note that teachers in WA Catholics schools appear to have a longer tenure; this may account for some of the jurisdictions early success with the DI /EDI program.

TEACHING ASSISTANTS

Work by Blatchford, Webster, and Russell (2012) demonstrates that the impact of support from teacher aides in the classroom is negligible on achievement as measured by standardised tests. They found that pupils receiving the most support from teaching assistants made less progress than similar pupils who received little or no support from teaching assistants, even after controlling for pupil factors related to attainment. The report suggests that this can be attributed to poor deployment, training, and practices, and the fact that often TAs take the child out of the learning (Blatchford et al., 2012). This is potentially true of the FLFRPSP.

While it can be argued that TAs are critical to the support of the FLFRPSP (and there are some reported examples of positive support for the use of TAs in delivery of the program), the evaluation highlights that many schools not only struggle to employ TAs, but that retaining them was also a challenge. Unsurprisingly, this negatively impacts the teachers' capacity to fully implement the program. In addition, TAs often struggle with supporting implementation of the program because of their low levels of literacy and English language knowledge. The interviews suggest that the TAs found the implementation of the program difficult. Conversely, the program was actually good for the development of the language levels of the TAs.

> ***"Lack of regular Assistant Teachers who can teach a group in the Language area. In remote areas it would be great to employ literate tutors to support our TA's just for the DI time." (teacher)***

STUDENTS

There are a number of drivers related to students that influence the program implementation and outcomes. Attendance by students is a critical issue for any educational venture. Similarly, the itinerant nature many of

the students educational experience will have a dramatic influence on continuous education. The nature of DI and EDI does allow students to pick up from the place where they discontinued their literacy learning. While the evaluation did not have access to quantitative student behavioural data, perceptions for teachers and principals indicated that behaviour management influenced the capacity of teachers to run the program effectively.

'Literacy and Numeracy (NAPLAN) is an annual literacy and numeracy test for all Australian students, and results from the test are disaggregated into a number of categories including LBOTE. For this and other categories, results on each section of the test are aggregated into state, territory and national means and standard deviations enabling comparison of performance. The NAPLAN data indicate that since the test began, in 2008, at a national level there is little difference between the results of LBOTE and non-LBOTE students on all domains of the test. This is a national result, and there is greater variation at state and territory level. However, these results defy a logic which might suggest that the LBOTE category will reflect the influence of English as a second language on test performance, rather suggesting that a second language background is not associated with test performance'.

COMMUNITY ENGAGEMENT

The school community regularly argued that buy in and support from community is essential to successful implementation. For example, the community perception is that 2.5 hrs of literacy training is too long for students to be engaged in. However, where results have been demonstrated, parent and community buy-in has increased.

> ***"Parents and the community can see that there is a light at the end of the tunnel now, and because they are learning other stuff much faster now, it's expanding their world view." (Principal)***
>
> ***"That was more initially, the parents now are all pretty much on board. They can see how much their kids are learning." (Principal)***

GGSA AND PROGRAM DEVELOPERS

While the support and responsiveness of GGSA and NIFDI/DataWorks was not brought to question, it was often felt that the program (DI or EDI) could take a more targeted approach to supporting schools in complex circumstances. Schools across the program were incredibly variable and it was noted that a one-size-fits-all was not conducive to full fidelity and high levels of dosage of the program. In summary:

- access to coaches was seen as a critical aspect of the program
- flexibility in terms of delivery, benchmarks and resources was required
- increased training and coaching
- capacity building for teachers on assessment and feedback

> ***"Manager who is based in America, who is coming from a very different context, who offers remedies or suggestions without really knowing the context like the teachers do- I think they find that very difficult." (Principal)***
>
> ***"The training was excellent, the initial training was run by Hollingsworth and Ybarra and other consultants from data works US- they came out and facilitated the development and it was first class" (Principal)***
>
> ***"Support from GGSA, DATAWORKS, School Principal, teacher coach & student attendance." (Teacher)***

EVALUATION

The evaluation from its early involvement with the FLFRPSP adopted the perspective that engagement of the program in the evaluation was indispensable to understanding the aims of the program and critical to adding value to the implementation process. Providing rigorous, accessible, and useful evaluative information about implementation is a necessary and ethical component of the implementation of any intervention.

Evaluation commenced with a view to increase evaluation capacity of those delivering the FLFRPSP. There is little doubt that from 2015 to 2016/17, the quality of data gathered and monitored by GGSA increased, and the amount of missing data decreased. Teachers, coaches and providers of the program increased their engagement and their willingness and readiness to engage in the evaluation processes considerably over the period of 18 months. It is important to note that

- databases were built with all program data uploaded
- attention was paid to analysis of the program data
- feedback was used to adapt the program
- there is a tangible increase in evaluation thinking

THE LIFECYCLE OF IMPLEMENTATION

It is also important to consider implementation from the perspective of time as well as complexity and context. The implementation literature (Berman, 1980) suggests three phases of implementation — mobilisation, implementation, and institutionalisation. It is fair to say that the FLRPSP is amidst the mobilisation and solid implementation phase. The life course model of implementation suggests that programs often move through phases of implementation that lead to a stage where full impact can be felt. In the early phases programs suffer from teething problems, with several iterations of program implementation ironing out these problems. Once program problems are resolved the full impact of the program should be felt and outcomes have a chance of being met.

