Evaluation of the Flexible Literacy for Remote Primary Schools Program

2015, 2016 and 2017 School Years

4 June 2018

Investigators:

Professor Janet

Georgia Dawson

Project Team:

Myra Koelle

Patrick Mclaren

Esther Kee

Katla Hower

Table of Contents

List of Figures

List of Tables

Abbreviations

Name	Abbreviation
Analysis of Variance	ANOVA
Centre for Program Evaluation	CPE
Context, Input, Process, Product Model	CIPP
Direct Instruction	DI
Explicit Direct Instruction	EDI
Early Years Literacy and Numeracy Data	EYLND
Flexible Literacy for Remote Primary Schools Program	FLFRPSP
Good to Great Schools Australia	GGSA
Language background other than English	LBOTE
Multivariate Analysis of Variance	MANOVA
Missing Completely at Random	MCAR
National Institute for Direct Instruction	NIFDI
University of Western Australia	UWA

1. Introduction

1.1. Flexible Literacy for Remote Primary Schools Program

In 2014, Good to Great Schools Australia (GGSA) was contracted by the Australian Government to implement the FLFRPSP using the principles of National Institute for Direct Instruction (NIFDI), Direct Instruction and Explicit Direct Instruction. The program's primary objectives are to:

- improve students' literacy abilities and results

- increase teacher pedagogical skills in teaching literacy using alphabetic teaching approaches, in particular, Direct Instruction (DI) or Explicit Direct Instruction (EDI).

Direct Instruction involves the use of explicit teaching techniques and is associated with the instructional approach and curriculum materials developed by Siegfried Engelmann and Carl Bereiter in the late 1960s. Teachers trained in DI implement a lesson-by-lesson approach to help students acquire a particular skill. Lessons are designed to work in sequence, and the pace is carefully controlled and designed to support incremental progress. Teachers implementing DI aim to maximise time on-task. Teachers also give students positive reinforcement to support their success at each level of attainment. The DI literacy curriculum is based on five components of reading: phonics, phonemic awareness, vocabulary, fluency and comprehension (National Institute for Direct Instruction, 2018).

Explicit Direct Instruction (EDI) is a step-by-step guide to creating and delivering lessons across the Australian Curriculum. The model was developed by John Hollingsworth and Dr Silvia Ybarra from DataWORKS, and is based on educational theory, neuroscience and DI.

1.2. Good to Great Schools Australia

Good to Great Schools Australia (GGSA) is a not-for-profit organisation that aims to help schools transition through the following categories: 'Poor to Fair', 'Fair to Good', 'Good to Great', and on to 'Excellent'. 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, each program is tailored to meet the diverse needs of schools and students, with schools having the autonomy to implement some or all elements of the program in relation to their specific needs and those of their students. Currently, GGSA is focused on addressing the long-term trend of underachievement in Australia by supporting schools who are in the 'Poor to Fair' category to transition into higher levels of achievement. Improving teaching practice is considered particularly important for enabling the improvement of low-performing schools and is therefore a key focus in working with these schools (Hattie, 2008). The FLFRPSP is a federal policy response to addressing low literacy performance; it is one of many programs that GGSA has undertaken to support school improvement among those in the 'Poor to Fair' category.

GGSA is responsible for introducing and implementing FLFRPSP, using either EDI or DI, in remote primary schools in NT, QLD and WA. In 2015, GGSA began working with 33 schools and as of July 2017, there were 35 schools in the program across the three states and territories. At the end of 2017, there were 34 schools in the program, as shown in Table 1. Following the submission of the Main Report in 2017 and additional submissions by other stakeholders, a review of program funding determined that 18 schools would continue with the program in 2018. The Main Report was submitted as part of the evaluation in June 2017. This report tabled the findings of the evaluation of the FLFRPSP for the school years 2015 to 2016 and used 2014 student progress as a baseline. This report provides summative information about the impact of FLFRPSP up to the end of 2017.

Table 1: Number of Schools in FLFRPSP by Jurisdiction and Type at the End of 2017

Jurisdiction	DI	EDI	Total
NT Government	17	0	17
NT Independent	3	1	4
WA Government	3	1	4
WA Independent	1	0	1
WA Catholic Education	4	3	7
QLD Government	0	1	1
Total	28	6	34

1.3. Evaluation of the Flexible Literacy for Remote Primary Schools Program

1.3.1. Project Background

The technical report submitted to GGSA in 2016 (Clinton, Au, McLaren, Dawson, & Afrin, 2016) documented the establishment of the evaluation, including the development of relationships with the project team, building the database, obtaining ethics approvals from relevant jurisdictions, and collecting 2015 program and performance data. Importantly, this stage provided the foundation for appraisal of data quality and access, as well as identifying data coverage to highlight where gaps were evident. This stage also helped to identify priorities for new data collection as well as the need to develop evaluation capacity within GGSA and schools. The technical report produced in 2016 was limited by the available data, but reported on implementation progress such as school fidelity, teaching fidelity, and student progress. CPE created a customised Microsoft Access® database, containing internal program data, which will be provided to GGSA at the end of the evaluation to allow them to continue monitoring and evaluating the program.

Overall, the first component of the evaluation (a) ensured that GGSA had the capacity for ongoing evaluation, and (b) incorporated a rigorous analysis of available data on the early impact of FLFRPSP on several outcomes, including teacher practice and student literacy as measured by internal program data.

In late 2016, management of the contract transferred from GGSA to the Australian Government, and the scope of the contract expanded to a full evaluation design, allowing for broader collection of data. This design enabled a more robust and nuanced analysis of the impact of FLFRPSP which would guide policy and provide credible evaluative information back to GGSA and participating schools. This evaluation design, described comprehensively in the 2017 report (Clinton, Dawson, McLaren, & Koelle, 2017), sought to investigate the program's effectiveness in (1) improving the literacy results of students in participating schools; and (2) improving teachers' pedagogical skills in teaching literacy in the DI and EDI models.

This report provides a final analysis of the impact of FLFRPSP in participating schools at the end of 2017 and complements quantitative data analysis and previous findings with richer qualitative data, including a triangulation of all data gathered. A more nuanced understanding of the implementation of FLFRPSP and its impact is presented in this report, based on 2017 student performance data, 2017 program data (Terms 1, 2 and 3), and interviews and surveys with school staff and key stakeholders, as well as an expanded literature review and synthesis.

1.3.2. Evaluation Framework and Approach

While the fundamental principle behind evaluation is to make a judgement of the merit, worth, and significance of an intervention or program, evaluation information should always be utilised to guide program development and hence add value. The current evaluation utilised an adapted form of the Centers for Disease Control and Prevention for Public Health Evaluation Framework (Centers for Disease Control and Prevention, 1999; see *Figure 1*) to embed evaluative principles within the evaluation. The framework establishes a continuous cycle of consultation and feedback between all stakeholders at each stage of the evaluation. Foundational to this cycle

are the evaluation standards of utility, feasibility, propriety, and accuracy (Yarbrough et al., 2011). Given the nature of this evaluation and the sensitivities that have become apparent in CPE's initial work with the jurisdictions, these standards are critically important.
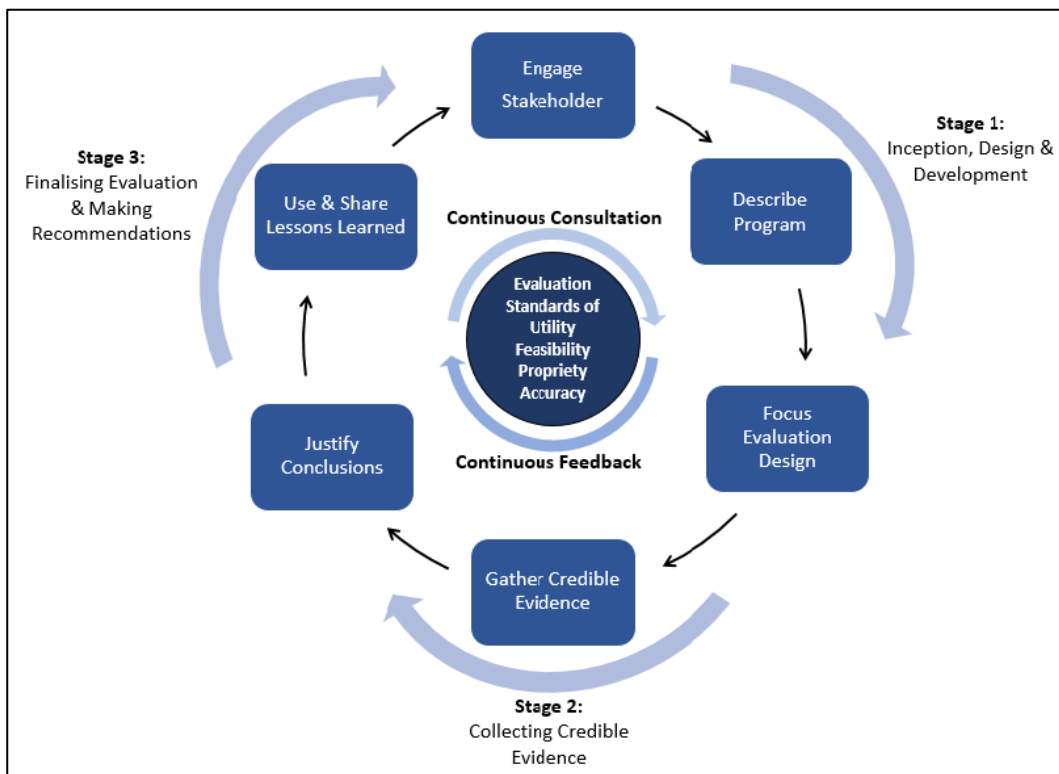
Figure 1: Evaluation framework

Phase two of the evaluation (the 2017 report) was focused on investigating the short-to-medium-term outcomes of implementation and impact of the FLFRPSP. The current phase is focussed on investigating the long-term effects of FLFRPSP, utilising both quantitative and qualitative data sources to provide information about the program's progressive impact through a theory and context-specific perspective.

Figure 2 below summarises the phases of the evaluation and key outputs, highlighting that this report is on phase two of the evaluation.
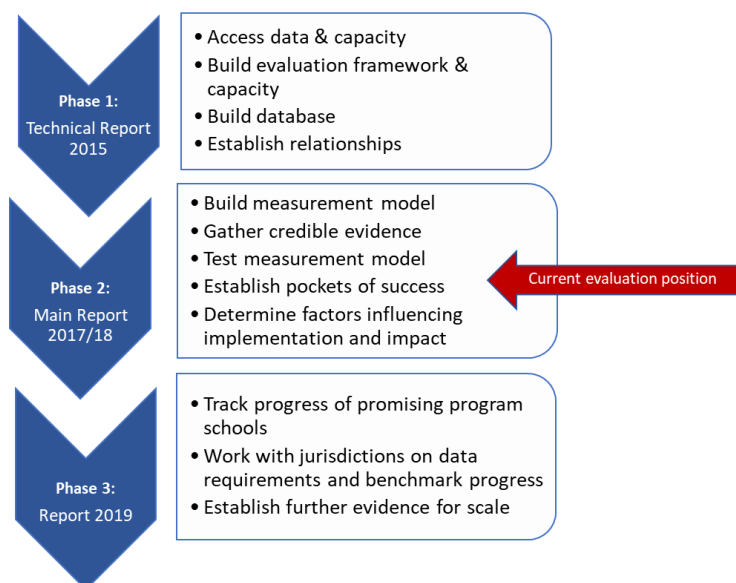
Figure 2: Evaluation progress

At this final phase of the evaluation, where data permits, the key focus is on testing the causal links defined by the theory of change between the implementation of FLFRPSP and literacy outcomes. At this phase of program implementation for all schools (see Figure 2) the measurement model should be tested to understand causal factors associated with the program's impact on student and teacher outcomes. This will be further extended in 2018 with the reduced schools involved in the program.

1.3.3. Developing a Measurement Model

The Context, Input, Process, Product (CIPP) model of evaluation (Stufflebeam, 2000) provides an underlying structural lens to view data collection and establish a long-term 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.
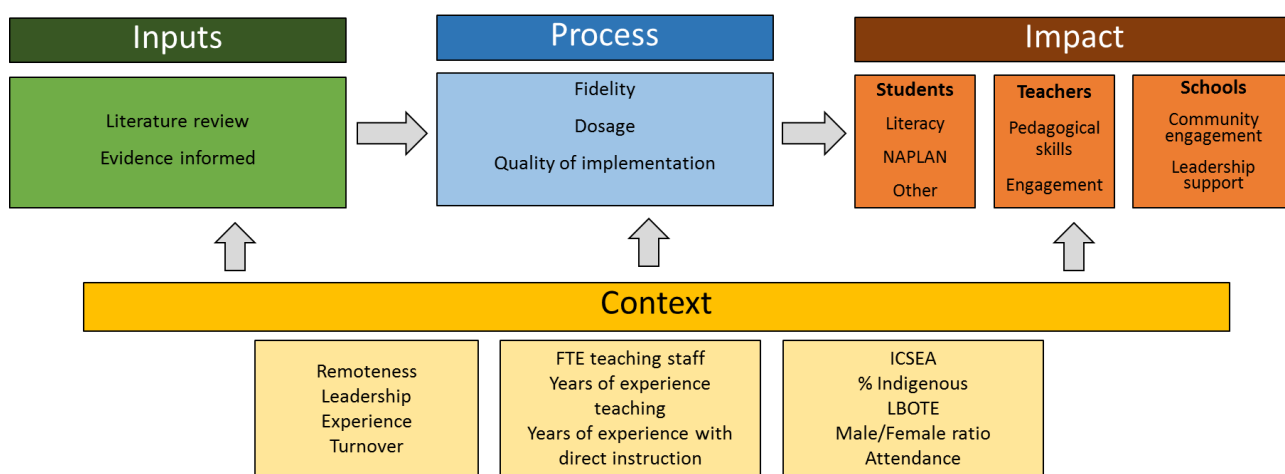
Figure 3: The CIPP model

1.3.4. Context

In undertaking this evaluation, it is critical that the contexts in which the program is implemented are considered and understood. One challenge is that the contexts are not homogenous at a student, school, or jurisdictional level. The evaluation is not only concerned with where and how the program is working, but equally, what factors might explain why the program works better in some contexts compared to others. Context therefore encompasses the environment within which FLFRPSP is implemented. 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 2*.

Table 2: Contextual Variables at Student, Teacher, and School Level.

Domain level	Contextual variable
Student	ICSEA; SEA; % Indigenous; LBOTE; Male-female ratio; Attendance
Teacher	Years' experience teaching; years' experience with DI; teacher turnover
School	School remoteness; leadership turnover; # FTE staff

1.3.5. Input

The input component of this evaluation focuses on understanding the program in the context of the problems it is seeking to address, namely, improving 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 can be expected by way of a benchmark or comparative degree of progress for these types of interventions. The evaluation questions for this component include:

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

1.3.6. Process

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

1.3.7. Product or Impact

Finally, the CIPP model describes product evaluation as the measurement, interpretation, and judgement of a program's outcomes (Stufflebeam & Coryn, 2014). Product, within the evaluation of the FLFRPSP, is referred to as impact and includes measuring the outcomes of the program among the following stakeholders:

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

2. This Report

2.1.   Purpose of this Report

The purpose of this report is to present the findings of the evaluation for the 2015, 2016, and 2017 school years. This report will contribute to the previous findings of the evaluation with regards to changes in student outcomes and perspectives of key stakeholders within FLFRPSP and will provide a culminated view of the program from its implementation through to the end of 2017.

2.2.   Structure of this Report

This report began with an executive summary and introduction, which gave an overview of the evaluation's background, framework, and methodology. Following this, the report provides the methodology section, which outlines the evaluation questions, data collection methods, and analytical methods. The successive results section first presents the evidence base of the program's components through a systematic literature review and meta-analysis. It then provides results from each state or jurisdiction including, where available:

1. Demographic information

2. Program implementation progress

3. Standardised student literacy outcomes

4. Impact evidenced through other data sources

5. Teacher- and school-level data

6. An overall summary

In the subsequent discussion, all findings are critically examined and contextualised within relevant research and theory.

3. Methodology

3.1. Rationale

The overarching goal of this evaluation is to provide rigorous information to the Australian Government regarding the impact of FLFRPSP on teacher pedagogical skills and student literacy outcomes. In doing so, CPE has endeavoured to:

- leverage and build on evaluation work to date, including data collected
- work with the Australian Government to 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

These aims are underpinned by the Joint Committee on Standards for Educational Evaluation (1999). The standards underscore four factors: utility, feasibility, propriety, and accuracy. Further, CPE has collected this data with adherence to both university and jurisdictional ethical requirements and guidelines.

There are two interrelated key aspects to the methodology driving this evaluation. First, FLFRPSP is endeavouring to affect educational reform in heterogeneous and sensitive contexts. This complexity must be a factor in deciding what data is collected, how it is collected, and how this data is analysed and interpreted. Second, the evaluation takes a long-term view, considering 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. These two aspects are interrelated because educational reform in any complex environment takes time. Evaluation plays a critical role in the provision of this time by futureproofing 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 evaluation seeks to answer two high-level questions:

- Does the program improve students' literacy abilities and results?
- Does the program increase teachers' pedagogical skills in teaching literacy through the use of alphabetic teaching approaches?

These questions will be addressed using a mixed-methods evaluation design, with data collected from a range of sources.

3.2. Description of the Data Sources

Table 3 below provides an overview of data sources in the evaluation, who was involved and how the data was used.

Table 3: Data Sources Used in the Evaluation

Data source	Description	Type	Use	Participants
FLFRPSP Program Data	Data pertaining to student placement in the program, student mastery levels, staffing number, records of program activities such as coaching sessions and training, and implementation and fidelity indicators.	Quantitative	Provides indicators of program implementation and fidelity, as well as staff turnover.	Students, teachers, school leadership.
State/Jurisdictional Literacy Data	Standardised literacy achievement data: Early Years Literacy and Numeracy Data (EYLND) – WA Catholic	Quantitative	An indicator of the impact of the program on literacy outcomes considered indicative of literacy achievement.	Students
NAPLAN	Standardised annual assessment for students in Years 3 and 5. Tests utilised in this evaluation were Reading, Writing, Spelling, and Grammar and Punctuation for program and control schools for the assessment years 2014, 2015, 2016, and 2017.	Quantitative	An indicator of the impact of the program on literacy outcomes considered more general than state/jurisdictional-based literacy measures.	Students
Teacher Survey	Teaching staff were invited to complete surveys relating to their perspectives and experiences with the program. Initially, these surveys were distributed online, but following a low response rate, a shorter paper-based version was posted to schools. Survey questions pertained to teacher perceptions of the following domains: general demographics, knowledge of the program, training, implementation, fidelity, attitudes towards the program, perceptions of student outcomes, teacher self-efficacy, job satisfaction, and wellbeing.	Mixed, purpose-designed	Stakeholder perceptions of the programs implementation and impact were triangulated with quantitative program and literacy achievement data. This enabled a more holistic understanding of how the program functions within schools.	Teachers (as respondents), students.
Principal Interviews	Principals were interviewed about their experiences with and perspectives of the program. More specifically, questions related to the program's initial	Qualitative	Stakeholder perceptions of the programs implementation and	Principals

Data source	Description	Type	Use	Participants
	implementation and ongoing practice, as well its impact on student, teachers, schools, and the community more broadly.		impact were triangulated with quantitative program and literacy achievement data. This enabled a more holistic understanding of how the program functions within schools.	
Teacher Interviews	Teaching staff were interviewed about their experiences with and perspectives of the program. More specifically, questions related to the programs initial implementation and ongoing practice, as well its impact on student, teachers, schools and the community more broadly.	Qualitative	Stakeholder perceptions of the programs implementation and impact were triangulated with quantitative program and literacy achievement data. This enabled a more holistic understanding of how the program functions within schools.	Teachers

3.3. Quantitative Methods

3.3.1. Analysis Procedures Utilised

Preparation of the quantitative data included analysing missing data and managing missing data using several methods:

- Missing Completely at Random Test: Little's Missing Completely at Random (MCAR) Test 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.

- Multiple Imputation: Multiple imputation is a statistical method used to overcome the problem of missing data points within a data set. Studies that have a large amount of data missing and exclude participants with missing values can have decreased statistical power and increased probability of bias. In the present evaluation, many students are missing one or more data points. Multiple imputation was chosen to increase the sample size and improve the power of analysis.

  This method replaces missing data values with a set of plausible values estimated using other information within the data set. Unlike other models of imputation, where a missing value is replaced by one other value, multiple imputation generates multiple values. This has been found to be more accurate than single imputation at estimating what the results would have been if values were not missing. During analysis these values are pooled to minimise bias and inaccuracies in the results.

After the initial data was cleaned and managed, several methods were used to analyse the data:

- Analysis of Variance (ANOVA): ANOVAs are a family of statistical tests used to compare the means of different groups and to determine if there is a statistically significant difference between them. In some instances, in the present report, one or more assumptions were violated. A non-parametric test, Welch's ANOVA, was used. Welch's ANOVA can be used when the assumption of homogeneity of variance has been violated.

- 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. 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.

3.3.2. NAPLAN Data Management and Cleaning

The initial NAPLAN sample consisted of 652 students. No anomalies[1] were detected using the 2015 and 2017 NAPLAN Equivalence Tables. However, some missing data was present in the data set; thus, Little's MCAR Test was conducted. This indicated that data was not missing at random (χ^2(97) = 276.17, p < .001), thus, data was investigated for patterns. Examination revealed 122 (18.7%) of students did not have NAPLAN scores in 2015 and 2017, 220 (33.7%) have data at only one time point (either 2015 or 2017), and 310 (47.5%) students have data at two time points.