There was substantial evidence that initial implementation was difficult for schools.

> ***"The implementation process was a big change, there was a little bit of upheaval and uncertainty." (Principal)***
>
> ***"The beginning was really tough, very different way of working." (Principal)***
>
> **"The initial change of pedagogy and mindset." (teacher - in response to difficulties during implementation)**

Figure 73 demonstrates the life course of the FLRPSP by plotting the predicted path of implementation as well as the current and predicted paths for literacy change across the various States and Territories and suggests when the program may release its full potential for outcomes (Year 5). Once implementation reaches higher levels of dosage, impact within the classroom for teachers and students should be visible. This of course assumes that time will be taken to achieve the desired result.

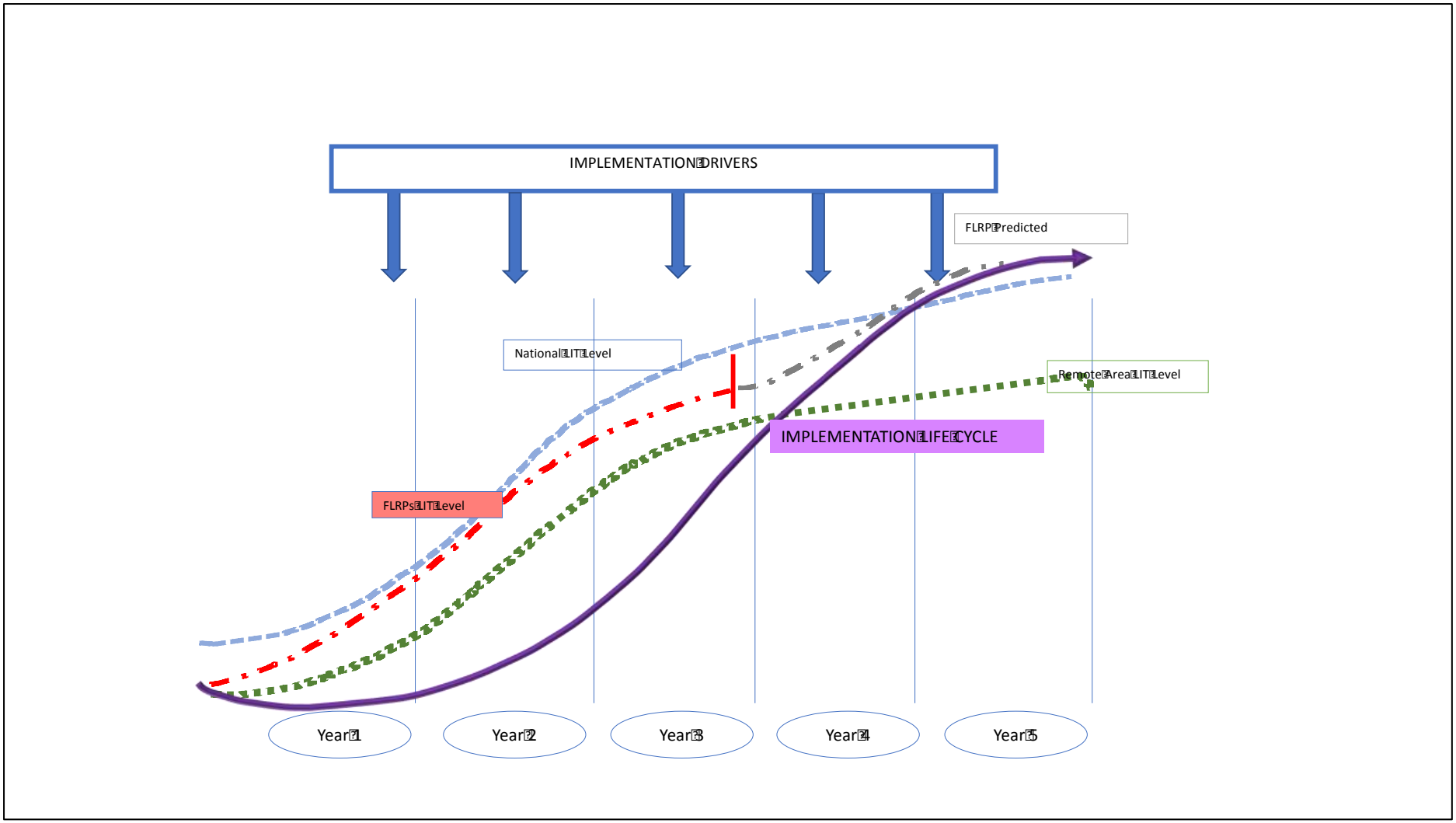


Figure 73. **Lifecycle of Implementation for FLFRPSP.**

We suggest that implementation is not yet at the peak of the curve and won't be until year 5. At that point, it is expected that broader educational outcomes will be evident, attributable to the program and sustainable. Year 5 is significant too because it is at this point that those students that have had continual exposure to the FLRPSP should demonstrate change in their NAPLAN scores.

Currently the program is positioned at emerging change, which suggests that there are pockets of success with scale of change yet to occur. For example, schools within NT and WA Government are showing promising results but overall for those jurisdictions more time is required to embed the program.

> ***"Please do not let the power of a few teachers allow this program to be stopped. Keep it going for 5 years and let the sceptics see what can be achieved." (Teacher)***
>
> ***"We are only beginning to see progress." (Principal)***

Continual and sustained exposure to the intervention is the critical variable; in particular, it is noted that currently individual students have been exposed to varying degree of implementation as well as differing levels of quality. Once this point of peak implementation occurs, impact that is attributable to implementation will be measurable and further conclusions can be drawn about the longitudinal impact of the program. The theory of change suggests that with appropriate levels of fidelity and support over the coming years, implementation will increase in depth, reach and quantity.

FINAL WORD

The implementation of FLFRPSP is having an impact on the students, teachers, and schools within the program with, not surprisingly, higher impact – in terms of intensity, complexity, and breadth – associated with more developed implementation processes and practices. The evaluation demonstrates that once training and implementation is achieved, and there is engagement by teacher and students, positive change in literacy learning can occur. Assuming that the current trajectory continues, literacy levels as a result of the implementation of the FLFRPSP are expected to increase over time and lead to improvements in teacher quality, greater consistency in the provision of the program, and better student learning and outcomes. The foundations for this change are firmly in place, evidenced by the increasing rate and quality of uptake of the program across the schools.

While the program is demonstrating clusters of success, one of the goals here in making an evaluative judgement is to predict the probability of successfully scaling the project. As previously suggested, an evaluation heuristic exists between the program, context, level of input, and process of the program, and this allows us to explore probability of ongoing success of the initiative. Identifying the programs active ingredients is crucial for unpacking the interventions 'black box'. More explicitly, we in this evaluation needed to determine the outcomes of the program that relate to the overall theory of change as described by the program logic. The argument proffered is that some form of explicit instruction (direct instruction or explicit direct instruction) will positively impact on literacy outcomes for children. Additionally, the question arises: is there a good fit between the assumptions behind direct instruction and the challenging contexts in remote areas?

This overall question should not focus on the opinions around pedagogical preference but more importantly whether there is a goodness of fit between the problem and the solution along with understanding any mediating factors. While the arguments about the success of direct instruction are long standing and there are opponents that will always be against the approach and similarly, some groups, communities, and teachers may simply not be enamoured with the approach. The evidence presented here, as well as an extensive literature gathered around direct instruction, demonstrates that this pedagogy can and should have a positive impact on children's literacy outcomes.

Equivalency of results or magnitude of impact also needs to be considered, the program produced positive results in some areas despite challenging implementation circumstances. Hence, it begs the question what could be achieved in more stable circumstances. Cohen and Ball provide a useful analogy:

"…… from this angle, implementation in ten percent of schools in New York City would seem small. But if innovation designs and schools have been as weak, and environments as unruly, as we claim, effective implementation of even a modestly complex intervention in ten percent of New York City schools would be an enormous achievement. Scale is relative not just to the universe of possible implementers, but to the scope and depth of what must be done to devise and sustain change." (pg 35)

This quote poses a useful perspective when considering the clusters of positive impact, we see as a consequence of the direct instruction program.

Are these results generalizable? This is a question that also needs to be taken into account in the evaluation, and, as discussed, it is too early to tell. Greater access to literacy data would have assisted in answering this question. The nature of influence by the program drivers must also be accounted for: what circumstances support the implementation of the program? The evaluation is partway to considering this question. The next phase of the evaluation is designed to facilitate explanations, close the data gaps, and refine the measurement model for future evaluation.

While the next level of evaluation will explore these questions further and in more depth, at this point there seems little doubt that the program is not only reproducible but also applicable to a high need, itinerant, low-attending population.

As described, the theory of change is solid. There appears to be a good fit with the context; this is not to say some form of direct instruction should be the only pedagogy students are exposed to. More importantly, it is using direct instruction at the appropriate time in the appropriate circumstance. Hattie and Donoghue (2016) demonstrated that certain learning strategies are ineffectual in certain conditions yet have enormous effects when other conditions are in place. It is more important to note that the evidence may suggest that this is not a one-size-fits-all approach for a one-size-fits-all context; the context varies considerably, thus so does need. While the context is significant, it is actually the relationship between the implementation fidelity and context that is the critical variable; context alone cannot explain varying results. Research on implementation and scaling up is relevant here, as it provides a lens to explore implementation and fidelity within context.