Given that 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). A one-way ANOVA was conducted to determine if attendances rates varied between these groups.

Assumption testing for ANOVA was conducted. While attendance data was normally distributed (skewness and kurtosis between ±1) with no 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* (2,424) = 23.90 p < .001. A Games-Howell post-hoc test revealed student attendance was significantly higher for students with data at *Two Time Points* (68% ± 25%, p < .001) when compared to the *One Time Point* (52% ± 32%) and *No Data* group (46% ± 31%). There were no statistically significant differences between the *One Time Point* or *No Data Points* groups. 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 4* [2]).
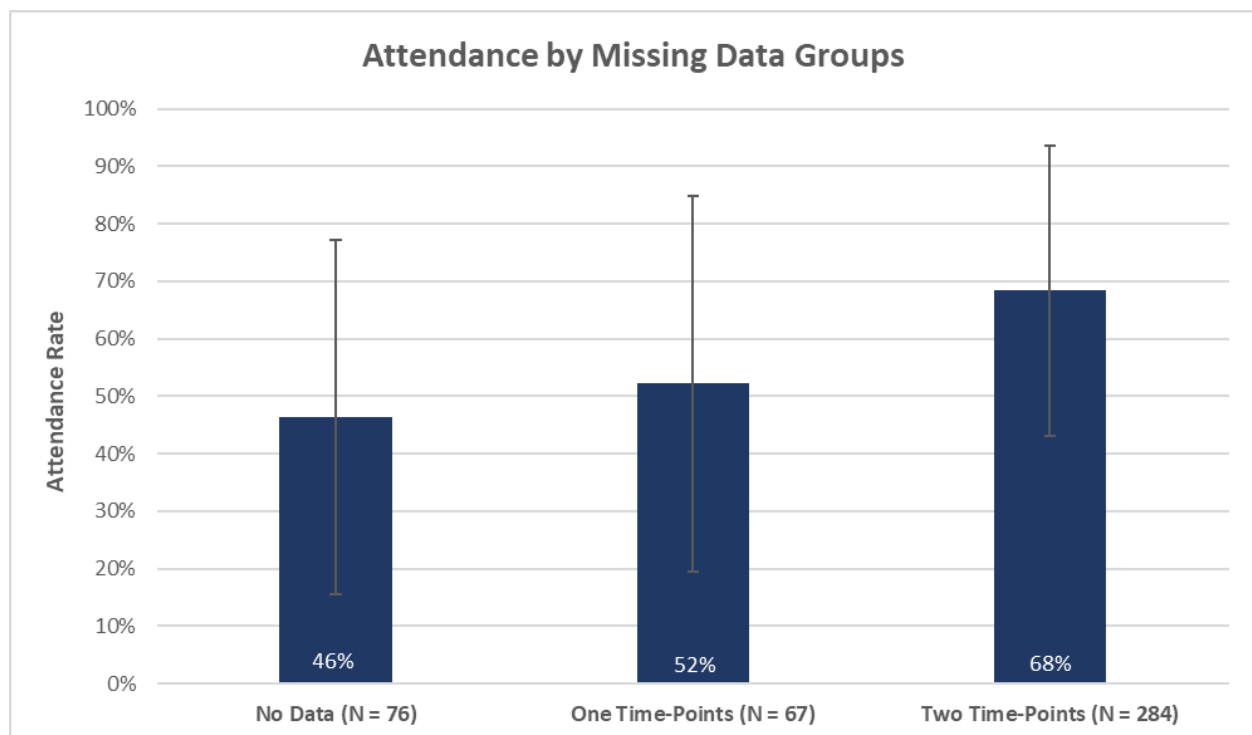
Figure 4: Mean difference in attendance (days) by missing data groups.

---

[1] Scale scores that are not possible according to the NAPLAN Equivalence Tables.

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

Due to the level and nature of missing data (non-random, 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 530. The remaining missing values are imputed based on attendance and other NAPLAN scores. Multiple imputation was conducted[3]. Throughout this report, if pooled multiple imputation results align with the original result, only the original results are presented.

3.4. Qualitative Methods

Qualitative data sources were analysed using the professional assistive research tool NVivo Pro. Analysis of the data comprised three stages:

1. **Descriptive Analysis:** Qualitative data sources categorised broadly according to topic or drive.

2. **Thematic Analysis:** Qualitative data sources in each category more closely analysed and coded into constituent themes, often specific ideas, concepts or points.

3. **Componential or Comparative Analysis:** Qualitative data sources iteratively analysed to check for missing information and to ensure resulting themes are coherent as a narrative structure until meeting adequacy criteria.

The outcome of this systematic process is the emergence of themes in the qualitative data, informed and refined by multiple stages of analysis. The resulting themes can be assessed against the evaluation questions.

3.4.1. Qualitative Data Sources

Surveys were sent out to participating school staff, who were asked to provide their perceptions of and experiences with the FLFRPSP. The survey questions included staff perception of knowledge of the program, training and implementation, attitudes towards the program, and perception of student outcomes and teacher self-efficacy. The aim of the surveys is to obtain stakeholder perceptions as indicators of the aforementioned measurement domains, which are to be triangulated with the quantitative program and literacy achievement data. Table 4 below shows the breakdown of survey respondents according to position and jurisdiction/state.

Table 4: Demographics of Survey Respondents

	Northern Territory - Government	Western Australia – Government	Western Australia – Catholic Education
Teachers	27	5	12
Teaching Assistants	1	3	4
Principals	3	-	4
Others	-	1	1
Total surveys	31	9	21
Total number of schools	8	2	7

In addition to the surveys, semi-structured interviews (n=26) were conducted with school-level practitioners (principals and teachers) according to a co-designed interview protocol by senior evaluation team members (see *Appendix 1)*. Participants (n=24) were from across three jurisdictions. Two participants from the NT were interviewed twice, once before and once after leaving the program. The breakdown and number of the types of

---

[3] 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.

interviews conducted are shown in Table 5 below. With participants permission these interviews were recorded, with recordings de-identified and partially transcribed by the evaluation team. In addition to interviews with school-level practitioners, a singular face-to-face semi-structured group interview was conducted with representatives of NIFDI.

Table 5: Summary of Qualitative Data Sources (Number of Semi-Structured Interviews)

	Northern Territory - Government	Western Australia – Government	Western Australia – Catholic Education
Teachers	2	0	3
Principals	12	3	4
Total interviews	16*	3	7
Total number of schools	12	2	6

*two participants from NT were interviewed twice each, once before and once after leaving the program

3.4.2. Descriptive Analysis

Qualitative descriptive analysis occurred at the parent nodal level in NVivo. This descriptive analysis, or a 'broad brush' descriptive coding, was derived directly from the interview protocol reflecting the topic structures contained within the protocol (*Appendix 2*). This descriptive structure generated from interview protocol topics generated a preliminary scaffold to both adapt and give further granularity to subsequent stages of thematic coding of child nodes.

3.4.3. Thematic Analysis

Thematic analysis of data sources occurred at the first-level and second-level child nodal levels. Thematic analysis was supervised by senior evaluation team members ensuring consistency and clarity of categorisation of first- and second- level child nodes and to ensure that constituent ideas, themes, and concepts coded for corresponded to indicators and outcomes of evaluative interest. A preliminary thematic narrative was generated by coders in Microsoft Word format to identify inconsistencies of input or omissions of data of interest. This preliminary thematic narrative also informed subsequent assessment by senior evaluation team members of the overall thematic coherence.

3.4.4. Componental or Comparative Analysis

The preliminary narrative generated by coders was appraised by senior evaluation team members for identification of inconsistencies as well as to ensure overall coherence. Feedback was provided to coders allowing for further creation of second-level child nodes, as well as the re-categorisation of data to reflect adjustments to overall nodal structure. Iterations of the overall thematic narrative were generated until senior evaluation team members were satisfied with the nodal structure and theme categories that adequately tracked outcomes of evaluative interest. A visual illustration of the iterative process is shown in *Figure* 5 below.
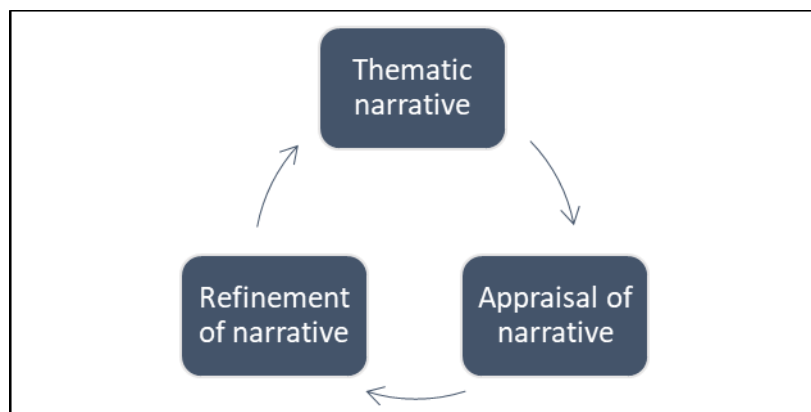
Figure 5: Iterative process of componental or comparative analysis

Results

3.5.   Literature Review

The previous 2017 report presented the results of a rapid literature review on DI pedagogies. The current report expands those findings with more recently published literature and provides research evidence for additional areas identified as important to the program's underlying foundation. These areas are component reading skills, positive behaviour support, and mastery learning.

3.5.1.   Summary of Previous Literature Review

The previous review of literature showed there was a growing body of evidence for the efficacy of direct instruction across a number of educational domains. Meta-analyses examining efficacy of teaching methods have found direct instruction to be one of the most effective teaching methods, with effect sizes ranging from d = .21 to d = .59 (Borman et al., 2003; Hattie, 2008). More specifically, studies investigating the effectiveness of direct instruction have found it to be an effective method for improving literacy, maths, and wellbeing outcomes. Direct instruction is also an effective teaching method 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 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 seems to be generally effective, studies in this area would benefit from more rigorous study design and larger sample sizes.

3.5.2.   Literature on Direct Instruction Since 2017

A review of recent research published since the previous report, released in July 2017, shows support for the established trends. Several studies have shown that direct instruction is statistically significant in primary and lower secondary students (Bedard, Bremer, Campbell, & Cairney, 2018; De Nigris, 2017; López, Torrance, Rijlaarsdam, & Fidalgo, 2017); however, these studies typically had small sample sizes, so their ability to generalise is limited. Other studies showed significant results in populations with learning difficulties, such as writing/literacy difficulties (Barth & Elleman, 2017, n=66), geometry difficulties (Zhang, 2017, n=4), and global delay (Özokcu, Akçamete, & Özyürek, 2017, n=3). Other studies compared the efficacy of direct instruction with other teaching approaches among various populations: Datchuk (2017) compared direct instruction and precision teaching, an iterative teaching process in which teaching practice and behaviours are altered to promote continued gains, among 15 students with writing difficulties in years 5-8. Datchuk found that both teaching approaches were effective for increasing written expression abilities. Chase and Klahr (2017) compared invention/'discovery-oriented' methods with direct instruction methods among 101 year 4 and 5 students, and found that both approaches yielded significant results with similar outcomes. Kuder (2017) compared studies examining various vocabulary instruction methods. One study, which examined direct instruction's effectiveness in teaching vocabulary, found that students using direct instruction learned more vocabulary words than using other methods, but that they read fewer words than in other conditions. However, direct instruction was similarly effective in vocabulary instruction compared to other methods such as mnemonic instruction, morphemic analysis, and multimedia instruction (Kuder, 2017).

A 2018 meta-analysis of the effectiveness of direct instruction indicates very positive results on all academic subjects examined, including reading, language, spelling, and math (Stockard, Wood, Coughlin, & Khoury).

Two reviews of direct instruction research have been published since the beginning of 2017. Hughes, Morris, Therrien, & Benson (2017) reviewed 68 publications which had explicit instruction as a primary focus and identified the most commonly used components of explicit instruction. Five main components, which were identified in at least 75% of the publications, were identified as essential components of explicit instruction. These components were a) segmenting complex skills, or 'chunking', in which complex tasks are broken down into more manageable units of instruction; b) drawing attention to important features through modelling or think-alouds, in which students are provided with clear and consistent descriptions and demonstrations of a skill; c) promotion of successful engagement by using systematically faded supports/prompts, in which a strategy or skills performed by a student is promoted through gradually decreasing levels of guidance and prompts, with an aim for student independence; d) providing opportunities for students to respond and receive feedback, in which

student responses allow the teacher to give corrections or affirmations, or modify instruction; and e) creating purposeful practice opportunities, in which students can practice and therefore retain and generalise new skills and knowledge.

Although Hughes et. al. (2017) distinguished between explicit instruction and direct instruction as being distinct but overlapping approaches, the authors also conceded that direct instruction and explicit instruction have been considered one and the same by much of the literature. Zhao (2017) conducted an overview of the historical and ongoing debates within DI studies and education, specifically identifying the potentially negative 'side effects' of direct instruction. Zhao critiqued the traditional approach of educational research, stating that "the lack of a tradition of considering side effects as an integral part of effectiveness in educational research has resulted in two irreconcilable bodies of literature: one proving [DI's] effectiveness…and the other condemning direct instruction" (2017, p. 5). Zhao suggested that the negative effects of direct instruction be acknowledged along with the positive effects, noting that, for example, direct instruction can improve student test scores while also suppressing creativity and problem solving.

While the studies published since 2017 contribute to the evidence suggesting that direct instruction is a generally effective approach in classrooms, when evaluating the effectiveness of direct instruction, it is also important to consider more closely the individual components of direct instruction that prove effective, as well as to consider the potential negative effects that direct instruction may produce. Nonetheless, while the FLFRPSP is a direct instruction based approach, other components of the delivered program need to be considered. These additional components are the five components of reading, positive behaviour support, and mastery learning, all of which contribute to the impact of the FLFRPSP. Literature-based evidence for their effectiveness is discussed below.

3.5.3.   5 Components of Reading Comprehension

In 2000, the United States National Institute of Child Health and Human Development (NICHD) established the National Reading Panel (NRP), which conducted a large-scale review of reading research and reading instruction, resulting in the definitive report Teaching Children to Read (National Reading Panel, 2000). This report synthesised decades of research and identified the five critical areas addressed by effective reading instruction: phonemic awareness, phonics, vocabulary, fluency, and comprehension (Collins & Collins, 2004). These areas have been accepted by educational practice as the 5 Components of Reading, and have had wide-reaching influence in literacy pedagogy (see Department of Education Science and Training, 2005; Langenberg, 2018; Shanahan, 2005; Centre for Education Statistics and Evaluation, 2017; Armbruster, 2001) and research (Konza, 2014; see Joshi et al., 2009; Wagner, Kim, & Foster, 2011). However, the NRP report has been widely criticised as being unscientific, having poor methodology or ignoring a majority of the existing research (Cunningham, 2001; Garan, 2001; Krashen, 2001). Regardless, the "Fab Five" reading components have been established as important factors in children's reading ability (Konza, 2014; Hattie, 2009). The influential report by Rose (2006), created for the US government, supports these five components as critical to literacy learning and stresses the importance of teaching synthetic phonics. The report also makes recommendations regarding best practice and the order in which literacy elements should be taught. Other research has cemented each of these five dimensions as important, if not critical, to the development of literacy ability (see Biemiller, 2005; Sedita, 2005; Hulme, Snowling, Caravolas, & Caroll, 2005; Caravolas, Hulme, & Snowling, 2001; and meta-analyses by Bus & van Ijzendoorn, 1999; Piasta & Wagner, 2010; Benner, Nelson, Ralston, & Mooney, 2010).

Phonemic Awareness

Phonemic awareness is "the understanding that spoken words are made up of separate units of sound", as well as the "skill at hearing and producing the separate sounds in words, dividing or segmenting words into their component sounds, blending separate sounds into words, and recognising words that sound alike or different" (Collins & Collins, 2004, p. 4). These separate sounds are called phonemes. Phonemes are individually recognisable speech sounds in a given language. They are not the same as letters, which are simply visual representations of speech sounds. Phonemic awareness in young children is an indicator of future literacy ability (Konza, 2014; Wimmer, Landerl, Linortner, & Hummer, 1991; but contra Castles & Coltheart, 2004; Castles, Coltheart, Wilson, Valpied, & Wedgwood, 2009). Sermier Dessemontet and de Chambrier (2015) found that "phonological awareness and letter-sound knowledge at 6-8 years old significantly predicted progress in word and non-word reading after one school year and two school years", with moderate effect sizes of $b=.34$ and $b=.35$ respectively. The National Reading Panel (2000) found that phonemic awareness training had significant

effects on reading ability (d = 0.53) and spelling (d = 0.59), which was retained at follow-up (d = 0.45 and d = 0.37, respectively). Its effect on reading comprehension was also measured (d = 0.32), although this effect was smaller than for word reading, likely due to the direct effect that phonemic awareness has on word reading, and word reading's effect on reading comprehension (National Reading Panel, 2000). A meta-analysis by Bus and van Ijzendoorn (1999) found that phonemic awareness instruction was effective in improving phoneme segmentation, phoneme blending, and sound deletion (d = 0.73).

There are several skills associated with phonemic awareness: isolating phonemes, in which individual speech sounds are identified in words; blending phonemes, in which individual speech sounds are blended together to create a word; phoneme identification, in which students identify common phonemes across several words; phoneme segmentation, in which words are divided up into their component phonemes; and phoneme addition or deletion, in which students are asked to either add or delete a phoneme into word (Collins & Collins, 2004; National Reading Panel, 2000).

Phonics

Phonics, also known as letter-sound knowledge, is the ability to associate speech sounds with the letters and combinations of letters used to represent them (Konza, 2014; National Reading Panel, 2000). This enables readers to "read and spell words accurately and rapidly" (Collins & Collins, 2004, p. 12). A meta-analysis by Piasta and Wagner (2010) found that alphabet knowledge was impacted to differing degrees depending on the type of instruction provided, but that school-based instruction was more effective than home-based instruction, and that small-group instruction was more effective than individual tutoring. This finding supports the small-group based design of DI/EDI. A meta-analysis by Ehri, Nunes, Stahl, and Willows (2001) found that phonics instruction had a moderate overall effect on reading (d =0.41), with larger effects for early phonics instruction (d = 0.55) than after year one (d = 0.27). Hattie (2009) found that phonics instruction had an effect size of d = 0.60, based on 14 meta-analyses and 425 studies.

One challenge to teaching phonics in English is the inconsistency of certain spelling and pronunciation rules, e.g. irregular words or loanwords, which need to be learned individually (Konza, 2014). One common approach to phonics instruction is synthetic phonics, in which students identify individual sounds from letters and then blend them to create words. One influential longitudinal study into synthetic phonics found the approach to be very effective (Johnston & Watson, 2005): the group of students learning using the synthetic phonics approach scored significantly higher on reading age, spelling age, emergent reading, letter knowledge, phonemic segmentation, rhyme skills, and nonwords than the other groups, who were taught using analytic phonics. Another approach to phonics instruction is larger-unit phonics, in which larger units than phonemes (e.g. syllables, spelling patterns) are identified and blended. The NRP also found this approach to be effective (d = 0.34; National Reading Panel, 2000), though less so than synthetic phonics instruction (d = 0.45). Hattie posits that direct instruction methods are the most effective in teaching phonics skills (2009).

Vocabulary

Vocabulary is the knowledge of words and their meanings. Collins and Collins identify four types of vocabulary usage: listening, speaking, reading, and writing (2004). If the spoken word is known, but its written counterpart is not, vocabulary knowledge allows readers to decode and recognise written words and match them with the meaning in their lexicon, therefore enabling readers to fill in meaning when reading. Vocabulary is typically learned indirectly and incidentally, resulting in a large vocabulary ability gap between children who are exposed to a large variety of words (e.g. children who are read to, children with educated parents) and children who are not exposed to these stimuli (e.g. lower-SES children) (Konza, 2014). Therefore, vocabulary deemed central to children's lives should be explicitly taught. An example of explicit vocabulary instruction is the three-tiered vocabulary model, where Tier 1 words are words commonly found in everyday life, rarely requiring instruction; Tier 2 words are more complex, abstract words, often occurring in academic settings; and Tier 3 words are domain-specific, specialised words which almost always need to be taught (Beck & McKeown, 1985).

Hattie (2009) found that vocabulary instruction was moderately effective in improving literacy (d = 0.67). The meta-analysis by Stahl and Fairbanks (1986) found that vocabulary instruction had a mean effect size of .97 on text comprehension. Vadasy, Sanders, and Nelson (2015) trialled a vocabulary instruction program with at-risk kindergarten ESL students and found that the experimental group made significant gains in both reading

vocabulary (d = 0.64) and decoding (d = 0.45), which were both noticeable at follow-up (d = 0.29 and d = 0.27 respectively).

Fluency

Fluency is "recognising the words in a text rapidly and accurately and using phrasing and emphasis" in a way that resembles natural speech and encourages comprehension (Collins & Collins, 2004, p. 17). It is the result of reading skills that are optimised to the point where "maximum cognitive energy is available to focus on meaning" (Konza, 2014, p. 161). For fluency to occur, the reader must be able to recognise familiar words on sight ('sight words'), either know or be able to discern their meaning from context, and be able to read the text at an adequate speed for comprehension (approximately 90-100 words per minute; Armbruster, Lehr, & Osborn, 2001). Wolf and Katzir-Cohen (2001) consider three types of processes as central to fluency: orthographic (e.g. letter-sound knowledge), phonological (phoneme awareness), and semantic (e.g. vocabulary). Klauda and Guthrie (2008) found that fluency, considered at the word, syntactic, and passage levels, related to reading comprehension performance. As such, reading comprehension and reading fluency are closely linked, and can be predictive of one another (Klauda & Guthrie, 2008; Kim, Wagner, & Foster, 2011).