Coburn (2003) identified four dimensions of scaling up: depth, sustainability, spread, and a shift in reform ownership. Most discussions of "scaling up" treat it as a quantitative problem, a matter of broad adoption and implementation. Our analysis suggests instead that scale is as much a qualitative as a quantitative problem.

> *'Scaling up is not only about increasing the number of users implementing a program, but also about the depth of the reform, sustainability of practices after initial implementation, and strong ownership of the intervention at the district and school levels.' (Coburn, 2003)*

There is much research on implementation (Barber et al., 2016; Honig, 2006; Spillane, Reiser, & Reimer, 2002) that demonstrates the importance of considering implementation and fidelity. In the background of evaluation, we must consider context, input, and process of this pilot in relation to change so as to inform policy. There are many models that relate here: an evidence based implementation model (Meyers, Durlak, & Wandersman, 2012), Getting To Outcomes (Chinman, Imm, & Wandersman, 2004), which provides a structure for development, implementation and evaluation; the recently released 'Scaling up Evidence Based Practices: Strategies from Investing in Innovation(i3)' (i3, 2017) suggests that scaling for success is dependent on 4 factors: use multiple methods to establish buy-in; building a regional and national infrastructure; adapting practice based on evidence; and plan for sustainability from day one. Levin (2013) also warns that 'The goal -

for scale - should be to avoid either excessive enthusiasm or excessive scepticism, but to maintain a reasoned approach that over time yields collective learning'.

While there is much advice about scaling-up, there is also considerable discussion about the hinderers to successful scaling. Interestingly, Engelmann and Engelmann (2004) wrote extensively about impediments to scaling up school reforms, and delivered a conundrum for education suggesting those models of reform that have numerous requirements are the hardest to implement, however they are more than likely the ones to have the greatest impact on student achievement.

The RAND manuscript (Chinman, et al., 2004) provides a very detail perspective on implementation and scale that goes to the heart of education. It offers two high level lessons:

- No matter the target of reform or the design construct, the scale-up process is necessarily iterative and complex and requires the support of multiple actors. This is likely to remain so for the foreseeable future. (pg. 647)

- If scale-up is to succeed, the actors involved—including developers, district officials, school leaders, and teachers—must jointly address a set of known, interconnected tasks, especially aligning policies and infrastructure in coherent ways to sustain practice.

This model of scaling illustrates the complexity and complicated nature of implementing for change in an education setting. The Mutual Adaptation Model of Scaling is designed for an American context but it assumes the context or the world view of the community considered. This model takes us to core of action of educational change –the classroom.

Figure 74. Mutual Adaption Model of Scale-Up presents a detailed "model that emphasizes the complexity of the organizational environment in which interactions between students and teachers take place and the need for mutual adjustment—of policies and practices—among the key actors in this Environment".

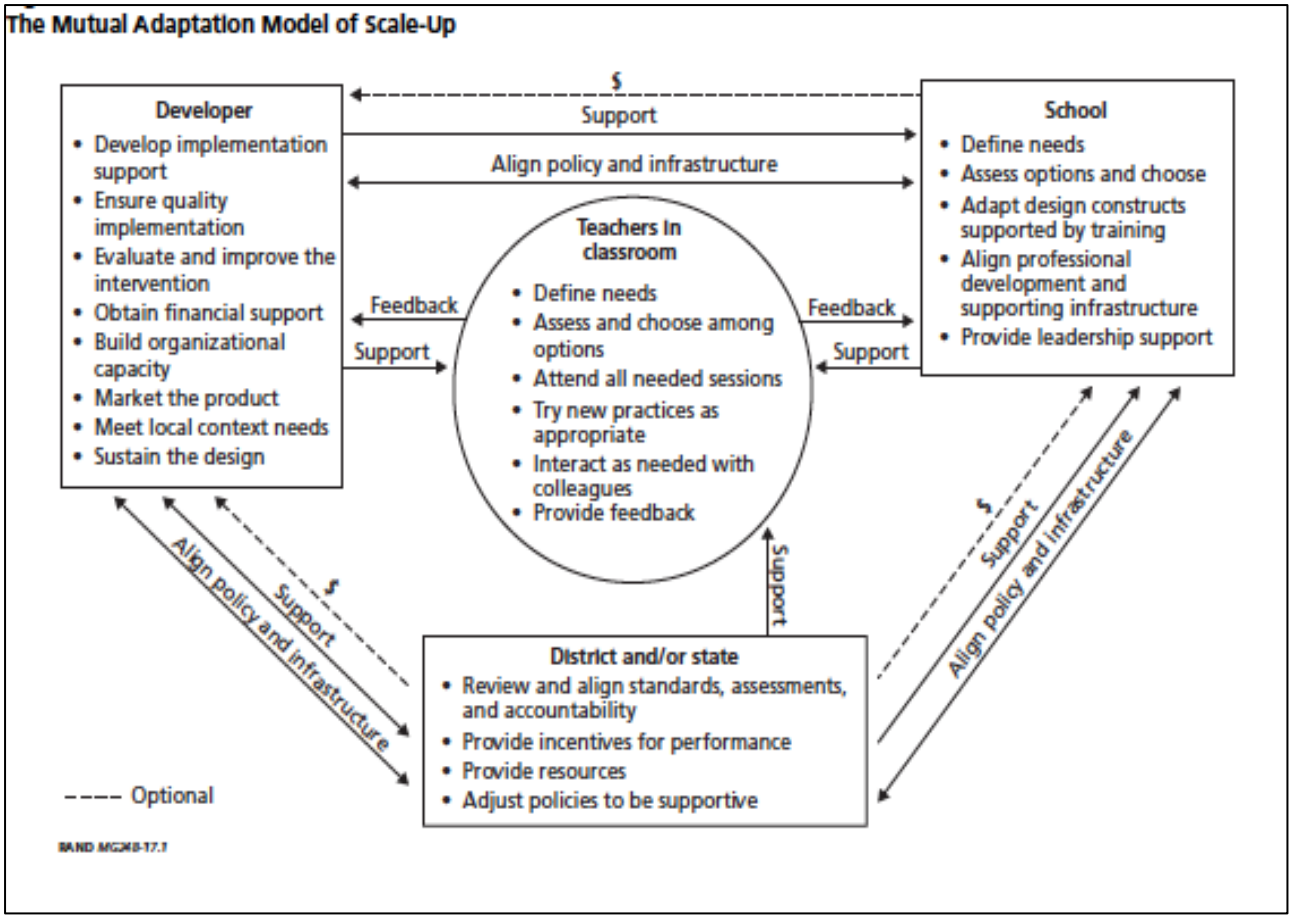


Figure 74. **Mutual Adaption Model of Scale-Up.**

This model of scaling-up and implementation articulates the drivers for implementation as well as the key relationship between the inputs, process, and program, hence providing a perspective on relationships for the FLFRPS program. Most importantly, it demonstrates the need for collective efficacy to achieve full implementation or scale of the program. It illustrates the complexity of the players and the concentration of agency they possess. The feedback loops between the classroom and the program providers (developer) and the school and system (district) are seen as essential. The authors suggest that elements within each area are not exhaustive lists; however, they are enough to grasp the key structure for each component. This model not only drills down into the relationships between the context, input process, and outcomes, but also demonstrates the collective nature of reform.

This model also presents a path for the next phase of the evaluation. The key point in this phase of evaluation is that we have been able to identify particular relationships that are significant. Going forward, we need to comprehend the nature of these relationships both the successes and failures. While as described the context for the program cannot be ignored, it does not totally define the success or failure of the program. Subsequently, it will be essential to understand the relationship between implementation fidelity and the context. Once this is established scale and implementation of the program can be fully explained.

There is little doubt that remote schools in Australia are at a disadvantage. Generally, remote area experience higher attendance challenges, behavioural issues, proportion of need, teacher attrition, principal turnover and resources need. This type of environment is education's 'wicked problem'. A wicked problem has innumerable causes, is tough to describe, and doesn't have a right answer, recently wicked problems were described as unsolvable but tameable (Camillus, 2008). As described there are a number of key variables that come into

play, teachers, students, schools and community, and systems have to bring different initiatives and elements to bear. In this case, we note that community disengagement with education is high, teacher turnover is high, classroom support is variable, principle turnover is high, and most importantly student engagement and attendance is significantly lower in remote regions. Attempting to implement in such circumstances raises considerable challenge, although education continues and programs are often successful despite the circumstances.

While the program may assist in taming the wicked problem, the causal chain connecting such an outcome is so elongated it may only be visible over a generation and considerable amount of data collection. Robinson (2001) also suggested that one of the greatest impediments for educational reform is that there is not always a direct relationship between the problem and the solution designed to elicit reform. Hattie (2008), however, suggests that student engagement relating to factors such as student attendance can be influenced initiatives that make the school more inviting; this of course includes quality teaching. In an evaluative sense, we know the literacy program has not as yet had an impact on student attendance levels, nor should it be an expectation of the program at this point; similarly, influence on teacher turnover or student behaviour are not considered outcomes of the program. It is not feasible at this stage to statistically assess the contribution of the FLFRPS program to these long-standing issues. However, the current evaluation has established a system of data collection that would support the analysis of this hypothesis in the long term.

The various elements of the model, e.g., the teachers and their influence, are explored in this evaluation and it highlights that specifically teachers, schools, and program delivery are key mediating factors. As the program logic suggests, it is the interplay between each element that will influence implementation. The current high level of turnover of teachers is one of the most significant impacts on the quality of the implementation of the program. In essence as a consequence of the constant change teachers may not always have the buy-in they require, an understanding of the need, the practices entailed in the program may be foreign and of course they may feel like they have been thrown in the deep end. We noted that the schedule of training can often be out of kilter with the fluidity of this teaching staff. The evaluation demonstrates that generally, TAs are not successful in implementing the program. Understanding the barriers for the successful use of TAs is a question for the second phase of the evaluation. However, the current findings relate to the work of Blatchford et al. (2012) that demonstrates TAs have a negligible and sometimes a negative impact on achievement. The explanation is a compelling one: Blatchford suggests TAs are provided with little training and are utilised incorrectly in the classroom, in this case literacy and English as a second language may also be of concern. This is a school teacher and community issue.