Fluency is developed naturally through reading, but other methods are needed to target its development, such as repeated reading and guided repeated oral reading (National Reading Panel, 2000). Repeated reading involves students reading and then rereading a text. Collins and Collins (2004) posit that the effectiveness of this approach is due to a shift of focus from word recognition, as students read a text the first time, to comprehension and meaning, as they reread the text. Words that are not known or immediately recognised upon first reading them will become more familiar as the text is read subsequent times. One review found that repeated reading did not have significant effects on any measures, including reading comprehension, alphabetics, or fluency (What Works Clearinghouse, 2014). However, others have found significant effects; in their case study, Swan, Leader-Janssen, and Conley (2017) found that repeated reading had a positive and sustained effect, and Therrien (2004) reviewed the research to find that repeated reading of at least three times aloud had an effect size of 1.37.

Guided repeated oral reading offers more instruction and support, which may include modelling fluent reading (effect size = 0.40; Therrien, 2004), supplying unfamiliar words and their meanings, or having students read along with the teacher (Collins & Collins, 2004).

Comprehension

Comprehension is the ultimate understanding of a text at a deep level, and requires the use of multiple skills in concert, such as vocabulary knowledge and fluency (Konza, 2014). Reading comprehension is a cognitive process in which readers use various comprehension strategies to understand texts of different types, purposes, and difficulty levels (Collins & Collins, 2004). Although comprehension can be detected and measured, the mechanisms by which it works and develops are not well understood (Mckeown, Beck, & Blake, 2009; Pearson, 2009).

Asking questions and engaging with a text is one way to increase comprehension of the text, both during and after reading. Asking students questions about the text before they read it allows them to focus on the content more closely, while group discussions can help students share background knowledge and important points, thereby deepening their understanding of the text (Collins & Collins, 2004). More independent readers will also use strategies to increase their comprehension by making predictions, discerning minor details from crucial content, and rereading or slowing down when they begin to 'lose meaning' (Konza, 2014).

The National Reading Panel identified vocabulary instruction and text comprehension instruction as two areas that are important in developing comprehension skills (National Reading Panel, 2000). However, Wright and Cervetti (2017) conducted a systematic review on the role of vocabulary in comprehension and found that vocabulary instruction in and of itself did not improve general comprehension, but that explicit instruction of words found in target texts increased comprehension of those texts. These findings do not link vocabulary instruction with improved comprehension skills. Other studies have found that teaching a variety of comprehension strategies had a positive effect on overall reading comprehension (Dole, Duffy, Roehler, & Pearson, 1991; Guthrie, et. al., 2004; Mastropieri et al., 2003; Spörer, Brunstein, & Kieschke, 2009). Hattie (2009) reported the effect sizes of various approaches on reading comprehension, including concept mapping (d = 0.57), questioning (d = 0.46), metacognitive strategies (d = 0.69), and vocabulary programs (d = 0.67).

Still other studies found no significant link between comprehension strategies and increased comprehension outcomes (Mckeown et al., 2009). Elleman and Compton (2017) stress that reading comprehension is a complex process made up of several abilities, and that comprehension instruction should include approaches that address this complexity.

3.5.4.   Positive Behaviour Management

Many existing approaches to changing or managing student behaviour in the classroom focus on encouraging a shift towards positive behaviour. Commonly mentioned behaviour management strategies include Positive Behaviour Support (PBS) and Whole-School Behaviour Management (WSBM). PBS is a general approach that involves redesigning the environment and curriculum to support students in changing or reducing their behaviour (Chitiyo, Makweche-Chitiyo, Park, Ametepee, & Chitiyo, 2011; Dunlap et al., 2010; Sugai et al., 2000). PBS differs from more 'traditional' behaviour management methods in that it focuses on the influence of larger systems instead of viewing the student as a 'problem' (Chitiyo et al., 2011). PBS is derived from applied behaviour analysis and often makes use of behaviour interventions to effect change in student behaviour (Grey, Lydon, & Healy, 2016). WSBM is a clear, comprehensive framework of consistent systems and policies designed to address behaviour issues (Scott, 2005). WSBM, while closely related to PBS, places more emphasis on top-down school systems for managing behaviour, and can provide consistency and support for teachers working to implement behaviour management strategies in their classrooms (De Nobile, London, & El Baba, 2015), as well as providing students with clear, consistent expectations and involvement in their own behaviour plans. Scott (2005) identifies belonging (among students), relationships (between staff and students and between students and students), and engagement (of students in their own academic life) as key principles required to make a whole-school behaviour approach effective.

Drawing from research in educational psychology, Hart (2010) identified core evidence-based elements of effective classroom behaviour management; these core elements include consistent and fair rules, reinforcement of appropriate behaviour, response to undesired behaviour, a focus on positive staff-student relationships, high expectations, procedures for chronic misbehaviour, and a conducive classroom environment. Hart then further identified the psychological approaches behind effective classroom behaviour management. The first and foremost is a behavioural approach, in which desirable behaviour is increased through positive reinforcement, and undesired behaviours are discouraged using a lack of positive reinforcement (or "extinction" in psychological terms) (Hart, 2010). This mechanism is at work in PBS and WSBM and can include activities such as, for example, assertive discipline; a school-wide prize program, in which positive behaviour is rewarded with tickets to enter into a prize draw; or "the 'Good Behaviour' game in which pupils are put into teams and compete" to receive rewards for the team with the fewest incidences of negative behaviour (Hart, 2010, p. 357). Other psychological approaches active in behaviour strategies are psychodynamic approaches, based on the capacity of close, trusting relationships to model good behaviour; systemic approaches, focusing on the complex interplay of environment and background and the effect these have on behaviour; and humanistic approaches, focusing on the relationship between the teacher and the student and the care that teachers are expected to display (Hart, 2010). Each of these psychological approaches is used in behaviour management strategies such as PBS and WSBM.

3.5.5.   Mastery Learning

Mastery learning is a teaching method wherein students are taught subjects according to a specific order, and do not progress until they have reached a certain level of achievement, or 'mastery', of the preceding steps. Assisting students to achieve this level of mastery often involves individualised teaching. Mastery learning as an approach was first developed by Benjamin S. Bloom in the 1960s in response to a large variation in student achievement among students with differing learning styles and background (Guskey, 2007). Bloom's approach involves organising subjects and skills into units that take one to two weeks to complete, assessing the students on their comprehension of the learning goals, and, where needed, providing further instruction and 'corrective' activities for students that have not reached the threshold for mastery (Bloom, 1968; Guskey, 2007). Another method of mastery learning is Keller's Personalised System of Instruction (PSI), which differs from Bloom's mastery in that PSI lessons are largely student-driven and rely on written materials, allowing the student to control their own learning and pace (Kulik, Kulik, & Bangert-Drowns, 1990). Mastery learning is frequently applied in a much less prescriptive manner, although the central ideas as developed by Bloom and Keller remain the same (Deweese, 2012; Guskey, 2007).

Meta-analyses of studies on mastery learning have found that mastery learning is generally an effective practice for achievement, retention, and motivation (Hattie, 2008; Kulik et al., 1990), although some studies have found no evidence for mastery learning's effectiveness (Slavin, 1987), or the effectiveness of certain components, such as correctives (Deweese, 2012). However, the vast body of research, while mostly conducted in the 1980s and 1990s, favours mastery leaning as an effective approach (Deweese, 2012).

3.5.6.   Similar Programs

MiniLit

MiniLit is an early literacy intervention program aimed at Year 1 students who fall into the bottom 25% of readers at the end of one year of literacy instruction. The program consists of 80 one-hour lessons over 20 weeks, and are delivered to groups of up to four children by an instructor trained in the program. Each lesson consists of three components: Sounds and Words Activities, Text Reading, and Story Book Reading. The MiniLit program focuses on the five components of reading; that is, phonemic awareness, phonics, vocabulary, fluency, and comprehension, and also incorporates elements of positive behaviour support, mastery learning, and direct instruction. Students are led through structured lessons and are explicitly taught content through modelling and guided reading. Behaviour expectations are made explicit and appropriate behaviour is praised, while inappropriate behaviour is not addressed unless necessary. Correct responses are also praised. Students are assessed regularly to test their understanding and do not move onto the next level unless their understanding is adequate.

Buckingham, Wheldall, and Beaman (2012) found large and significant mean differences between students who received the intervention and students who did not; those in the experimental group outperformed in phonological recoding and single word reading.

Cape York Aboriginal Australian Academy

The Cape York Aboriginal Australian Academy (CYAAA) is a program run jointly by Education Queensland and Good to Great Schools Australia, which currently operates two primary school campuses in Queensland. The original pilot program began in 2010 with two schools and included an additional school in 2011. The CYAAA education model organises curriculum into three domains:

- Class: literacy, numeracy, and English language are taught using a direct instruction approach, with the aim of helping students to reach the national minimum standards for English literacy and numeracy.
- Club: provides students with artistic, musical, and sport activities to increase personal development and physical activity.
- Culture: targets Indigenous culture and language with the aim of giving students "fluency in their own culture" and providing "support for their identity as bicultural people" (ACER, 2013).

The 2013 ACER evaluation of the pilot program was unable to determine if the CYAAA initiative had a significant impact on student learning due to large amounts of unavailable or missing data. Evidence, while weak, indicated that more students were at or above the national minimum standards for Grammar and Punctuation, one literacy sub-strand within NAPLAN. Other outcomes showed that student behaviour had improved, teacher expectations of their students were higher, and staff turnover had reduced. Student attendance, however, had not increased, despite this being one main goal of the initiative.

In 2016, CYAAA reported 36 students (out of approximately 150) at their grade level in Reading and 47 students at their grade level in Numeracy, the highest number ever for the Academy (CYAAA, 2016).

Spelling Mastery

Spelling Mastery is a direct instruction program designed by NIFDI to help students from Year 1 to 6 improve their spelling skills. The program teaches phonemic, morphemic, and whole-word skills using a highly structured direct instruction method. Lessons are 15-20 minutes per day, and systematically review relevant skills to encourage long-term retention. The program is distributed in Australia by ACER. A review by What Works Clearinghouse found that Spelling Mastery had potentially positive effects on writing for students with learning disabilities; studies reporting findings in other domains were not included due to standards of evidence (What Works Clearinghouse, 2014). Darch, Eaves, Crowe, Simmons, and Conniff (2006) found that students taught using Spelling Mastery outperformed students taught using traditional spelling methods with moderate to large effect sizes on various tests.

3.6. Northern Territory Government Schools

There are 20 schools in the NT that are part of the FLFRPSP, 16 of which are government and four are independent schools. Of these, 19 schools run the DI program and one school runs EDI.

School selection into the program was centrally managed by the NT Department of Education, which nominated eligible schools to participate in the program. Since the program's inception, two schools have withdrawn. Based on conversations with GGSA and other key stakeholders, there appear to be 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 the Progressive Achievement Tests Reading (PAT-R). PAT-R measures reading comprehension, vocabulary knowledge and spelling, which are taught to Australian students. Assessment of comprehension covers Foundation to Year 10 and assessment of spelling ranges from Year 2 to Year 10. 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.

3.6.1. Demographic Profile

Table 6 shows the demographic information regarding FLFRPSP schools in the NT and all other participating schools. NT program schools have a lower mean ICSEA value than that of the other program schools. On average this is also a higher percentage of students in the bottom SEA quarter than other program schools. There was no reported SEA data for two NT program schools. NT program schools also have a higher percentage of Indigenous students and students with a LBOTE. All NT program schools are geographically classified as 'Very Remote'.

Table 6: Key Demographic Variables (Mean Values) for NT Program Schools

	Teachers*	Staff*	ICSEA	Students in bottom SEA quarter %	Total Students	Girls	Boys	Indig %	LBOTE%
NT PROG	8.07	7.63	597.35	91.37	101.12	51.24	49.88	98.53	93.88
Overall PROG	9.35	8.63	652.65	82.75	91.12	44.88	46.24	92.26	86.68

*full time equivalent

3.6.2. Student Level Data

Impact: Standardised Assessment

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

Jurisdictional literacy data

No literacy data (PAT-R results) was provided by the NT government to the evaluation team. Thus, standardised assessments (NAPLAN results) have been used instead to analyse impact of the FLFRPS program on student literacy outcomes.

National Assessment Program for Literacy and Numeracy (NAPLAN)

NAPLAN assessment scores formed part of five separate analyses:
- State-level cohort tracking comparing program with control schools between 2015 and 2017

- School-level average change in NAPLAN scores between 2015 and 2017

- Change in each year level NAPLAN average scores over time (see Appendix 3: NAPLAN Average Scores over Time, NT Government Program Schools)

- Percentage of students below national Minimum Standard (NMS) (see Appendix 4: Percentage of Students below NMS, NT Government Program Schools)

- NAPLAN participation rates by year and grade level (see Appendix 5: NAPLAN Participation Rates, NT Government Program Schools)

NAPLAN: Program and Control Schools

To explore change in NAPLAN over time, two approaches were used: understanding direction of growth and understanding growth differences between program and control schools.

Direction of Growth

The first approach reports the average NAPLAN score of the program schools for year 3 2015 and year 5 2017 to demonstrate the direction of growth for program and control schools, compared to national and very remote averages (Figure 6). It should be acknowledged that the national and very remote average NAPLAN scores presented as a comparison represent the same students linked from year 3 to year 5, however the averages for the program schools may not be the same students. Use of matched data for the program schools would have resulted in a significant reduction in data available for analysis. As can be seen in Figure 6, both program and control schools experienced growth. However, control schools started slightly higher and ended slightly higher than program schools, with both control and program schools falling below the national and remote schools average.

Growth Differences between Program and Control

The second approach utilised NAPLAN scores to explore the differences in growth between the program and control schools using a change score.

A series of independent t-tests were conducted to determine if program schools experienced *more* growth compared to control schools between 2015 and 2017. This was accomplished by comparing NAPLAN data from year 3 in 2015 and year 5 in 2017, for the NAPLAN domains of Reading, Writing, Spelling, and Grammar and Punctuation. Using this data, a growth variable (2017 – 2015) was created for each NAPLAN domain. Growth scores were then compared between program and control schools.

Preliminary assumption testing did detect some assumption violations, with the presence of minor outliers, however, normality was within acceptable limits, skewness/kurtosis between with ±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 analyses 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[4].

Overall, the analysis demonstrated some differences in the level of growth between control and program schools within the NT Government Schools cohort across the NAPLAN domains. Program schools experienced significantly more growth between 2015 and 2017 (M = 98.74, SD = 58.44) compared to control schools (M = 98.74, SD = 58.44) on NAPLAN spelling $t(133) = 3.89$, $p < 0.001$, Cohen's d = 0.67. However, there were no significant differences in the level of growth in reading $t(128) = 0.54$, $p = 0.59$, writing $t(132) = -0.44$, $p > 0.66$, and grammar and punctuation $t(133) = -1.22$, $p = .22$ between program and control schools.

---

[4] Where results were the same between original and imputed data, original results were reported.
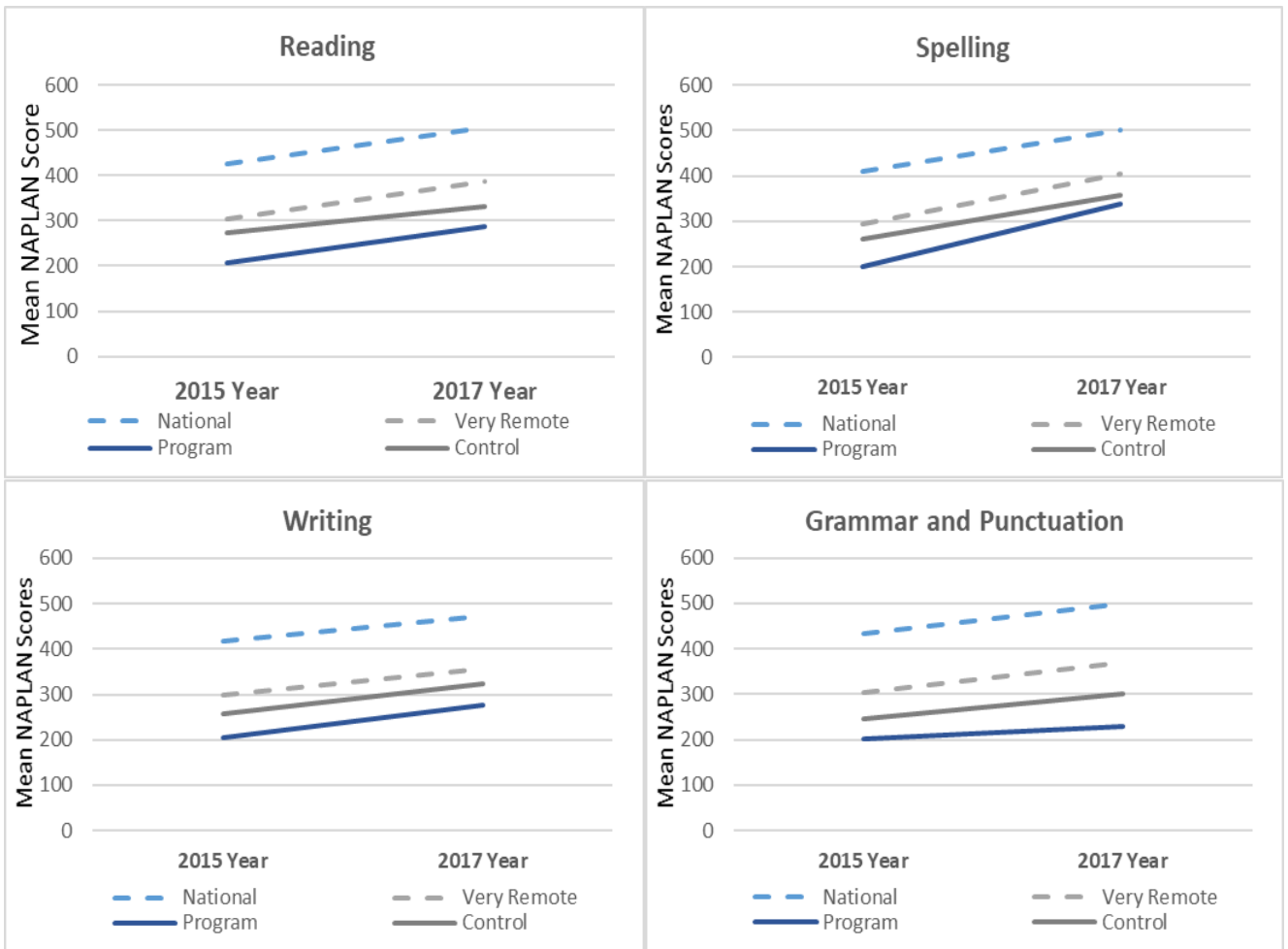
Figure 6: Change in mean NAPLAN Scores for NT government program schools alongside control schools, National Average, and Very Remote Average.

NAPLAN: Individual school analysis

At the individual school level, a NAPLAN analysis was conducted to compare the change in mean scores between Year 3 and Year 5 for the same cohort (unmatched). This was undertaken for all NAPLAN assessments and is presented in *Figure* 8. The majority of NT program schools 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 two years of program implementation for increases between Year 3 and Year 5. While some schools show a negative mean change in Grammar and Punctuation scores, it is worth noting that confidence intervals for these schools included zero—in effect, the mean changes in scores are unlikely to be significantly lower than zero.

An average effect size for all NT government program (n=13) and control schools (n=11) was also calculated based on the difference between the mean (average) change in scores between Year 3 (2015) and Year 5 (2017).