Leadership is also critical, every educational reform needs a champion, the preliminary surveys and interviews from the evaluation support this premise. Leadership turnover is high; hence, addressing the continuity of implementation is a significant concern for this program.

System factors, such as resourcing, policies, and assessment and monitoring are also significant to implementation of the program. The capacity to monitor and evaluate and classroom, school, system and program level cannot be ignored

Evidence for implementation and scale are essential to ongoing adaptation and improvement of any intervention or innovation. In this case there is clear evidence that the collection of evaluative data has improved considerably and the provider organisation GGSA have made considerable efforts to strengthen the collection and flow of information, as well as devised an infrastructure for storage and reporting on information. The increase in evaluation capacity is evident. It should also be noted that schools within the program have also committed to increased collection and collation of assessment materials related to the program. There is, of course, some way to go. This program information and assessment information collected at a school in system level will ensure that causal modelling will be feasible in the coming years. At a system

and school level there are highlighted concerned in this evaluation with some states and territories, and jurisdiction inability to gather, link, and share data that could determine the progress of the program. This must be rectified.

In these highlighted mediating factors, we have the crux of an explanation for the variable degree of implementation fidelity and quality and within-school differences. These factors will be explored further in the next phase of the evaluation

Much of the literature on implementation and scale of programs suggest that program sustainability is a key element for scaling up. The recent work released by i3 (2017) suggests that planning for sustainability from day one is key if a program is to ensure ongoing success and adoption through continued engagement and buy in. The complexity of the environment suggests that greater flexibility may be more appropriate for the program; the current benchmarks for implementation and progress are hindered by the environment. The program needs to align with the current environment and circumstances while at the same time pushing for progress. For example, activities in the classroom are key to implementation. Teachers suggest that it is taking longer to get through the lessons, so the question here is, in what ways can a continuum of implementation that is seen as progressive and appropriate in pace be developed? Training is another area where the program may be able to adjust. Any such adaptions must be completed in collaboration with teachers, schools, and the system. Ball and Cohen (2006) suggest that '*to solve the problem of "scaling up" requires "scaling in" – by this we mean developing the designs and infrastructure needed to support effective use of an innovation. That, in turn, requires consideration of the problems that have made some sorts of innovation difficult, and taking these into account in deciding what to change, and how to design the means to do so. It also requires significant attention to designing the use of innovations by practitioners, in the environments in which they work (p 35)*'. It is always essential to consider multiple ways to engage and get buy-in from schools, communities, students, and teachers in the implementation of any innovation or initiative, and the FLFRPSP is no different.

The evaluation points to the idea that change in literacy proficiency for students in remote areas relies on a collective efficacy, the student, teacher, senior leaders, school community and of course the system are responsible for some level of current contribution to the success of the program. There is little doubt that collective impact is critical to the sustainable impact of educational initiatives.

There is little doubt that the program has already had a positive impact. To consider the sustainability and scale of the program, a collective approach is necessary. Systems, communities, and schools can disrupt the current challenges that may be restraining the program or causing variable results. The program and its communities may need to be more adaptive to the environment and the agile in relation to the implementation drivers. The next phase of the evaluation will explore these issues.

The theory of change for this program is solid. The need for the program is clear, while the will to implement is variable. The program has already accrued clusters of successful change in a complex environment. Undoubtedly, the program is disrupting falling literacy results in schools and is on track to generate further change.

WHAT'S NEXT?

Given the implementation life course model and expected timeframe for fully realised impact of 5 years, the evaluation is in the mid-phase data collection. The aim of this phase was to collate available and accessible data and determine relationships between the context, the FLFRPS, the process of implementation, and outcomes relating to changes in literacy. The ultimate aim is to establish a causal relationship between the input of the program and now plan scores.

The methodology and capacity building applied within the past 18 months aims to establish the foundation such that longitudinal causal analysis can take place to determine the overall impact of the program. The next phase of the evaluation will explore through observation, interviews and further data collation the impact and implementation of the program. The closing is report to be delivered in February 2018, focusing on understanding an explanation the relationships between the program and outcomes.

The results indicate that the implementation of the FLFRPSP has been, to some extent, a complex and intricate solution for a 'wicked problem' of how best to enhance quality literacy learning in remote Australia.

REFERENCES

Aikman, A. (2016, September 2). Labor will let remote schools choose Direct Instruction: Gunner. The Australian.

Al-Makahleh, A. A. (2011). The Effect of Direct Instruction Strategy on Math Achievement of Primary 4th and 5th Grade Students with Learning Difficulties. International Education Studies, 4(4), 199-205.