Program schools demonstrated a medium to large effect size for change over the two years in all domains (Reading: Hedge's g = 1.16; Writing: Hedge's g = 1.08; Spelling: Hedge's g = 2.94; Grammar and Punctuation: Hedge's g = 0.45). Control schools demonstrated a similar range of effect sizes (medium to large) over the two years in all domains (Reading: Hedge's g = 0.86; Writing: Hedge's g = 0.77; Spelling: Hedge's g = 2.03; Grammar and Punctuation: Hedge's g = 0.49). However, as can be seen in *Figure* 7, control schools demonstrated lower effect sizes across Reading, Writing and Spelling when compared to program schools.
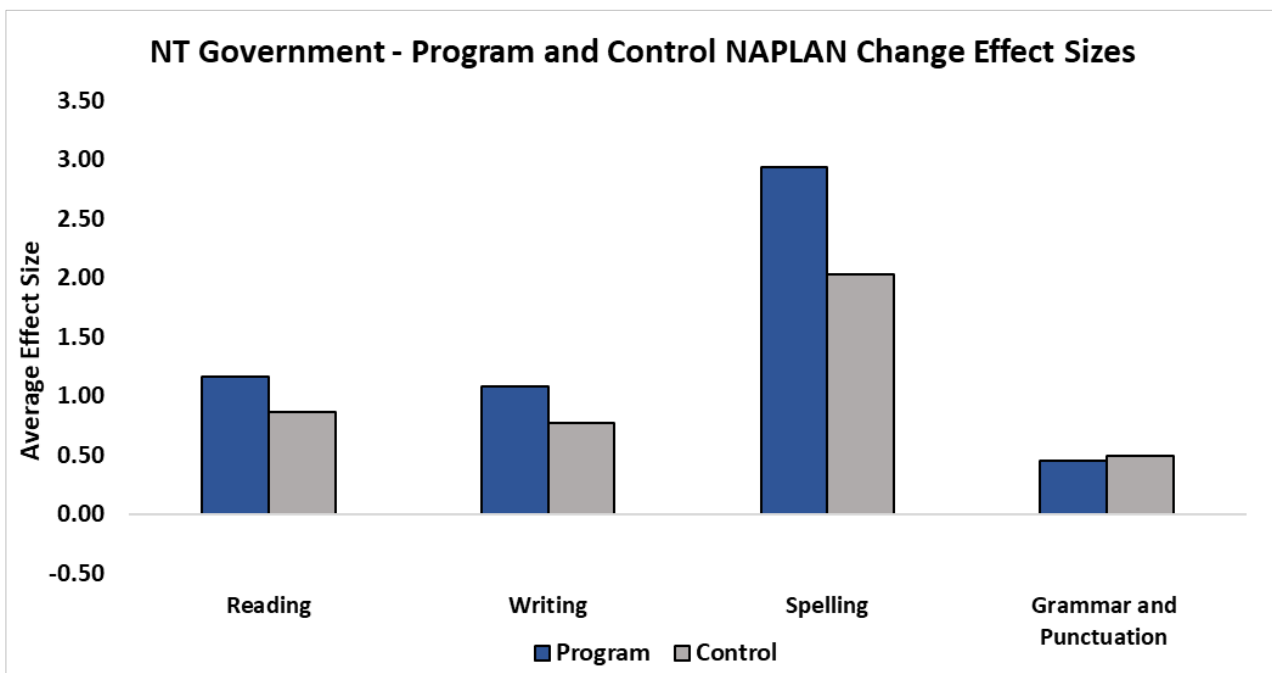
Figure 7: NT government program and control school change effect sizes.
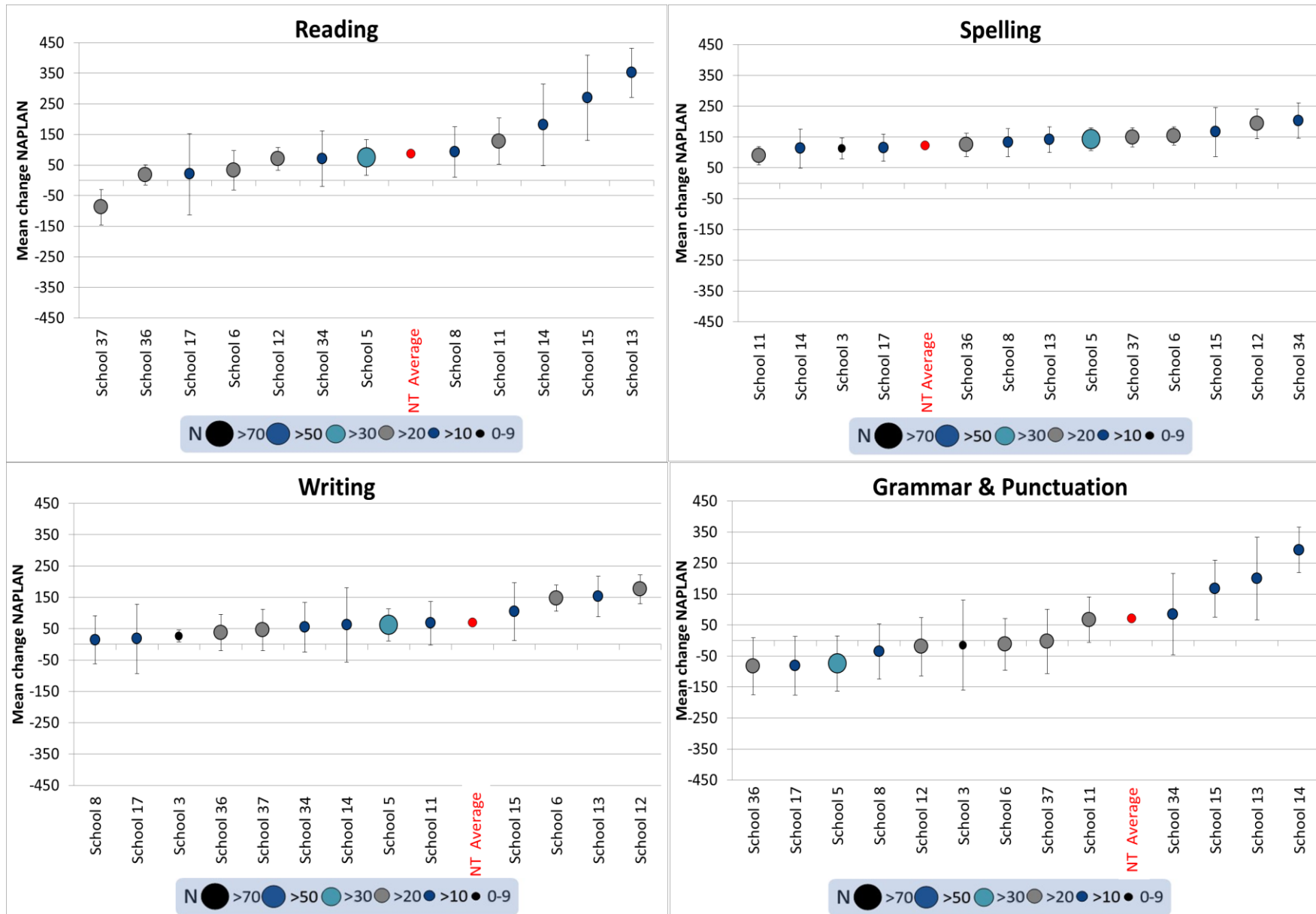
Figure 8: Individual schools' mean NAPLAN change for NT government program schools

Impact: School Staff Perceptions
School staff surveys

Overall, 31 school staff in the NT participated in the survey. The staff surveyed comprised of 27 teachers, one teaching assistant, and three principals. Staff were asked to provide their perspectives and experiences of the FLFRPSP in the context of their school and community. This section presents the results of questions relating to student outcomes.

Figure 9 presents teaching staff attitudes to student outcomes from NT government participating schools. The majority of teaching staff surveyed either strongly agreed or agreed that the program had improved students' literacy (58%) and that program had made a difference for students (55%). Fewer teachers agreed with statements about incidental changes relating to the program, such as its impact on student wellbeing (29%) and attendance (26%).

Please indicate the extent to which you agree with the following statements regarding the flexible literacy program

Statement	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
The program has improved student wellbeing	10%	19%	19%	19%	6%	10%	16%
The program has improved student attendance	13%	13%		29%	6%	19%	19%
The program has made a difference for the students	32%	23%	13%	19%		6%	6%
Students have engaged with the program	13%	39%	16%	16%	6%	3%	6%
The program has helped improve the students' literacy	42%	16%	13%	19%	3%		6%

Figure 9: Teaching staff attitudes to program's impact on students (n=31).

These findings were mirrored in the long answer survey questions, with many teachers commenting on improvements in students' literacy abilities.

> *"It works! I have seen children who would not normally be able to read starting to read."*

Other benefits mentioned included helping students develop resilience and feelings of success, increasing student enjoyment and motivation, and giving students an opportunity to practice English, especially when English was spoken as a second language. However, teaching staff emphasised the importance of context and variation in how both different students and teachers responded to the program:

> *"Our school has exited the DI program now. While using it, the approach was very successful for a number of students. A handful of students in my current [class] had returned to grade level according to PM testing [standardised literacy assessment] by the conclusion of the program. A few students for many reasons, however, under DI did not progress and of these some have found more success in another program."*

Negative aspects of the program were also acknowledged, with some teachers noting behavioural issues and boredom, especially among older students. Attendance emerged as a strong theme, with low attendance negatively impacted student progression and engagement.

School staff interviews

School staff, including both teachers (n=2) and principals (n=12), were interviewed by the evaluation team about their perception the programs impact on students' literacy and other outcomes. These interviews were conducted with the intention of sourcing evidence about the program from those practitioners directly involved in its implementation and delivery.

School staff within the Northern Territory typically perceived the program as improving the overall literacy of students, as one principal recounted:

> *"I had a student come up to me at the beginning of the term and say, 'This boring.' And at the end of the term, the same child came up to me and said, '[Name of principal], this subject is really boring. I don't like it.' And I thought you are now talking full sentences."*

Other staff felt that the program had positively influenced student learning and engagement beyond literacy:

> *"You were seeing students using more correct standard English in the classroom and students who could maintain focus a lot longer than they had previously demonstrated—because of the structure of the program."*

Despite the general improvements in student literacy reported by Northern Territory practitioners, progress was perceived as dependent on attendance, staff retention, and the schools' capacity to initially implement and sustain the program within their community context.

Relatedly, some participants felt that student improvement could not be attributed to the program:

> *"The kids who are progressing well are the ones who come to school every day and who would progress regardless [of the program]."*

3.6.3. Teachers

For the Northern Territory program schools, 43% of surveyed staff (n=30) had between 2-5 years of teaching experience while 27% had over 15 years of teaching experience. However, the majority of the teachers (84%) had only been at their current school for less than 5 years.

Implementation

As part of the program implementation, teachers and school staff were provided with program support. Program support refers to the implementation and monitoring assistance provided by GGSA, in-school coaches or program developers (NIFDI, DataWorks). The average number of training sessions per school is taken as the training data and includes behaviour, DI program and support training for teachers, teaching assistants and other leadership positions. Observation data is the number of observations per school and consists of two-minute, five-minute, and extended observations that were conducted by either an implementation manager, principal, instructional coach, teacher coach, or teaching principal. As seen in *Figure 10*, the NT program school data shows there is a relatively high number of observation sessions and training sessions across all terms.
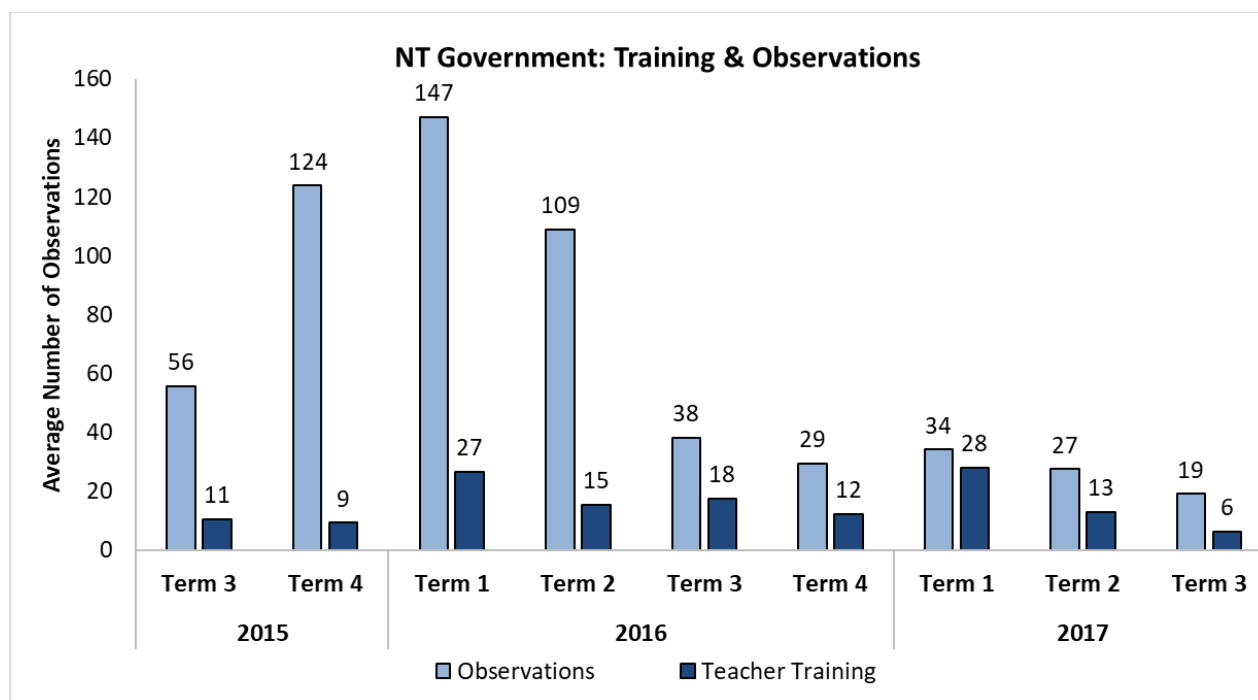
Figure 10: Average number of training sessions attended and coaching observations by term for NT government schools.

Impact: Program Data

The DI program data provides a measure of teacher effectiveness across three dimensions that include: classroom organisation, instructional delivery, and behaviour management and engagement. Classroom organisation entails recording of data, material use, and time allocation. Instructional delivery entails the use of pedagogical activities, including assessments, corrections, use of repetition, pacing, and use of signals. Behaviour management entails task transitions, teacher engagement during independent and group work, maintaining motivation, and positive behavioural support.

Figure 11 shows the development of teacher effectiveness trends over Terms 1, 3 and 4 in 2015, Terms 2, 3 and 4 in 2016, and Terms 1, 2, and 3 in 2017. Data was not available for Term 2, 2015 or Term 1, 2016 or Term 4 2017. Initial improvements were apparent in Term 4, 2015 in all three dimensions, especially classroom organisation. These dimensions decreased comparatively in subsequent terms, although rates of teacher effectiveness have stayed stable across 2016 and 2017.

Figure 11: Teacher effectiveness measures in NT program schools from DI program data.

Impact: School Staff Perceptions

School staff survey

As part of the staff survey, teaching staff were asked to rate their perception of the impact of the program on their pedagogy, role, and work as teachers. The results of these questions as discussed below.

Teaching staff were asked about how the program had impacted their pedagogy, with *Figure* 12 illustrating that there were mixed responses to questions relating to the programs impact on their teaching ability and practice.
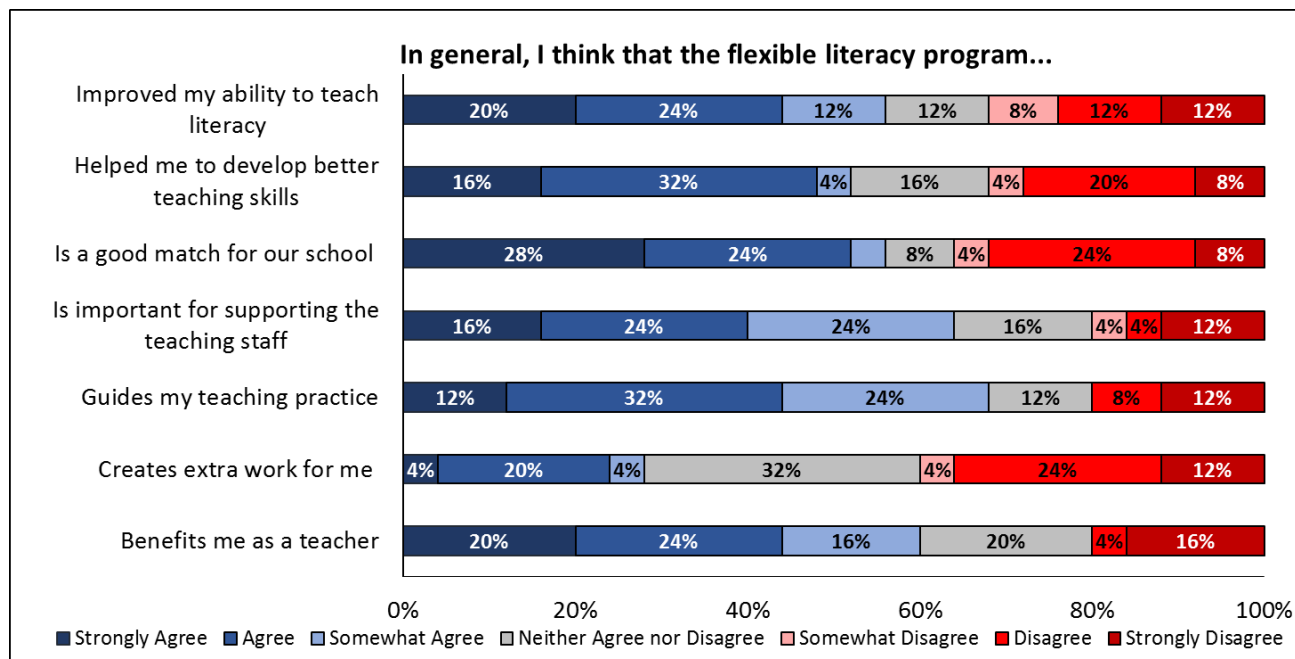
Figure 12: Teacher perception of the program's impact on their pedagogy (n=25).

In their responses to the long answer questions, some teachers indicated that the programs structure meant they spent less time on lesson preparation and planning and were more able to focus on teaching and delivery.

> *"I am no longer spending hours each week trying to decipher which skills I should be focusing on teaching for my students."*

However, some teachers mentioned that the data collection and administrative tasks within the program added to their workload without adding value to their teaching. As one teacher noted:

> *"At first I found it quite difficult to record the information where it was wanted. More time could have been used to show us how to do this. I had a lot of paper records for a while which I had no idea what to do with them."*

Figure 13 shows that almost all teaching staff surveyed strongly agreed or agreed (93%) that they have a clear understanding of the types of activities and tasks. Eighty-seven per cent also indicated (strongly agreed and agreed) that they have a strong understanding of how to teach literacy more generally.

I feel as though I have a good understanding of...

Category	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree
The specific tasks and activities students are required to do	43%	50%		3%	3%
My role within the EDI/DI program	37%	47%	13%	3%	
The pedagogical approach specific to EDI/DI	40%	40%	13%	3%	3%
How to teach literacy	37%	50%	7%	7%	

■ Strongly Agree ■ Agree ■ Somewhat Agree ■ Neither Agree nor Disagree ■ Somewhat Disagree ■ Disagree ■ Strongly Disagree

Figure 13: Teacher perception of the program's impact on their role and work as teachers (n=30).

Interestingly, they also acknowledged that the program deviated from their previous teaching practices. As commented by one teacher:

> *"It has introduced me to new sequences of teaching literacy."*

School staff interviews

Despite the perception of overall improvement of student literacy levels, the school staff interviewed had mixed opinions on the programs impact on teacher pedagogy and practice. Many participants acknowledged there is variation in how teachers and teacher's assistants within the same school perceived the program:

> *"You get the teachers who love it, and the teachers who hate it."*

While some teachers reported the structure of the program supported their teaching and preparation, the prescriptiveness of the program was seen as a drawback for many teachers, who wanted more autonomy over lesson planning. More specifically, some teachers felt that they were unable to incorporate as much project-based learning. As commented by one principal:

> *"They [teachers] cannot do normal [project-based] stuff—it is a prescriptive program."*

Relatedly, some teachers felt that the time taken to implement the program extended beyond the recommended 2.5 hours a day, which impacted their capacity to teach other content.

> *"[The program] was supposed to be 2.5 hours and rotate, but you need a break and you need to do it again. It is impossible."*

3.6.4.   School Level Data

At the school level, understanding of implementation is measured with DI program data. Program data includes measures of school fidelity as well as program support through training and observations for school staff. Perception of impact from school staff after training is also included in the previous section.

Staff Retention

As part of the program implementation, staff numbers were tracked using school reporting tools. Each semester a list of staff who were present for that semester was provided. The data was used to create estimates of staff retention across 2016-2017. A Teacher and principal retention value was created for each school, based on the number of semesters each staff member was listed across 2016 and 2017. Higher values represent higher staff retention.

The results indicated that the majority of schools (65%, n=13) retained the same principal for both 2016 and 2017, however, several (30%, n=6) changed principals at least once, with one school (5%) changing principals three times in two years. Teacher retention, as seen in *Figure 14*, varied greatly across schools with retention rates varying between 18% and 70%.
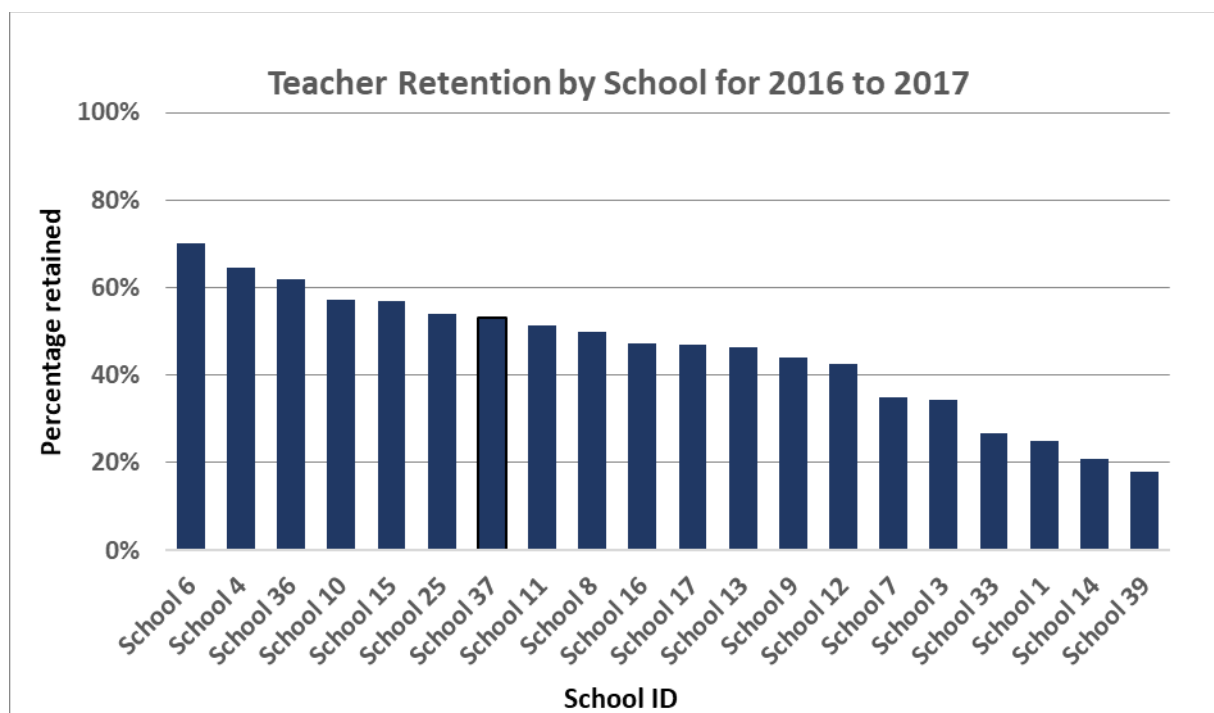
Figure 14: Teacher retention 2016 to 2017.

Implementation

School fidelity measures comprise teacher readiness, instructional leadership, classroom readiness, resource readiness and GGSA readiness, which are combined to give an overall fidelity score. A summary of each fidelity measure can be found in Appendix 6. *Figure 15* shows scores for all factors and overall school fidelity scores across 2015 (Terms 3, 4), 2016 (all terms) and 2017 (Terms 1, 2, 3). There is some variability across all four factors over time. Although there was a decline in GGSA readiness in Term 2, 2017, it increased and had the highest score across all five factors in the following term and overall was consistently high in previous years. This indicates that support and monitoring from GGSA is well implemented. Teacher readiness scored the lowest of the five factors in the end of 2015 and start of 2016. However, there was great improvement in teacher readiness in Term 3 of 2016 showing a more stable trend after this point. Generally, overall school fidelity improved in mid-2016 and has been relatively stable since.
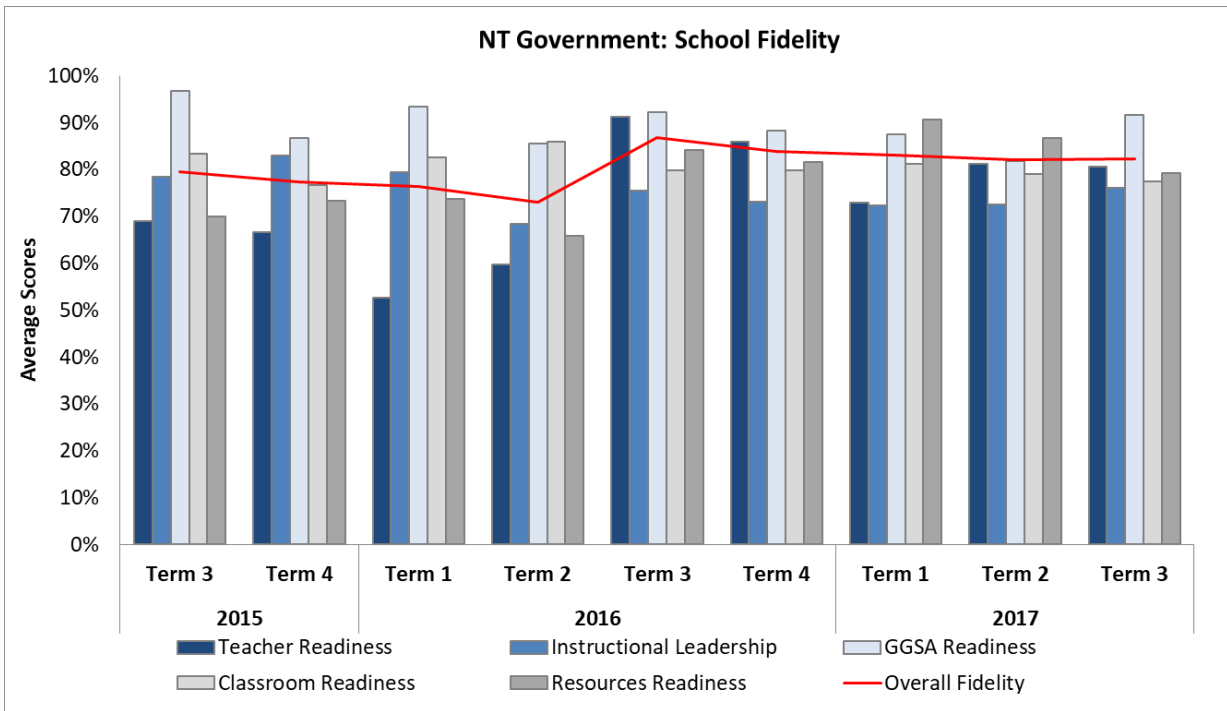
Figure 15: School fidelity measures for NT program schools, 2015-17.

Impact: School Staff Perceptions

Overall, teaching staff indicated a positive experience with the training. As shown in *Figure 16*, the majority of teaching staff surveyed strongly agreed that they felt prepared to implement the program after undertaking initial training (66%) and that they had enough resources to implement it (81%).
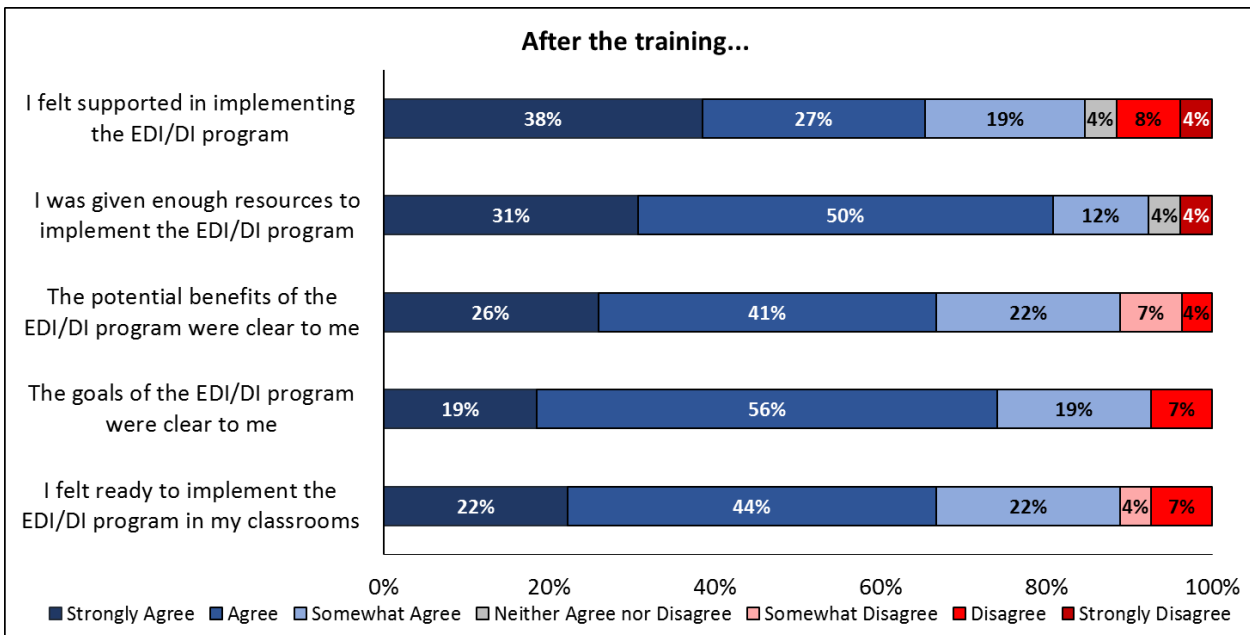
Figure 16: Teachers' perception of their understanding of the program following initial training (n=26-27)

However, in the long answer responses a small number of teachers emphasised a lack of time to get paperwork done, which was considered a barrier to them being able to implement the program with a high degree of fidelity. One teacher noted:

> *"At first I found it quite difficult to record the information where it was wanted. More time could have been used to show us how to do this. I had a lot of paper records for a while which I had no idea what to do with."*

Besides the issue of time for documentation and paperwork, there were minimum negative responses from teaching staff regarding training and implementation.

Overall, teachers felt that the whole-school approach of the program enabled staff to work more smoothly and resulted in a positive school climate where staff are consistent, students have a clear understanding of the expectations placed on them, student attendance is more consistent, and principal support is felt across the school. When asked whether they would continue with the DI/EDI program in their schools given the opportunity, the majority (55%) of the surveyed staff indicated that they would like to continue implementing the program, as shown in Figure 17 below.
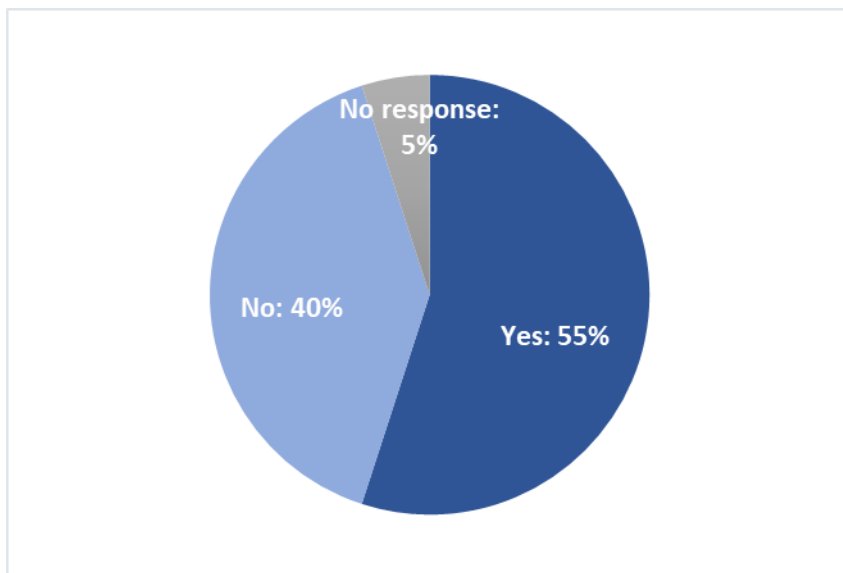
Figure 17: Staff responses to the continuation of the DI/EDI program in NT government program schools, (n=20).

Staff who wished to continue emphasised that the program had been successful within their classrooms and schools:

> "Yes! Yes! YES! It works! I have seen children who would not normally be able to read starting to read. I am a literacy-based teacher so I have invested a lot of time in the learning of different ways to teach reading. This meant if one method of teaching reading didn't work with a particular child I would implement another way hoping that this might end in success."

> "Yes, most definitely. For this school it has been a huge success. It is inspiring to see young Indigenous learners have success with English speaking and reading."

Participants who responded that they would choose not to continue, referenced a lack of engagement from students and teachers, as well as other key stakeholders within the school and community.

> "No, personally I have seen the benefits but it is not supported by the community so it is no longer viable here."

> "No as it creates a lot of behavioural issues throughout the school."

School staff interviews

Participants noted that the program demanded more teacher time and support than their standard teaching practices:

> "You need an assistant teacher there. Otherwise you have a poor teacher battling."

> "I don't think we can understate the importance in assistant teachers having an equal partnership in the program"

Relatedly, there was a strong emphasis on the importance of adequate resources to implement the program effectively and with a high degree of fidelity. As one principal noted:

> "With the program [DI], it is heavy with the reporting—daily reporting, weekly reporting—it is cumbersome. If the team is well-trained, then it is successful if it is supported heavily. The funding that came with DI we were able to do that, but when the funding stopped we were able to do that. When the funding stops, the program falls flat on its face [laughs]."

Schools also had difficulties in balancing culturally appropriate literacy and fidelity of practice often adapting the program to incorporate more Indigenous literacy elements:

> *"Ultimately, I think, in terms of allowing for cultural needs in the school, the time required for DI [to English language] limits that time [for Indigenous literacy]. I think that is ultimately a decision for our school community and our Indigenous staff at the school… These conversations need to be held with the communities as well."*

3.6.5.   Overall Summary of Collective Results for NT Government

Students

- NAPLAN spelling scores showed more growth between 2015 and 2017 for program schools compared to control schools, while all other NAPLAN data showed progression at a comparable rate.

- The majority of NT program schools demonstrated positive mean change in NAPLAN scores, with several exhibiting scores above the NT state average, depending on the relevant NAPLAN assessment.

- Across both the surveys and interviews school staff reported that the program had improved students' literacy. However, behaviour management, attendance, staff retention and other contextual factors were considered critical to its success.

Teachers

- Program data indicated that teacher effectiveness has remained stable and high across the dimensions of classroom organisation, instructional delivery, and behaviour management and engagement.

- Across both the surveys and interviews, school staff were mixed in their reports of how the program had impacted teacher practice. Some teachers felt that the program enhanced their pedagogy, whereas others found it too prescriptive. However, there was a consistent belief that teacher engagement in the program was central to it being successful.

School

- While most of schools retained the same principal across 2016 to 2017, teacher retention during this period varied between 18% and 70%.

- Surveyed staff were mixed in their desire to continue with the program, with 55% indicating they would prefer to continue to use the program, relative to 40% who would prefer to leave.

- Interview participants emphasised the that the program demanded more teacher time and support, and therefore additional resourcing.

3.7. Western Australia Government Schools

There are four WA government schools that are part of the FLFRPSP, with three schools running DI and one school running EDI. Since the inception of the program, one school has withdrawn.

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

3.7.1. Demographic Profile

Table 7 provides an overview of demographic characteristics of FLFRPSP in WA government schools and compares them with the same data for all schools in the program. WA government schools have a higher mean ICSEA value than the mean ICSEA value than the mean for all program schools. WA government schools have a higher percentage of students in the bottom SEA quarter; SEA data was not reported for two WA government schools. WA government schools have a lower percentage of Indigenous and LBOTE students. All schools are classified geographically as Very Remote.

Table 7: Mean Values for Key Demographic Variables for WA Government Schools in the Program

	Teachers*	Staff*	ICSEA	Students in bottom SEA quarter %	Total Students	Girls	Boys	Indig %	LBOTE%
WA GOV PROG	8.86	3.57	676	84.50	88.5	44.75	43.75	90.5	75
Overall PROG	9.35	8.63	652.65	82.75	91.12	44.88	46.24	92.26	86.68

*full time equivalent

3.7.2. Student Level Data

Impact: Standardised Assessment

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

Jurisdictional literacy data

No PM Benchmark data was collected by the evaluation team, standardised assessments (NAPLAN results) have been used instead to interpret the impact of the FLFRPS program.

National Assessment Program for Literacy and Numeracy (NAPLAN)

NAPLAN assessment scores formed part of five separate analyses:

- State-level cohort tracking comparing program with control schools between 2015 and 2017

- School-level average change in NAPLAN scores between 2015 and 2017

- Change in each year level NAPLAN average scores over time (see Appendix 7**:** NAPLAN Average Scores over Time, WA Government Program Schools)

- Percentage of students below National Minimum Standard (NMS) (see Appendix 8: Percentage of Students below NMS, WA Government Program Schools)

- NAPLAN participation rates by year and grade level (see Appendix 9: NAPLAN Participation Rates, WA Government Program Schools)

NAPLAN: Program and Control Schools

To explore change in NAPLAN over time, two approaches were used: understanding direction of growth and understanding growth differences between program and control schools.

Direction of Growth

The first approach reports the average NAPLAN score of the program schools for year 3 2015 and year 5 2017 to demonstrate the direction of growth for program and control schools, compared to national and very remote averages (Figure 18). It should be acknowledged that the national and very remote average NAPLAN scores presented as a comparison represent the same students linked from year 3 to year 5, however the averages for the program schools may not be the same students. Use of matched data for the program schools would have resulted in a significant reduction in data available for analysis. As can be seen in Figure 18, both program and control schools experienced growth. For spelling, and grammar and punctuation control schools started slightly higher and ended slightly higher then program schools, with both control and program schools falling below the national and remote schools average. Program schools started slightly lower than control schools in writing and finished slightly higher. Reading and writing for control and program schools also fell below the national and remote schools average.

Growth Differences between Program and Control

The second approach utilised NAPLAN scores to explore the differences in growth between the program and control schools using a change score.

A series of independent t-tests were conducted to determine if program schools experienced *more* growth compared to control schools between 2015 and 2017. This was accomplished by comparing NAPLAN data from year 3 in 2015 and year 5 in 2017, for the NAPLAN domains of Reading, Writing, Spelling, and Grammar and Punctuation. Using this data, a growth variable (2017 – 2015) was created for each NAPLAN domain. Growth scores were then compared between program and control schools.

Preliminary assumption testing did detect some assumption violations, with the presence of minor outliers, however, normality was within acceptable limits, skewness/kurtosis between with ±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 analyses 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[5].

Overall, no differences were detected in overall growth on NAPLAN between control and program schools within the WA Government School cohort between 2015 and 2017 on NAPLAN reading $t(42) = 0.77$, $p = .44$.; NAPLAN writing $t(45) = -1.33$, $p = .19$; NAPLAN spelling $t(47) = -1.60$, $p = .11$; or NAPLAN Grammar and Punctuation $t(47) = 0.03$, $p = .97$ between program and control schools. These results suggest that both the control and program schools progressed at a comparable rate for students moving from year 3 in 2015 to year 5 in 2017[6].

---

[5] Where results were the same between original and imputed data, original results were reported.
[6] Sensitivity testing confirmed all results.
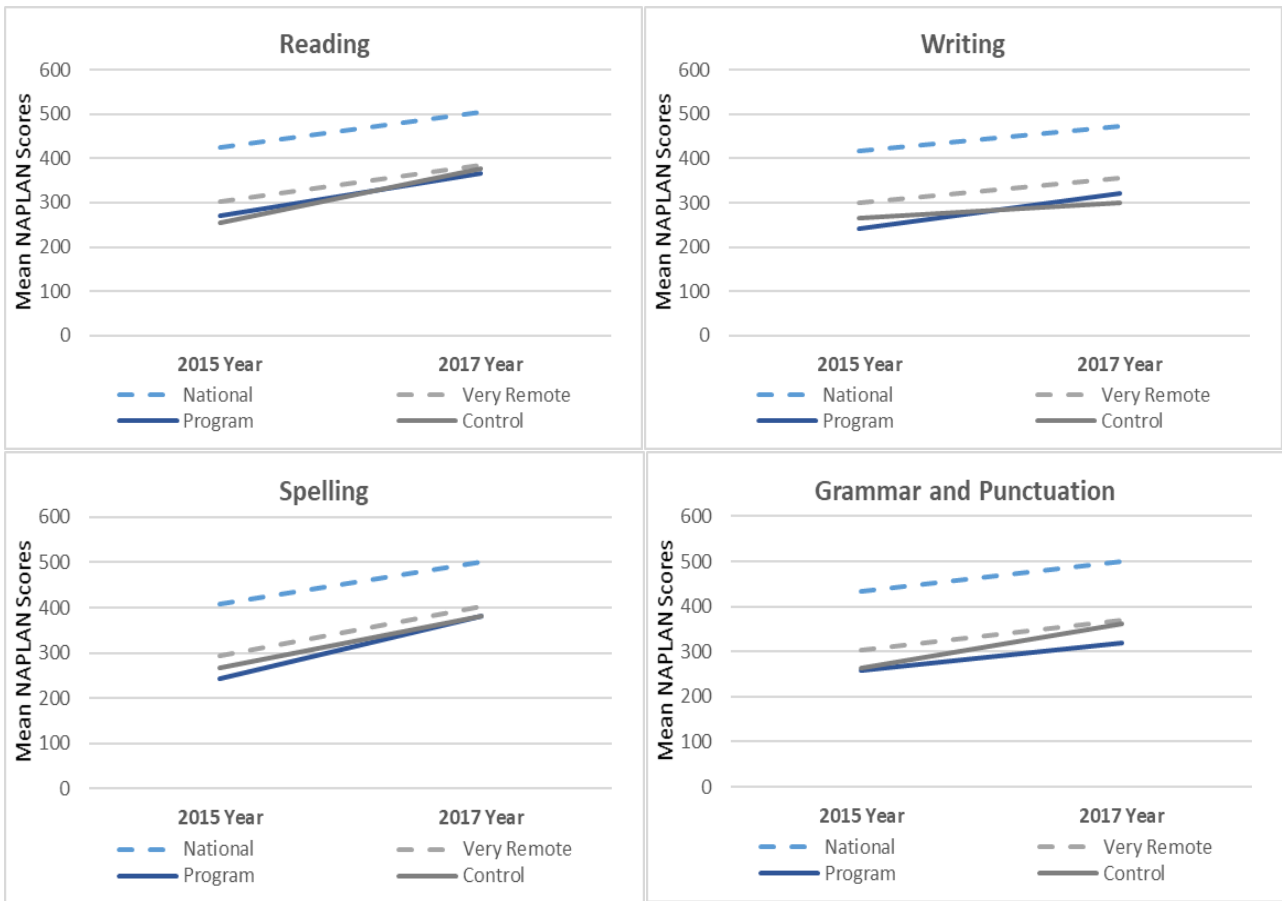
Figure 18: Change in mean NAPLAN scores for WA government program schools alongside control, National average, and Very Remote average.

NAPLAN: Individual school analysis

At the individual school level, NAPLAN analysis focused on the mean (average) change in scores between Year 3 (2015) and Year 5 (2017) for the same cohort (unmatched), as well as confidence intervals. 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 20*. A majority of government program schools in WA demonstrated positive mean change in NAPLAN scores with some exhibiting scores above the WA state average, depending on the relevant NAPLAN assessment. These results reflect expected outcomes for NAPLAN after two years of program implementation and for changes between Year 3 and Year 5.

An average effect size for all WA government program (n = 4) and control schools (n = 6)[7] was also calculated based on the difference between mean change in scores between Year 3 (2015) and Year 5 (2017).