Ashdown, D., & Bernard, M. (2012). Can Explicit Instruction in Social and Emotional Learning Skills Benefit the Social-Emotional Development, Well-being, and Academic Achievement of Young Children? Early Childhood Education Journal, 39, 397-405.

Australian Council for Educational Research. (2013). Evaluation of the Cape York Aboriginal Australian Academy Initiative. Retrieved from

Barber, M., Rodriguez, N., & Artis, E. (2016). Deliverology in Practice: How Education Leaders Are Improving Student Outcomes. USA: Corwin.

Berman, P. (1980). Toward an Implementation Paradigm of Educational Change.

Blatchford, P., Webster, R., & Russell, A. (2012). Challenging the role and deployment of teaching assistants in mainstream schools: The impact on schools. Final report on findings from the Effective Deployment of Teaching Assistants (EDTA) project. Retrieved from http://www.teachingassistantresearch.co.uk

Blik, H., Harskamp, E., & Naayer, H. (2016). Strategy Instruction versus Direct Instruction in the Education of Young Adults with Intellectual Disabilities. Journal of Classroom Interaction, 51(2), 20-35.

Borman, G., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive School Reform and Achievement: A Meta-Analysis. Review of Educational Research, 73(2), 125-230.

Botts, D., Losardo, A., Tillery, C., & Werts, M. (2014). A Comparison of Activity-Based Intervention and Embedded Direct Instruction When Teaching Emergent Literacy Skills. The Journal of Special Education, 48(2), 120-134.

Cadette, J., Wilson, C., Brady, M., Dukes, C., & Bennett, K. (2016). The Effectiveness of Direct Instruction in Teaching Students with Autism Spectrum Disorder to Answer ‘‘Wh-’’ Questions. J Autism Dev Disord, 46(2968-2978).

Camillus, J. (2008). Strategy as a Wicked Problem. Harvard Business Review: Strategic Planning(May).

Celik, S., & Vuran, S. (2014). Comparison of Direct Instruction and Simultaneous Prompting Procedure on Teaching Concepts to Individuals with Intellectual Disability. Education and Training in Autism and Developmental Disabilities, 49(1), 127-144.

Centers for Disease Control and Prevention (1999). *Framework for Program Evaluation in Public Health*. US Department of Human Services.

Chandler, P. (2016a). $2 million invested in new works and upgrades to Arnhem school with Boosting Our Economy package [Press release]. Retrieved from http://newsroom.nt.gov.au/mediaRelease/18400

Chandler, P. (2016b). Direct Instruction improving Northern Territory student outcomes [Press release]

Chinman, M., Imm, P., & Wandersman, A. (2004). Getting To Outcomes™ 2004. Retrieved from http://www.rand.org

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. Educational researcher, 32(6), 3-12.

Considine, T. (2016a). Indigenous Education Strategy Newsletter, Issue 3. Retrieved from https://education.nt.gov.au/

Considine, T. (2016b). Indigenous Education Strategy Newsletter, Issue 4. Retrieved from https://education.nt.gov.au/

Considine, T. (2016c). Indigenous Education Strategy Newsletter, Issue 9. Retrieved from https://education.nt.gov.au/

De Bortoli, L., & Thomson, S. (2010). Contextual factors that influence the achievement of Australia's Indigenous students: Results from PISA 2000–2006. Victoria: ACER Press.

Dow, L. (2011). Spiders Are Mammals: Direct Instruction in Cape York. Literacy and Numeracy Studies, 19(1), 50-65.

Education HG News Team. (2015, February 20). Even cyclone can't disrupt NT schooling. Education HQ Australia.

Engelmann, S., & Carmine, D. (1982). Theory of Instruction: Principles and Applications. New York: Irvington Publishers.

Engelmann, S., & Engelmann, K. E. (2004). Impediments to scaling up effective comprehensive school reform models Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions (pp. 107-133).

Ewing, B. (2011). Direct Instruction in Mathematics: Issues for Schools With High Indigenous Enrolments: A Literature Review. Australian Journal of Teacher Education, 36(5).

Firdaus, F., Wahyudin, & Herman, T. (2017). Improving primary students' mathematical literacy through problem based learning and direct instruction. Educational Research and Reviews, 12(4), 212-219.

Fisher, D., Frey, N., & Hattie, J. (2016). Visible Learning for Literacy, Grades K-12: Implementing the Practices That Work Best to Accelerate Student Learning: Corwin.

Gaston, A., Martinez, J., & Martin, E. P. (2016). Embedding Literacy Strategies in Social Studies for Eighth-Grade Students. Journal of Social Studies Education Research, 7(1), 73-95.

Grossen, B. (2004). Success of a Direct Instruction Model at a Secondary Level School With High-Risk Students. Reading & Writing Quarterly, 20(2), 161-178.

Haas, M. (2005). Teaching methods for secondary algebra: A meta-analysis of findings. NASSP Bulletin, 89(642), 24-46.

Harris, A., Day, C., Hadfield, M., Hopkins, D., Hargreaves, A., & Chapman, C. (Eds.). (2013). Effective leadership for school improvement (2nd ed.). London: RoutledgeFalmer.

Hattie, J. (2008). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Hoboken Taylor and Francis.

Hattie, J., & Donoghue, G. (2016). Learning strategies: a synthesis and conceptual model. npj Science of Learning, 1.

Holden, A. (2016). Interim Evaluation of Flexible Literacy for Remote Primary Schools. Retrieved from

Hollingsworth, J., & Ybarra, S. (2009). Explicit Direct Instruction: The Power of the Well-Crafted, Well-Taught Lesson. USA: Corwin Press.

Honig, M. I. (2006). New Directions in Education Policy Implementation: SUNY Press.

Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. Educational researcher, 33(8), 16-30.

i3. (2017). Scaling Up Evidence-Based Practices: Strategies from

Investing in Innovation (i3). Retrieved from https://i3community.ed.gov

Kanfush, P. (2014). Dishing Direct Instruction: Teachers and Parents Tell All! The Qualitative Report, 19, 1-13.

Lencioni, G. M. (2013). The Effects of Explicit Reading Strategy Instruction and Cooperative Learning on Reading Comprehension in Fourth Grade Students. (Doctor of Education), The University of San Francisco.

Levin, B. (2013). What Does It Take to Scale Up Innovations? An Examination of Teach for America, the Harlem Children's Zone and the Knowledge is Power Program. Retrieved from http://nepc.colorado.edu/publication/scaling-up-innovations

Louis, K. S., Leithwood, K., Wahlstrom, K. L., Anderson, S., Michlin, M., Mascall, B., & Moore, S. (2010). Learning from leadership: Investigating the links to improved student learning. Minneapolis, MS: University of Minnesota.

Lowe, G., & Balcher, S. (2012). Direct instruction and music literacy: one approach to augmenting the diminishing? . Australian Journal of Music Education, 1, 3-13.

Luke, A. (2014). On Explicit and Direct Instruction: Australian Literacy Educators' Association.

Luke, A., Dooley, K., & Woods, A. (2011). Comprehension and content: planning literacy in low socioeconomic and culturally diverse schools. Aust. Educ. Res., 38, 149-166.

Lund, E., & Douglas, W. (2016). Teaching Vocabulary to Preschool Children With Hearing Loss. Exceptional Children, 83(1), 26-41.

MacTiernan, A. (2016). Inquiry into educational opportunities for Aboriginal and Torres Strait Islander students.

McCollow, J. (2012). The Cape York Aboriginal Australian Academy Three Years On: What is the Evidence? What Does It Indicate? . Paper presented at the AARE APERA International Conference, Sydney.

Meyers, D., Durlak, J., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps in the implementation process. American journal of community psychology, 50(3-4), 462-480.

Mourshed, M., Chijioke, C., & Barber, M. (2010). How the world's most improved school systems keep getting better. New York: McKinsey & Company.

Rebar, M. (2007). Academic Acceleration in First Grade Reading Using the Direct Instruction Model, Technical Report #2007-1. Retrieved from

Robinson, K. (2001). Mind the gap: The creative conundrum. Critical Quarterly, 43(1), 41-45.

Shippen, M., Houchins, D., Steventon, C., & Sartor, D. (2005). A Comparison of Two Direct Instruction Reading Programs for Urban Middle School Students. Remedial and Special Education, 26(3), 175-182.

Snyder, S. (2013). The Simple, The Complicated, and the Complex: Educational Reform Through the Lens of Complexity Theory. OECD Education Working Papers, 96. doi:http://dx.doi.org/10.1787/5k3txnpt1lnr-en

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. Review of Educational Research, 72(3), 387-431.

Stewart, J. (2002). The Relevance of the “Learning Styles Debate” for Australian Indigenous Students in Mainstream Education. The Australian Journal of Indigenous Education, 30(2), 13-19.

Stufflebeam, D. L. (2000). Evaluation Models. In D.L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), (2nd ed.). Boston: Kluwer Academic Publishers.

Taylor, P. (2015, March 21). Students blossom as Noel Pearson's way hits Pilbara. The Australian. Retrieved from http://www.theaustralian.com.au/

Terzon, E. (2015, May 13). Remote Peppimenarti Indigenous school looks to Direct Instruction trial to solve literacy gap. ABC Radio Darwin. Retrieved from http://www.abc.net.au/

Wheldall, K., Beaman, R., & Langstaff, E. (2002). ‘Mind the Gap’: Effective Literacy Instruction for Indigenous Low-Progress Readers. Australian Journal of Special Education, 34(1), 1-16.

Wolgemuth, J., Savage, R., Helmer, J., Lea, T., & Harper, H. (2011). Using Computer-Based Instruction to Improve Indigenous Early Literacy in Northern Australia: A Quasi-Experimental Study. Australasian Journal of Educational Technology, 27(4), 727-750.

Young, T., & Lewis, W. D. (2015). Educational Policy Implementation Revisited. Educational Policy, 29(1), 3-17. doi:10.1177/0895904815568936.