Program schools demonstrated a medium to large effect size for change over two years (NAPLAN Reading: Hedge's g = 1.48; NAPLAN Writing: Hedge's g = 1.03; NAPLAN Spelling: Hedge's g = 1.67; NAPLAN Grammar and Punctuation: Hedge's g = 0.77). Overall, control schools demonstrated a relatively similar range of effect sizes (medium to large), over the two years in all domains, except for Writing, which was considered a small effect (Reading: Hedge's g = 1.09; Writing: Hedge's g = 0.28; Spelling: Hedge's g = 1.17; Grammar and Punctuation: Hedge's g = 0.64). Despite a similar pattern in effect sizes, as can be seen in *Figure 19*, control schools demonstrated smaller effects across all domains; however, this was particularly noticeable for Writing, which was much lower in control schools.
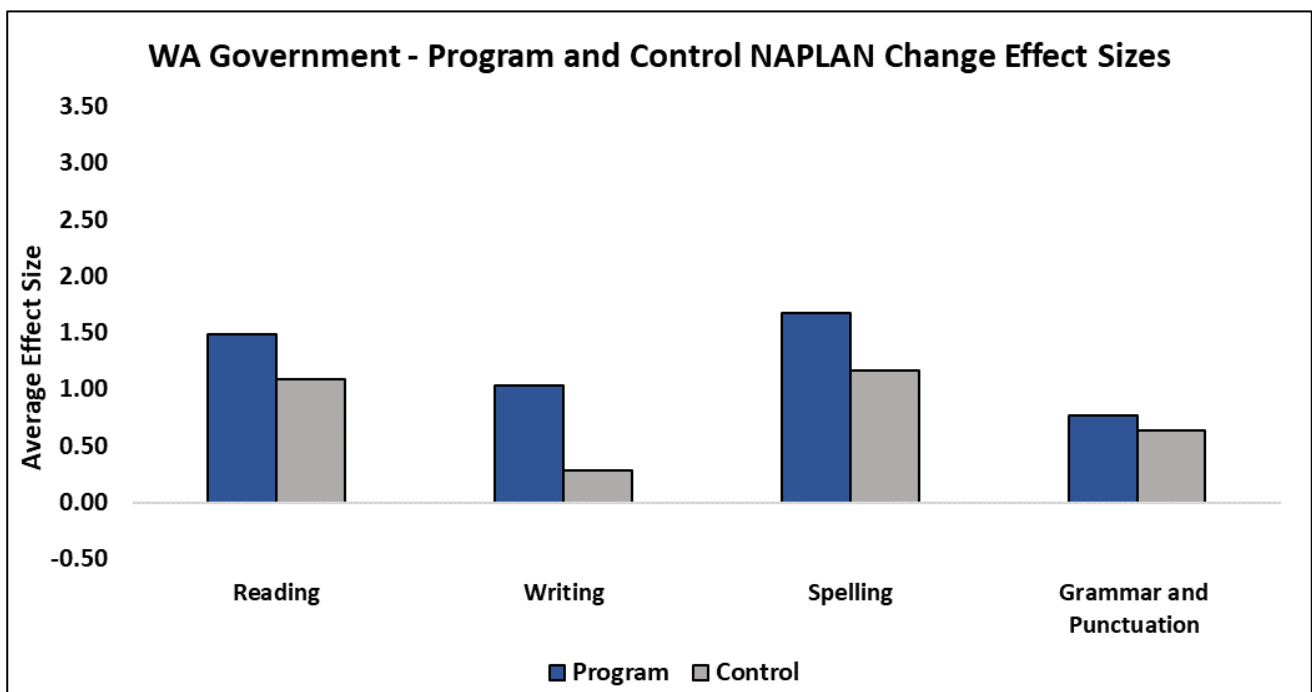
Figure 19: WA government program and control school change effect sizes.

---

[7] Differences in program and control school numbers was typically the result of school not having NAPLAN data for either 2015 or 2017.
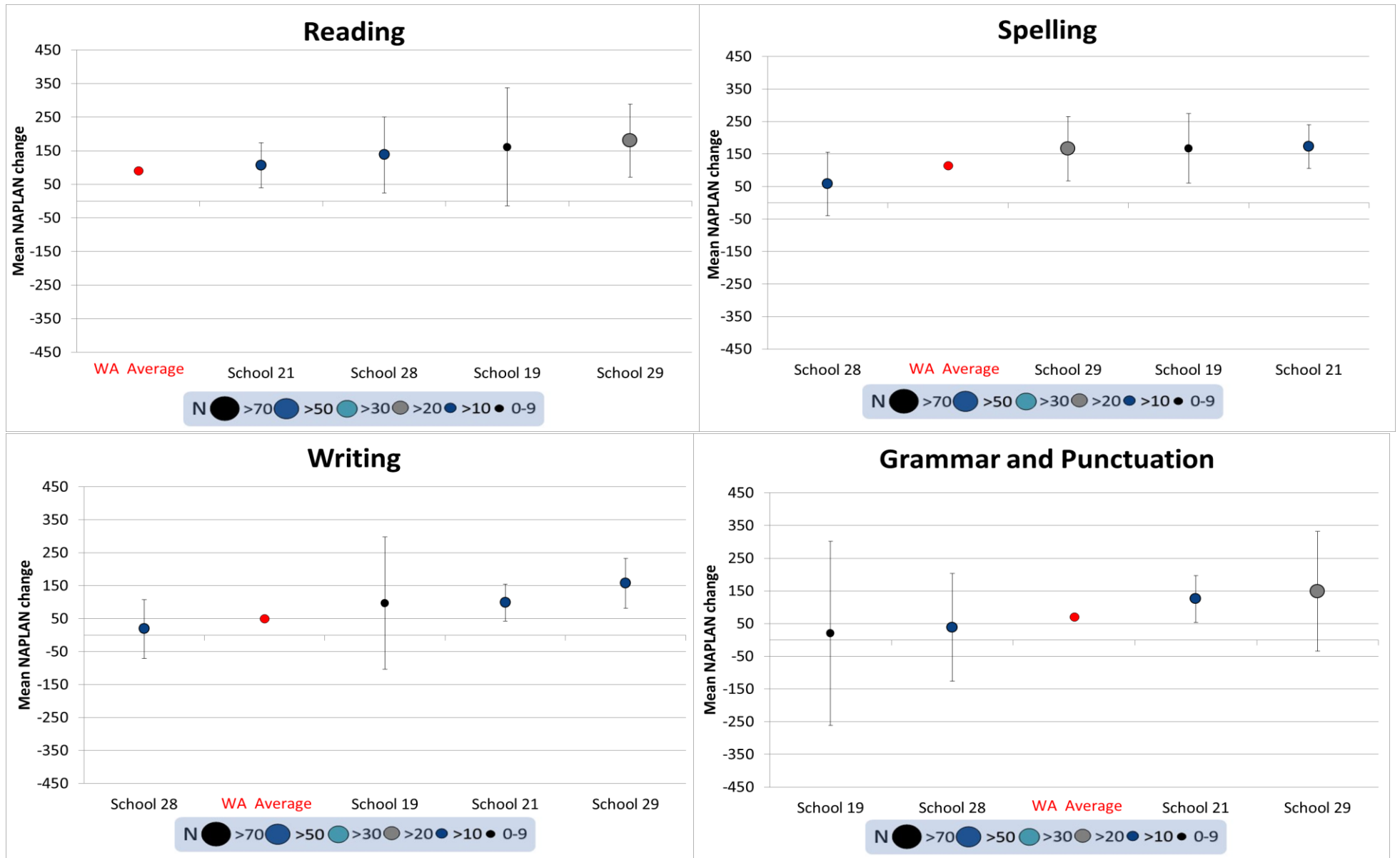
Figure 20: Individual schools' mean NAPLAN change for WA government program schools

Impact: School Staff Perception

School staff survey

Overall, nine school staff from WA government program schools participated in the survey. The staff surveyed comprised five teachers, three teaching assistants, and one other school staff member. They were asked to provide their perspectives and experiences of the FLFRPSP on student outcomes in their school. Questions in the survey considered factors such as student literacy skills, student engagement, and wellbeing, and whether the program had improved student attendance.

Figure 21 shows the results of these questions for the WA government participating schools. The majority of teaching staff surveyed agreed or strongly agreed that the program had improved students' literacy (76%) and that the program had made a difference for the students (76%).

Please indicate the extent to which you agree with the following statements regarding the flexible literacy program
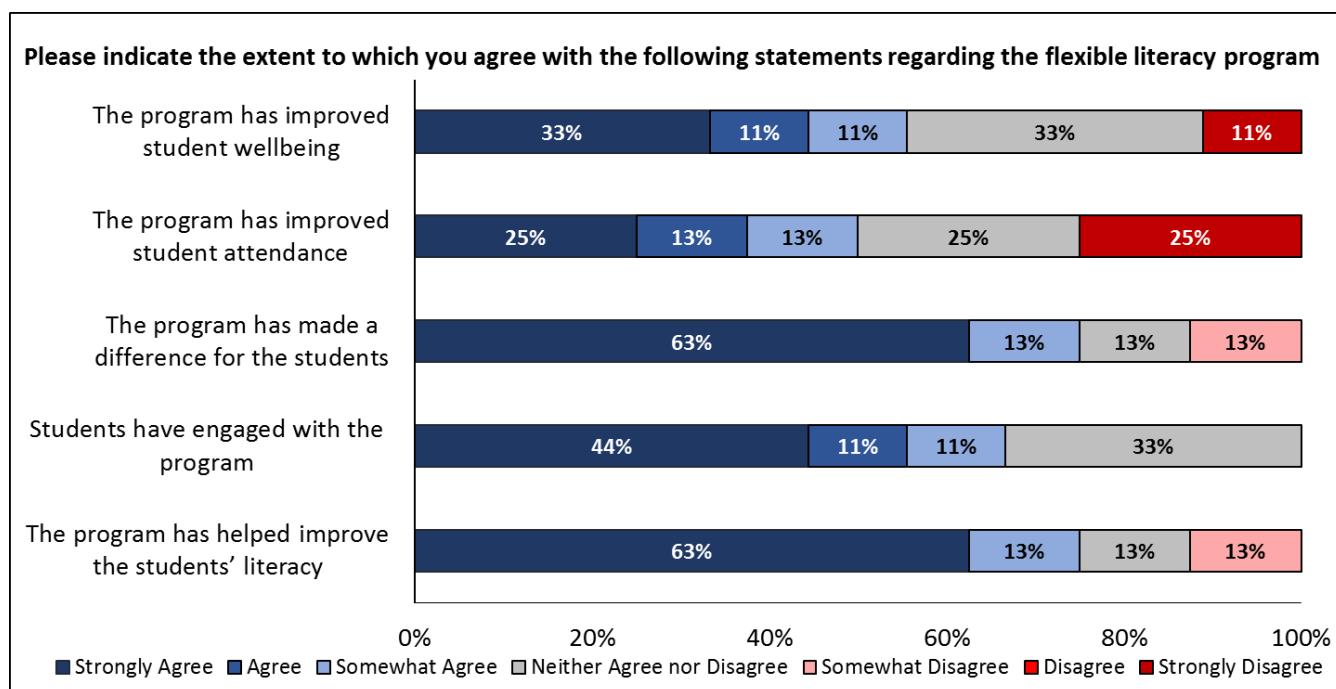
Figure 21: Teaching staff attitudes to program's impact on students (n=8-9).

These sentiments were mirrored within the long answer survey questions, which were thematically analysed. Teachers noted that the program resulted in excellent student results:

> *"I have seen excellent results in our school. I have worked at this school for many years and haven't seen a literacy program that has worked as well, producing very positive results, as DI."*

> *"It has provided a solid basis of an interesting and structured literacy program with which our children have engaged with and learned from."*

However, negative aspects of the program were also acknowledged, with some teachers noting behaviour issues and boredom amongst some students, indicating that the program was not engaging for some students:

> *"Boredom and monotony. The program really isn't that exciting for kids. They try and flip through the book to see how much longer they have to endure sitting still in one place."*

School staff interviews

Teaching staff (n=3)[8] from participating government schools in WA were interviewed by the evaluation team about their perceptions of the impact of the program on student literacy and other outcomes. These interviews were gathered with the intention of sourcing the many experiences and perspectives of school staff on the program. Due to the limited interviews at the time of writing, it is worth recognising that the findings contained in this section reflect a preliminary understanding of the WA government's implementation of the program and its perceived impacts.

Overall, interviewees felt that the program had a positive impact on student literacy, with one principal claiming

> "they have moved a whole level up after doing a year of DI."

However, participants emphasised the importance of attendance for progress in student literacy and other potential positive outcomes, such as improved motivation and attitudes towards learning:

> *"All of the students who attend all of the time have made progress, and their English skills have improved dramatically. I think that when we first started we were very much reward motivated—now kids have an intrinsic motivation - they can't wait to finish one book and then move on to the next book. And they come in here and read things to me and all that kind of stuff […] They are enthusiastic about their own progress."*

However, participants felt that students' attitudes towards the program varied, with one principal noting differences between cohorts:

> *"The early years students did not like it because it was more phonic-based than they were used to. The older children in the high school did like it."*

Students lack of engagement was also attributed to the programs approach to grouping, opposed to groupings on age:

> *"Having primary school and high school students together because of ability levels… It is quite a confused package, really."*

---

[8] Three principals.

3.7.3. Teachers

For the WA government program schools, 63% of the surveyed staff (n=8) had over 15 years of teaching experience and 25% had 2-5 years of teaching experience. However, the majority (75%) had only been at their particular school for less than five years.

Implementation

As part of the program implementation, teachers and school staff were provided with program support. Program support refers to implementation and monitoring assistance provided by GGSA, in-school coaches, and program developers support (NIFDI, DataWorks).

Figure 22 below shows the average number of observations and average teacher training for WA government schools in 2015 (Terms 3 and 4), 2016 (all terms) and 2017 (Terms 1, 2, and 3). The average number of training sessions per school is taken as the training data and includes behaviour, DI program and support training for teachers, teaching assistants, and other leadership positions. Observation data is the number of observations per school and consists of two-minute, five-minute, and extended observations that were conducted by either an implementation manager, principal, instructional coach, teacher coach, or teaching principal.

The WA government program school data shows there to be a relatively high number of observation sessions and training sessions across all terms, with a spike in the number of observations earlier in the year.
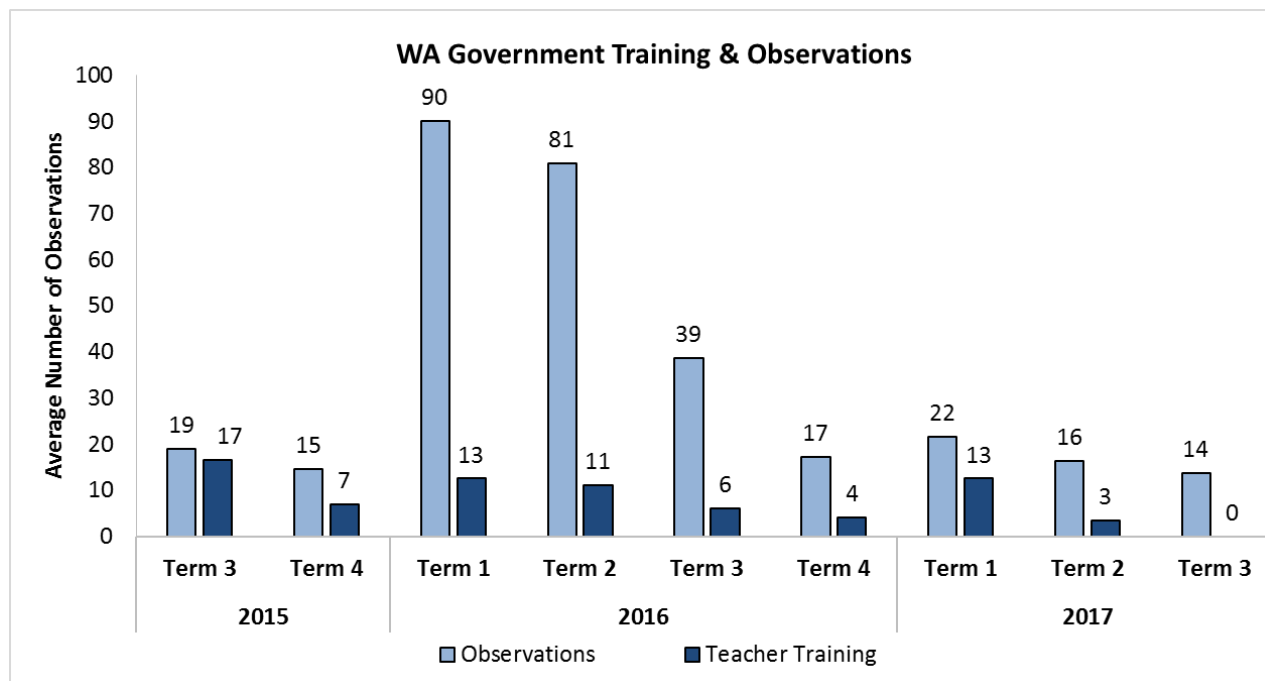
Figure 22: Average number of training sessions attended and coaching observations by term for WA government program schools.

Impact: Program Data

The DI program data provides a measure of teacher effectiveness across three dimensions that includes classroom organisation, instructional delivery, and behaviour management and engagement. Classroom organisation entails recording of data, material use, and time allocation. Instructional delivery entails the use of pedagogical activities, including assessments, corrections, use of repetition, pacing, and use of signals. Behaviour management entails task transitions, teacher engagement during independent and group work, maintaining motivation, and positive behavioural support.

Figure 23 shows the development of teacher effectiveness trends over Terms 1 and 3 in 2015, Terms 2, 3 and 4 in 2016, and Terms 1, 2, and 3 in 2017. Data was not available for Term 2 and Term 4, 2015, Term 1, 2016 or Term 4 2017. Beyond an initial dip during 2015 in all three dimensions, rates of teacher effectiveness have remained relatively stable across 2016 and 2017.
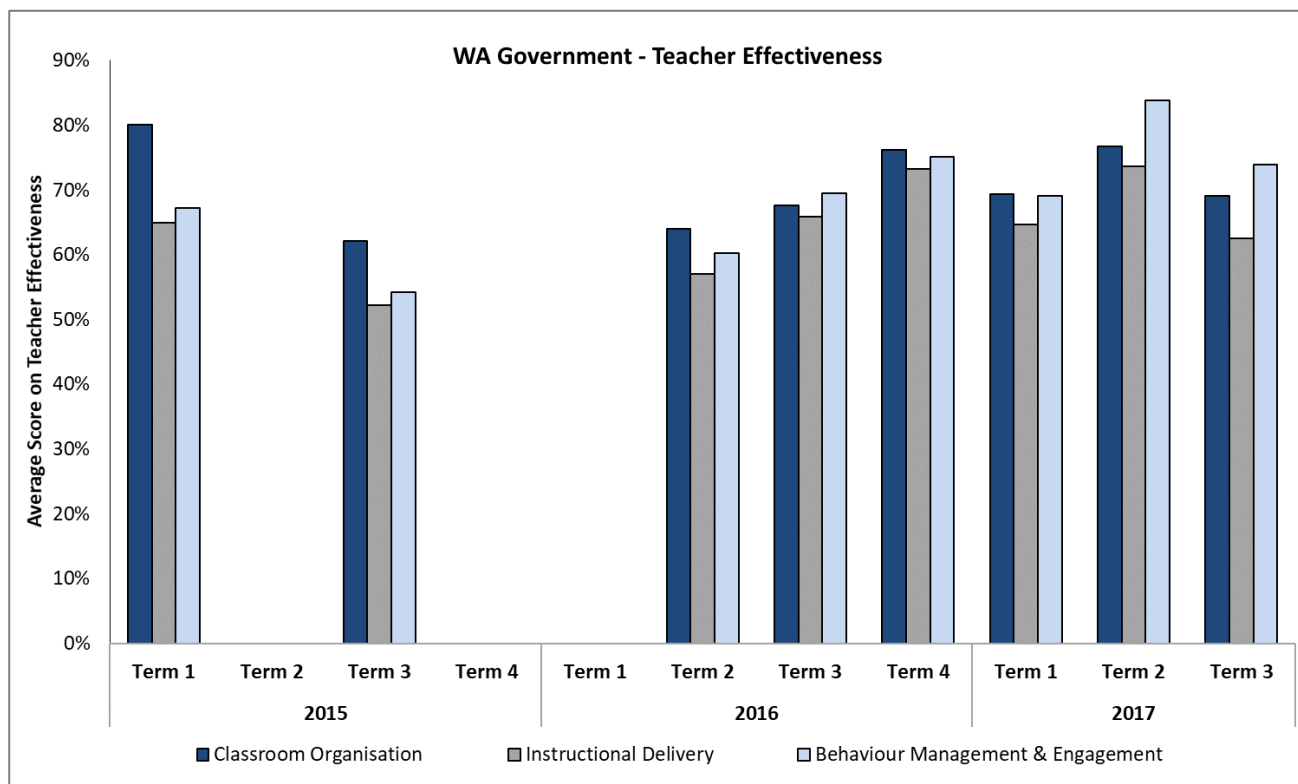
Figure 23: Teacher effectiveness measures for WA government program schools.

Impact: School Staff Perception

School staff survey

A total of nine school staff from WA government program schools participated in the survey. The staff surveyed comprised five teachers, three teaching assistants, and one other school staff member. They were asked to provide their perspectives and experiences of the FLFRPSP on their pedagogy and the impact of the program on their work as teachers.

School staff had mixed views regarding how the program had impacted their teaching practices, as shown in *Figure 24*. Most teachers (67% neither agree nor disagree) were indifferent about the program benefitting them as a teacher.

However, within the long answer responses, the importance of engaged teaching staff emerged as a strong theme:

> *"If you have a committed, hardworking, and dedicated team this program gets results. All the staff at [my school] are behind this program 100% and would love to keep moving our students forward next year."*

Although some teachers mentioned wanting additional training and support:

> *"feel that I needed more training and that it should be available during the year, maybe a mid year session for brush up. I also felt a bit let down since our coach left the program and left us needing support."*

In general, I think that the flexible literacy program...

	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
Improved my ability to teach literacy	25%	13%		25%	13%	13%	13%
Helped me to develop better teaching skills	25%	25%		25%		25%	
Is a good match for our school	50%		13%	25%	13%		
Is important for supporting the teaching staff	14%	43%	14%	14%		14%	
Guides my teaching practice	13%	25%	13%	25%	13%	13%	
Creates extra work for me	17%	17%		17%		50%	
Benefits me as a teacher	17%			67%		17%	

Figure 24: Teacher perception of the program's impact on their pedagogy (n=6-8).

However, there was a strong consensus amongst the staff surveyed that they understood the program and associated teaching strategies. As shown in *Figure 25*, most teachers surveyed felt that they had a good understanding of the program, including the tasks and activities required of students (100% agree and strongly agree), their own role within EDI/DI (100% agree/strongly agree), the pedagogical approach used in EDI/DI (100% agree and strongly agree), and how to teach literacy in general (67% agree and strongly agree).

I feel as though I have a good understanding of...

	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
The specific tasks and activities students are required to do	67%	33%					
My role within the EDI/DI program	67%	33%					
The pedagogical approach specific to EDI/DI	17%	83%					
How to teach literacy	50%	17%	17%			17%	
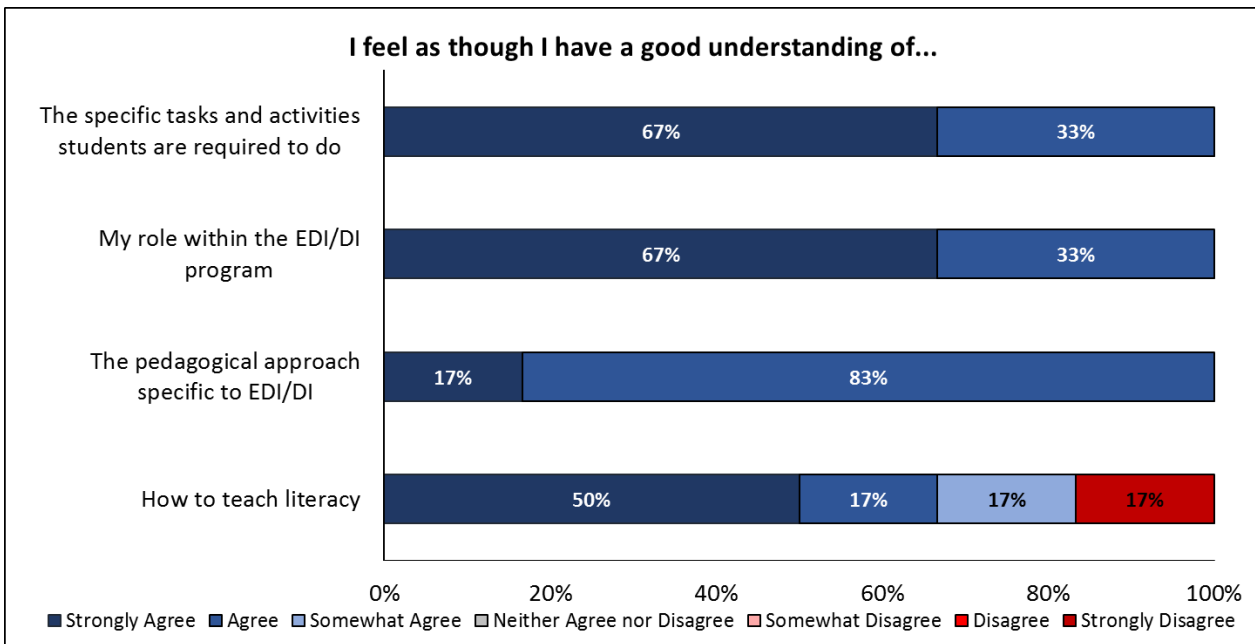
Figure 25: Teacher perception of the program's impact on their role and work as teachers (n=6).

There was also positive feedback from teachers regarding their involvement in the program. Several teachers noticed a change in their practice in general, with one teacher noting that EDI/DI had directly impacted their approach:

> *"I have become more direct in my teaching and in the way I instruct and prompt the children. I have used the student teacher game of giving points out to groups in my other lessons as well. I use the same instructions and rules that we use in DI."*

School staff interviews

As with the surveys, interview participants varied in their views on how, if at all, the program had influenced teacher pedagogy and classroom practices.

School staff typically discussed teaching changes in relation to student learning, with some explaining that they were adapting their classroom content and approaches as students developed foundational literacy skills:

> *"What's happening now - because the children have more skills and can complete more school work - they are able to engage with other aspects of the curriculum and work in more areas."*

Interview participants also noted that teachers differed in their attitudes towards the program. One principal saw the program as having a mutually beneficial relationship between student success and teacher engagement:

> *"The teachers can see the results themselves, so they have bought in and they want to pursue it."*

Contrastingly, another principal saw the program's stringent structure as problematic for some teachers, with one principal believing it disconnected teaching staff from their work:

> *"It was very prescriptive … it lends itself to a lethargic approach to teaching… It is all done for them pretty much… There are no real success criteria in it and no real checking of it. When it is done correctly there is a lot of research, a lot of background to it, but I didn't actually see that by anybody at the school at that time. […] I am not against DI, the teachers that we had at the time were not the strongest and not very personable. It was probably not in the same mode as it is done in most other schools… [Teachers] could read from the book the day before or they could read from the book as they were doing it [teaching]."*

However, there was a common belief that teacher engagement and belief in the program was critical to it being implemented effectively and with a high degree of fidelity.

3.7.4.   School Level Data

At the school level, understanding of implementation is measured with DI program data. Program data includes measure of school fidelity as well as program support through training and observations (for school staff). Perception of impact from school staff after training was included in the previous section.

Staff Retention

As part of the program implementation, staff numbers were tracked using school reporting tools. Each semester a list of staff who were present for that semester was provided. The data was used to create estimates of staff retention across 2016-2017. Teacher and principal retention values were created for each school, based on the number of semesters each staff member was listed across 2016 and 2017. Higher values represent higher staff retention.

The results indicated that the majority of schools (75%, n = 3) retained the same principal for both 2016 and 2017, with only one school (25%) changing principals across the two years. Teacher retention, as seen in *Figure* 26, varied across schools with retention rates varying between 33% and 72%.
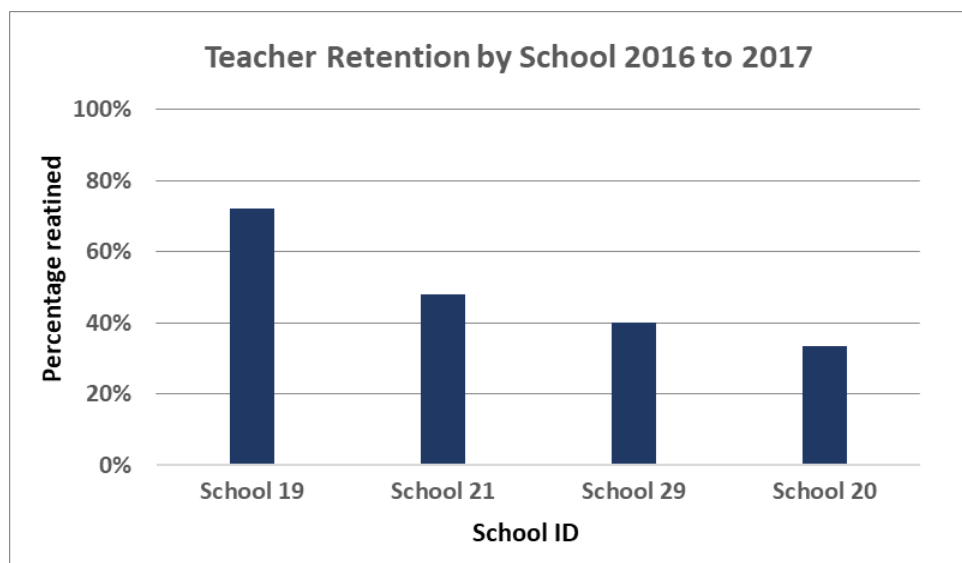
Figure 26: Staff retention 2016 to 2017.

Implementation

School fidelity measures comprise teacher readiness, instructional leadership, classroom readiness, resource readiness and GGSA readiness, which are combined to give an overall fidelity score. A summary of each fidelity measure can be found in Appendix 6. *Figure 27* shows scores for all factors and overall school fidelity scores across 2015 (Terms 3, 4), 2016 (Terms 1, 2, 3, 4) and 2017 (Terms 1, 2, 3). There is variability across all four factors over time and no obvious trend for each factor can be identified. However, the overall school fidelity showed a spike in Term 4, 2016 before declining and becoming relatively stable in 2017.
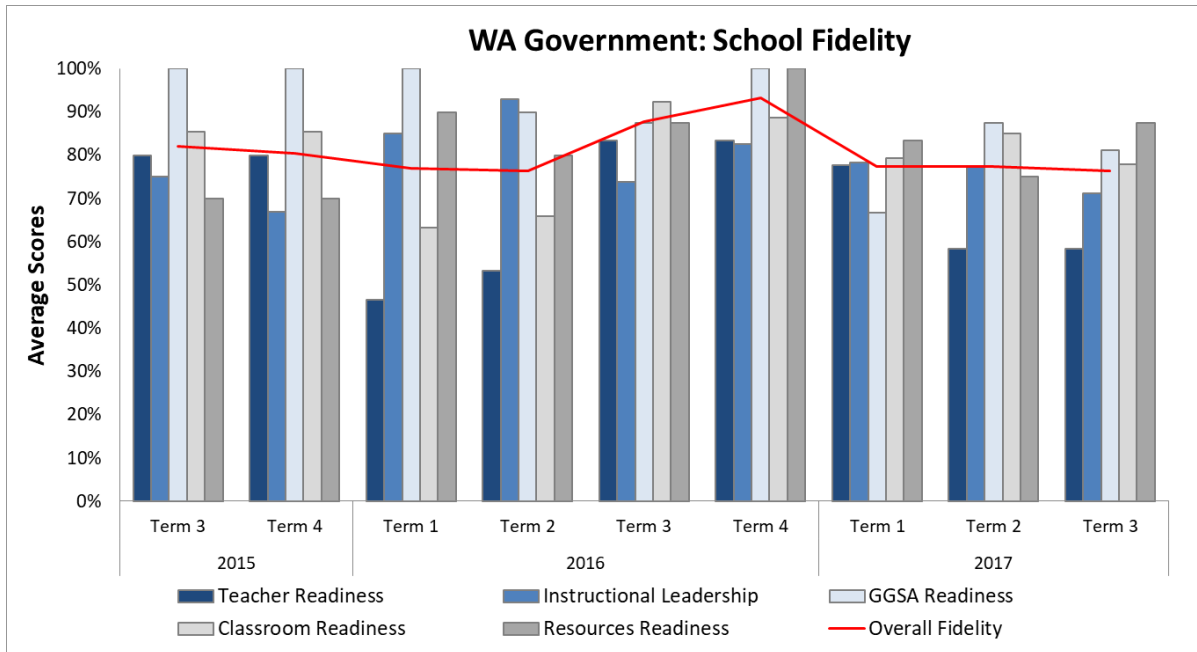
Figure 27: School fidelity measures for WA government program schools by term, 2015-2017.

Impact: School Staff Perception

School staff survey

When asked whether they would continue with the DI/EDI program in their schools given the opportunity, the majority (60%) of staff surveyed (n=5) indicated that they would like to continue implementing the program.

However, some teaching staff felt that the program was too restrictive:

> *"No, I would prefer to use the Norms in my own learning program for literacy as I am able to differentiate and plan for the individual. I would also be able to assess and report easier as I've planned the lessons."*

> *"No. It has minimal flexibility; little to no reading and writing involved and the students know they just need to repeat what the teacher say."*

Collectively, the survey results indicated that from the teacher perspective DI/EDI has resulted in a positive whole-school mindset, in which DI coordinators and school staff alike show "great enthusiasm for our kids to do well", as well as support for the program and its students. Adding to the positive school climate is "the students' realisation that they are beginning to or can read".

Staff were also asked about their experience with the training, and as shown in *Figure 28*, the majority of teachers who underwent the initial training felt that they were provided with sufficient resources to implement the program (76% agree and strongly agree). As noted by one teacher,

> *"I like the fact program resources are prepared as I can just focus on pedagogy."*

Overall, surveyed staff indicated a positive experience with the initial training, and a majority (75% agree and strongly agree) felt that they were more prepared to implement the program in their classrooms post-training. One teacher, however, felt a need for ongoing training:

> *"[I] feel that I needed more training and that it should be available during the year, maybe a mid-year session for brush-up."*
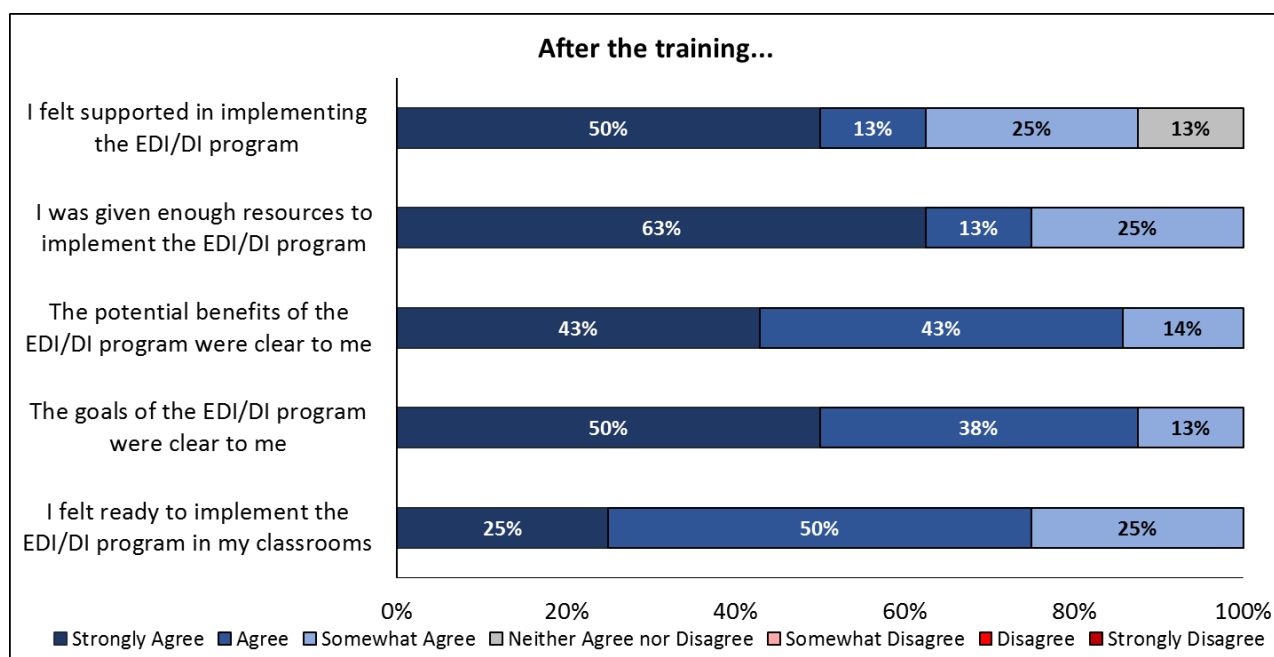
Figure 28: Teacher perception of their understanding of the program following initial training (n=7-8).

School Staff Interviews

While staff within WA government schools were largely positive about the program and its impact on student learning outcomes, they all noted barriers to its implementation and effectiveness. These issues related to the programs fit with the current approaches to teaching and learning, teacher turnover, as well as a perceived lack of program and financial support.

Participants repeatedly emphasised the importance of teacher engagement as critical to the program's success:

"It's definitely something that teachers need to buy into, otherwise it's not going to work."

Also noting that teacher turnover impacted student learning outcomes:

"Quite a few staffing issues… We found anecdotally, that since we changed the staff, our children actually fell behind."

The associated costs of implementing the program and training staff were considered to be a barrier to one principal, who stated:

"A lot of expense with it. I get the feeling as well, in the Kimberley, they were getting success but it has dropped off quite a bit."

A teaching aide, who felt overall very positively about the program, commented that additional supported was needed to more effectively embed the program within the school:

"A lot more support. We had a lot of people that quit one year. We had emails out that weren't returned. We had questions. I had a specific question about testing these new kids that are coming in. I had to figure it out myself.

3.7.5.   Overall Summary of Collective Results for WA Government Schools

Students

- No differences were detected in overall NAPLAN growth between control and program schools between 2015 and 2017, suggesting that both the control and program schools progressed at a comparable rate for students moving from Year 3 in 2015 to Year 5 in 2017.
- The majority of schools demonstrated positive mean change in NAPLAN scores, with some exhibiting scores above the WA State average, depending on the relevant NAPLAN assessment.
- Most of the teaching staff surveyed agreed or strongly agreed that the program had improved students' literacy (76%) and had made a difference for students (76%).
- Overall, the school staff interviewed felt that the program had a positive impact on student literacy, but emphasised student attendance and engagement as being critical to its success.

Teachers

- Data collected as part of the program indicates high numbers of training and observations; however, both decreased in 2017 relative to 2016.
- Program data also showed very high rates of teacher effectiveness across the domains of classroom organisation, instructional delivery, and behaviour management and engagement.
- School staff were mixed in their reports of how the program had impacted their teaching strategies, but were consistent in their belief that they understood the program and its associated pedagogy.

School

- The majority of schools retained the same principal for both 2016 and 2017, with only one school changing principals across the 2 years.
- Teacher retention varied across schools with retention rates ranging from 33% to 72%.
- Interviews with school staff revealed there to be a perception that the program's success is dependent on teacher turnover and adequate resourcing.
- School fidelity was high and relatively stable across the fidelity measures, which are comprised of teacher readiness, instructional leadership, classroom readiness, and resource readiness.

3.8. Western Australia Catholic Education Office

Seven WA Catholic education schools participate in the FLFRPSP. Currently all seven schools are running EDI, although four schools initially began with DI before switching to EDI.

Schools within the WA Catholic education jurisdiction undertake literacy assessments as part of the Early Years Literacy and Numeracy Data (EYLND), administered by the University of Western Australia (UWA) in conjunction with the WA Catholic Education Office. The assessments are designed for Years 1 to 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.

3.8.1. Demographic Profile

Table 8 summarises the relevant demographic information regarding FLFRPSP schools under the WA Catholic Education Office and compares them with the same data for all schools in the program. WA Catholic education program schools have a higher mean ICSEA value than the mean for all program schools. WA Catholic education schools have a lower percentage of students in the bottom SEA quarter; there was no data reported for one WA Catholic education School. There is also a lower percentage of Indigenous students and LBOTE students, compared with the other program schools.

Table 8: Mean Values for Key Demographic Variables for WA Catholic Education Schools in the Program

	Teachers*	Staff*	ICSEA	Students in bottom SEA quarter %**	Total Students	Girls	Boys	Indig %	LBOTE%
WA Catholic Education PROG	8.55	7.93	763	64.5	72.86	37.14	35.71	77	73.57
Overall PROG	9.35	8.63	652.65	82.75	91.12	44.88	46.24	92.26	86.68

*full time equivalent

**for WA Catholic education schools there was a large range of students in the bottom SEA quarter (28%-90%)

3.8.2. Student Level Data

WA Catholic education schools practicing the EDI program 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 29* below graphically shows the average student mastery scores for students in EDI Catholic education schools over Terms 3 and 4, 2015, all terms in 2016, and Terms 1, 2 and 3 in 2017.
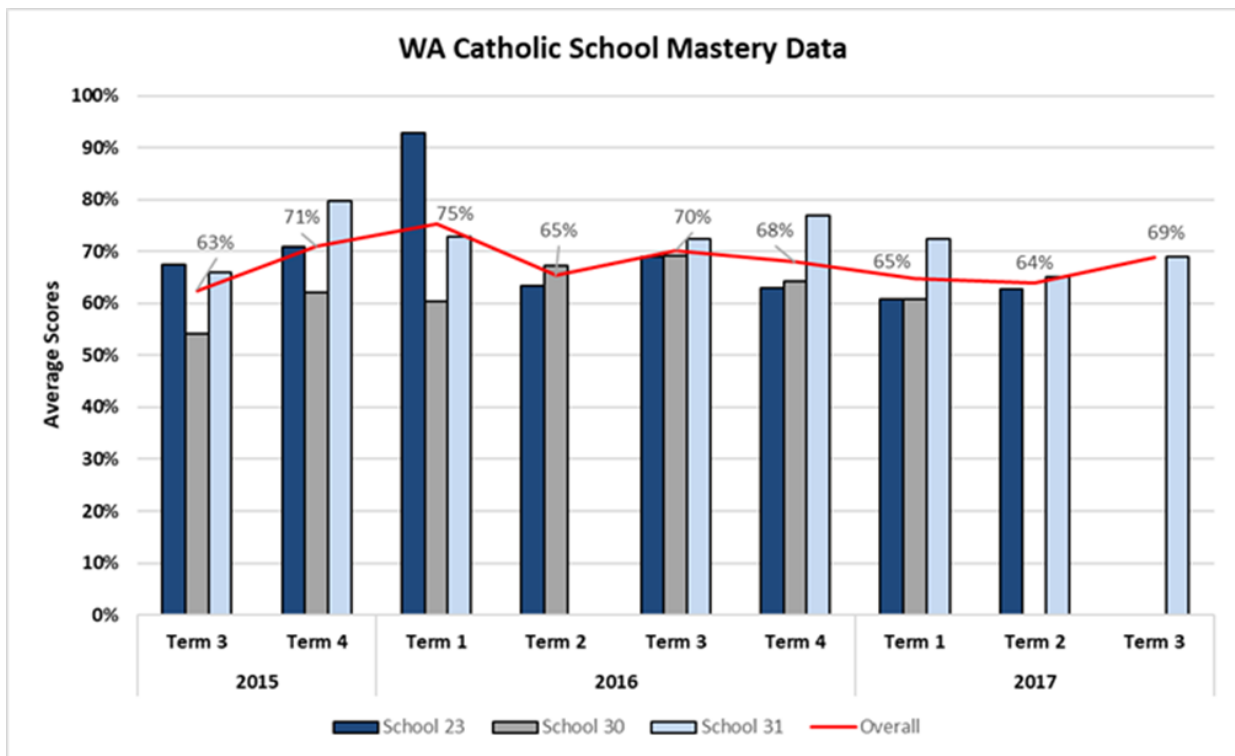
Figure 29: Student mastery scores for students in EDI Catholic education schools.

Impact: Standardised Assessment

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

Jurisdictional literacy data

To explore literacy growth over time between program and control schools, student EYLND data from all participating WA Catholic education schools, in addition to a sample of control schools, was collected and matched. This process focused on matching students who were in Year 1 in 2015, Year 2 in 2016 and Year 3 in 2017. Due to the EYLND testing timeframe, data from 2015 to 2016 reflects the 2015 year, while data from 2016 to 2017 reflects the 2016 year.

In total, 159 students were matched between 2016-2017 across 12 schools (six controls and six program). Three students were excluded from the analysis due to changing enrolment or a lack of data.

Analysis 1: Control and program differences on overall EYLND results

Prior to performing analyses, the data was inspected for missing values. A Little's Missing Completely at Random (MCAR) Test was conducted, which indicated that data was missing at random (chi-square (612) = 564.26, p=.917). As a result, an Expectation Maximisation was conducted to impute missing values[9]. Mixed ANOVA (two groups over time) 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. The analysis indicated a significant interaction effect of time and condition (i.e. program and control) (F (1.94,298.77) = 7.69, $p < .001$, η_p^2 = .048). Overall, literacy levels increased significantly in both program and control schools (main effect of time: F(1.90, 298.77) = 471.17, p < .001, ηp2 = .75). However, there was no significant difference between program and control schools at each time point. The mean difference between program schools and control schools changed from -6.07 to 14.31 for the 2015 and 2016 years. As seen in Figure 30, control schools started slightly higher than program schools in 2015; however, program schools achieved higher scores in 2016. Post-hoc testing revealed that program schools demonstrate a significant change between every time point (T1 vs T2, T2 vs T3). However, control schools have significant change between T1 and T2, but not between T2 and T3.

---

[9] A description of these procedures can be found in the Analysis Procedures Utilised section 3.3.1 page 18.
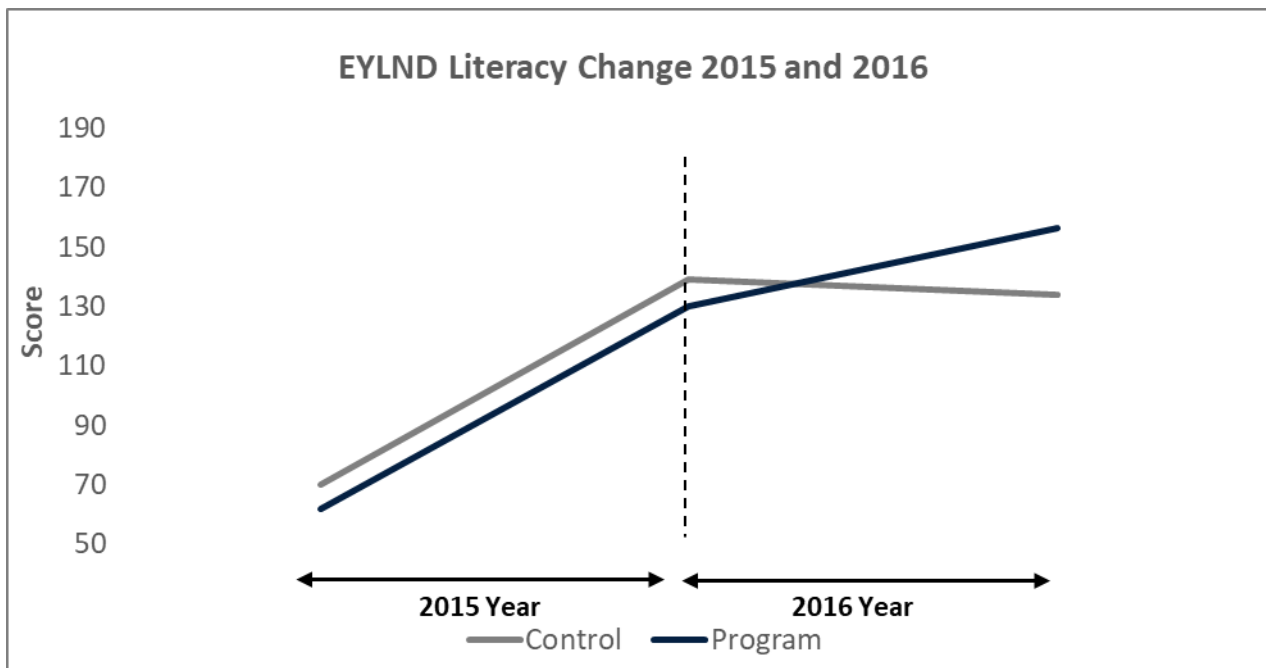
Figure 30: Change in EYLND literacy scores for program and control schools for 2016 and 2016

It is important to note that, before performing this test, a series of analyses were conducted to check whether the data passed key assumptions. No outliers were present as assessed by no greater than ±3 standard deviations. The distributions of the literacy scores were not normally distributed, as assessed by the Kolmogorov-Smirnov test of normality. However, as skewness were within ±1 and kurtosis were within ±1.5, the violation of the underlying assumption of normality was not deemed problematic, particularly given the robustness of ANOVA to this assumption breach. As the data did not meet the assumption of sphericity (p < .05), Huynh-Feldt correction is used to interpret the result. Student gender, ATSI, and LBOTE status were controlled for in this analysis.

Analysis 2: Control and Program Differences on Individual EYLND Subtests (2017)

A second series of one-way ANOVAs were performed to examine potential differences between program schools and control schools on individual EYLND subtests. Changes score were created representing the amount of change between 2015 and 2017[10]. As seen in Figure 31, program schools scored significantly higher on Burt Word Test $F(1, 154) = 7.97$, p < .01, Concepts About Print $F(1, 154) = 7.60$, p < .01, Dictation $F(1, 154) = 4.04$, p < .05, Duncan Word Test $F(1, 154) = 9.76$, p < .01, and Writing Vocabulary $F(1, 154) = 4.30$, p < .05.

---

[10] Assumption testing revealed breaches of the assumption of homogeneity of variances for Burt Word Test and Duncan Word Test; thus, the result of Welch's test is used instead of the original p value of F test for these subscales.

Differences Between Control and Program Schools on EYLND Subtests Score Changes 2015-2017
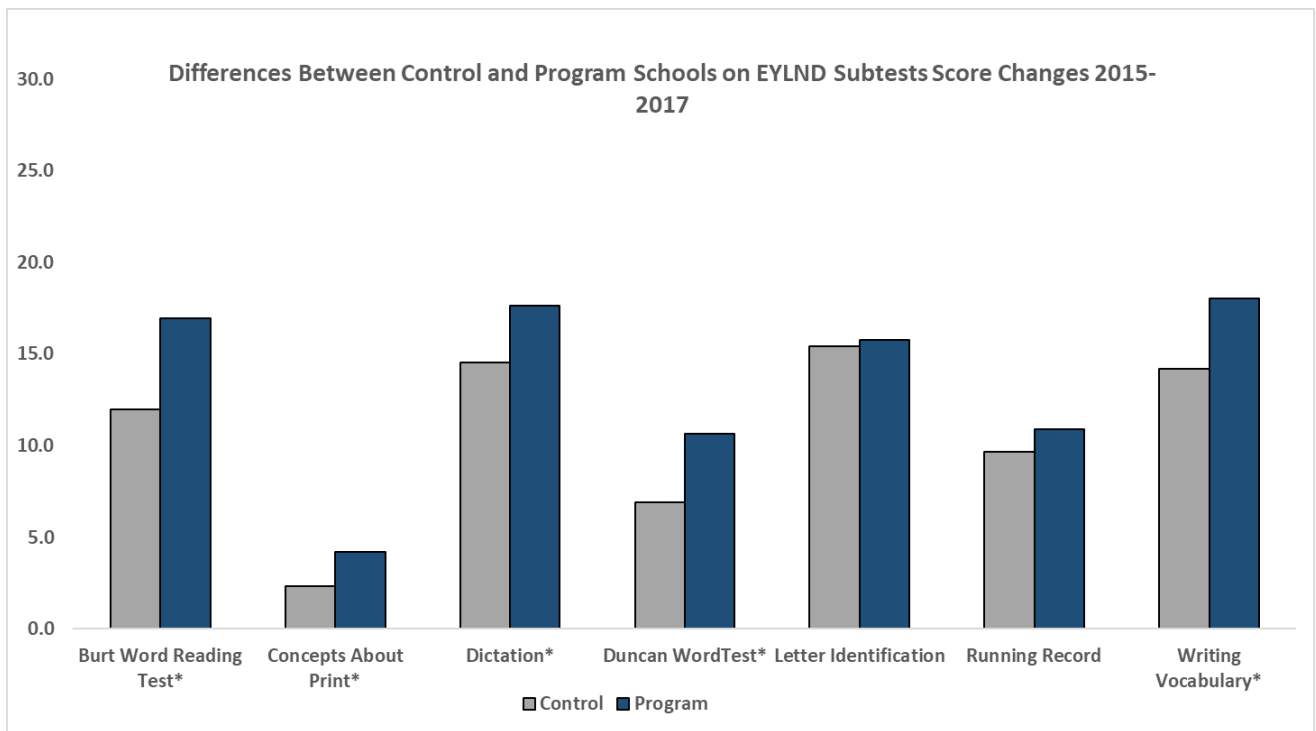
*Significant difference detected, p < .05

Figure 31: Difference between control and program on EYLND subscales 2017.

An average effects size for all WA Catholic education schools (n = 6)[11] was also calculated based on the difference between on the mean (average) change in scores between 2015 and 2017. As can be seen in *Table 9* schools demonstrated a large effect size for change over two years.

Table 9: EYLND Effect Sizes by School

School	EYLND Hedge's G 2015-2017
School 18	0.97
School 23	1.60
School 24	1.53
School 26	4.36
School 30	2.29
School 32	2.39

National Assessment Program for Literacy and Numeracy (NAPLAN)

NAPLAN assessment scores formed part of five separate analyses:

- State-level cohort tracking comparing program with control schools between 2015 and 2017

- School-level average change in NAPLAN scores between 2015 and 2017

- Change in each year level NAPLAN average scores over time (see Appendix 10)

- Percentage of students below national Minimum Standard (NMS) (see Appendix 11)

- NAPLAN participation rates by year and grade level (see Appendix 12)

---

[11] One school did not have EYLND results.

NAPLAN: Program and Control Schools

To explore change in NAPLAN over time, two approaches were used: understanding direction of growth and understanding growth differences between program and control schools.

Direction of Growth

The first approach reports the average NAPLAN score of the program schools for year 3 2015 and year 5 2017 to demonstrate the direction of growth for program and control schools, compared to national and very remote averages (Figure 32). It should be acknowledged that the national and very remote average NAPLAN scores presented as a comparison represent the same students linked from year 3 to year 5, however the averages for the program schools may not be the same students. Use of matched data for the program schools would have resulted in a significant reduction in data available for analysis. As can be seen in Figure 32, both program and control schools experienced growth. However, program schools started slightly higher and ended slightly higher then control schools, with both control and program schools falling below the national average and above the remote school average.

Growth Differences between Program and Control

The second approach utilised NAPLAN scores to explore the differences in growth between the program and control schools using a change score.

A series of independent t-tests were conducted to determine if program schools experienced more growth compared to control schools between 2015 and 2017. This was accomplished by comparing NAPLAN data from year 3 in 2015 and year 5 in 2017, for the NAPLAN domains of Reading, Writing, Spelling, and Grammar and Punctuation. Using this data, a growth variable (2017 – 2015) was created for each NAPLAN domain. Growth scores were then compared between program and control schools.

Preliminary assumption testing did detect some assumption violations, with the presence of minor outliers, however, normality was within acceptable limits, skewness/kurtosis between with ±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 was 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[12].

Overall, no differences were detected in overall growth on NAPLAN between control and program schools within the WA Catholic Schools cohort between 2015 and 2017 on NAPLAN reading $t(60) = 1.16$, $p = .25$; NAPLAN writing $t(64) = -2.16$, $p = .03$; NAPLAN spelling $t(64) = -1.25$, $p = .22$; or NAPLAN Grammar and Punctuation $t(64) = -.146$, $p = .15$ between program and control schools. These results suggest that both the control and program schools progressed at a comparable rate for students moving from year 3 in 2015 to year 5 in 2017[13].

---

[12] Where results were the same between original and imputed data, original results were reported.
[13] Sensitivity testing confirmed all results.
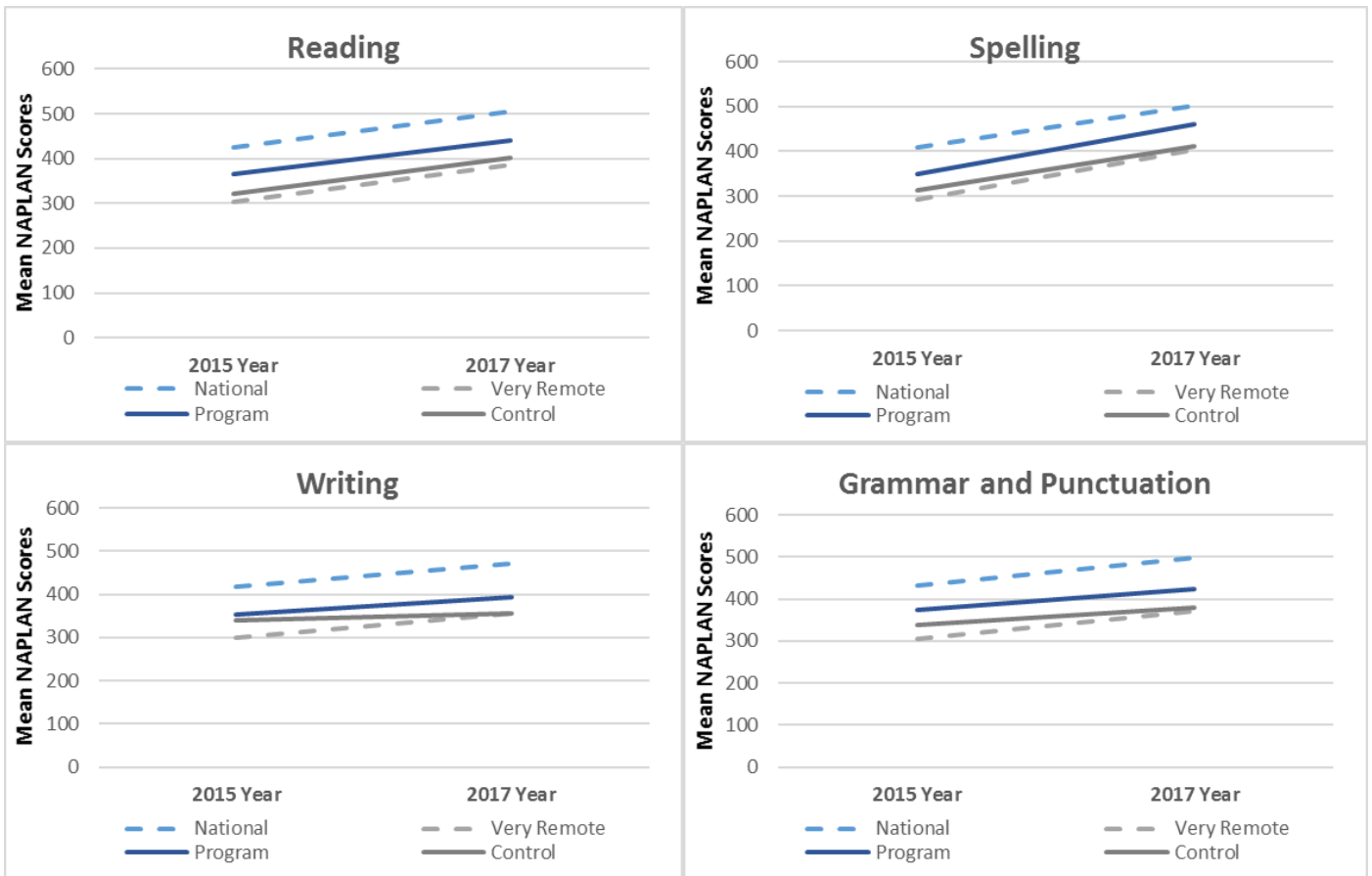
Figure 32: Change in mean NAPLAN scores for WA Catholic education program schools alongside control, National Average, and Very Remote average

NAPLAN: Individual School Analysis

At the individual school level, NAPLAN analysis focused on mean (average) change in scores between Year 3 (2014) and Year 5 (2017) for the same cohort (unmatched), as well as confidence intervals. 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 34* (below). A majority of WA Catholic schools demonstrated a positive mean change in NAPLAN scores with some exhibiting scores above the WA State average.

It should be noted that many of the schools that have approximately the same level or lower mean change in scores compared to the WA State average also have large error margins; this occurs as a result of small samples sizes and large variability among the students.

An average effects size for all WA Catholic education program (n = 7) and control schools (n = 4)[14] was also calculated based on the difference between on the mean (average) change in scores between Year 3 (2015) and Year 5 (2017).

Program schools demonstrated a medium to large effect size for change over two years (NAPLAN Reading: Hedge's g = 0.85; NAPLAN Writing: Hedge's g = 0.74; NAPLAN Spelling: Hedge's g = 1.72; NAPLAN Grammar and Punctuation: Hedge's g = 0.42). Overall, control schools demonstrated a much lower range of effects sizes (small to large) over the two years in all domains, particularly Writing and Grammar and Punctuation, which both had small negative effect sizes (Hedge's g = 0.63; Writing: Hedge's g = -0.15; Spelling: Hedge's g = 1.08; Grammar and Punctuation: Hedge's g = -0.11). This pattern is further demonstrated in *Figure 33.*
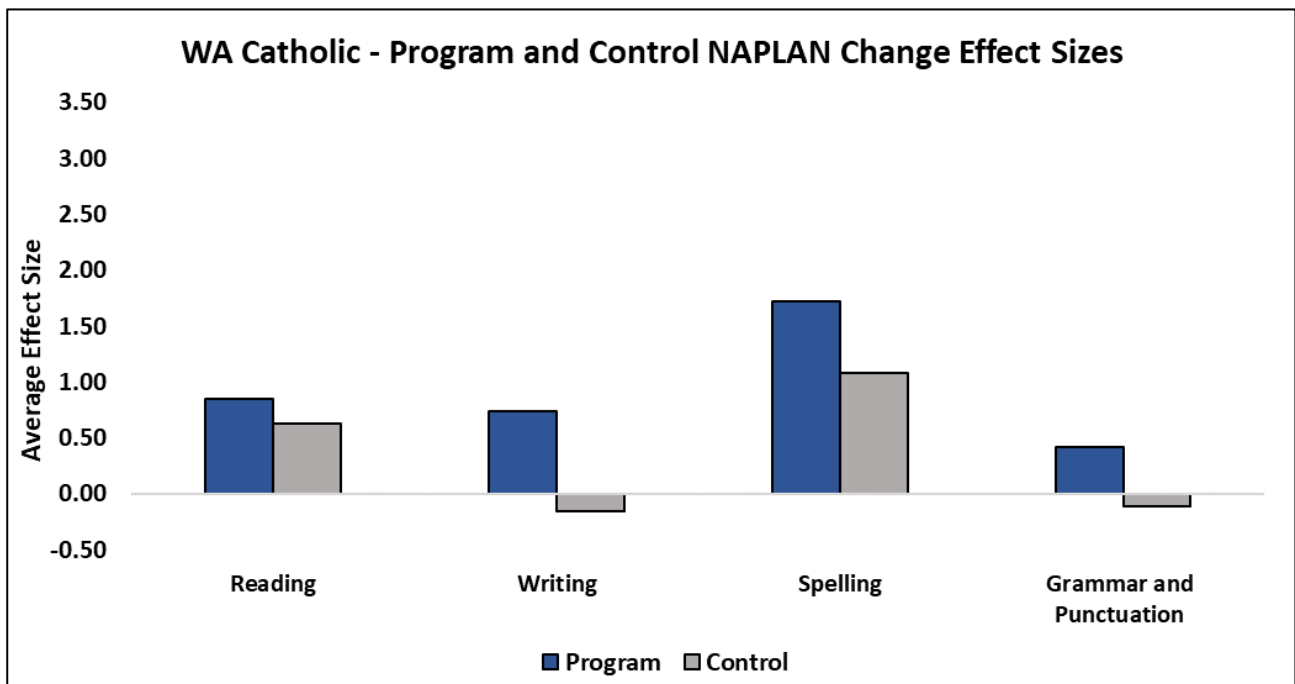
Figure 33: WA Catholic program and control school change effect sizes.

---

[14] Differences in program and control school numbers was typically the result of schools not having NAPLAN data for either 2015 or 2017.
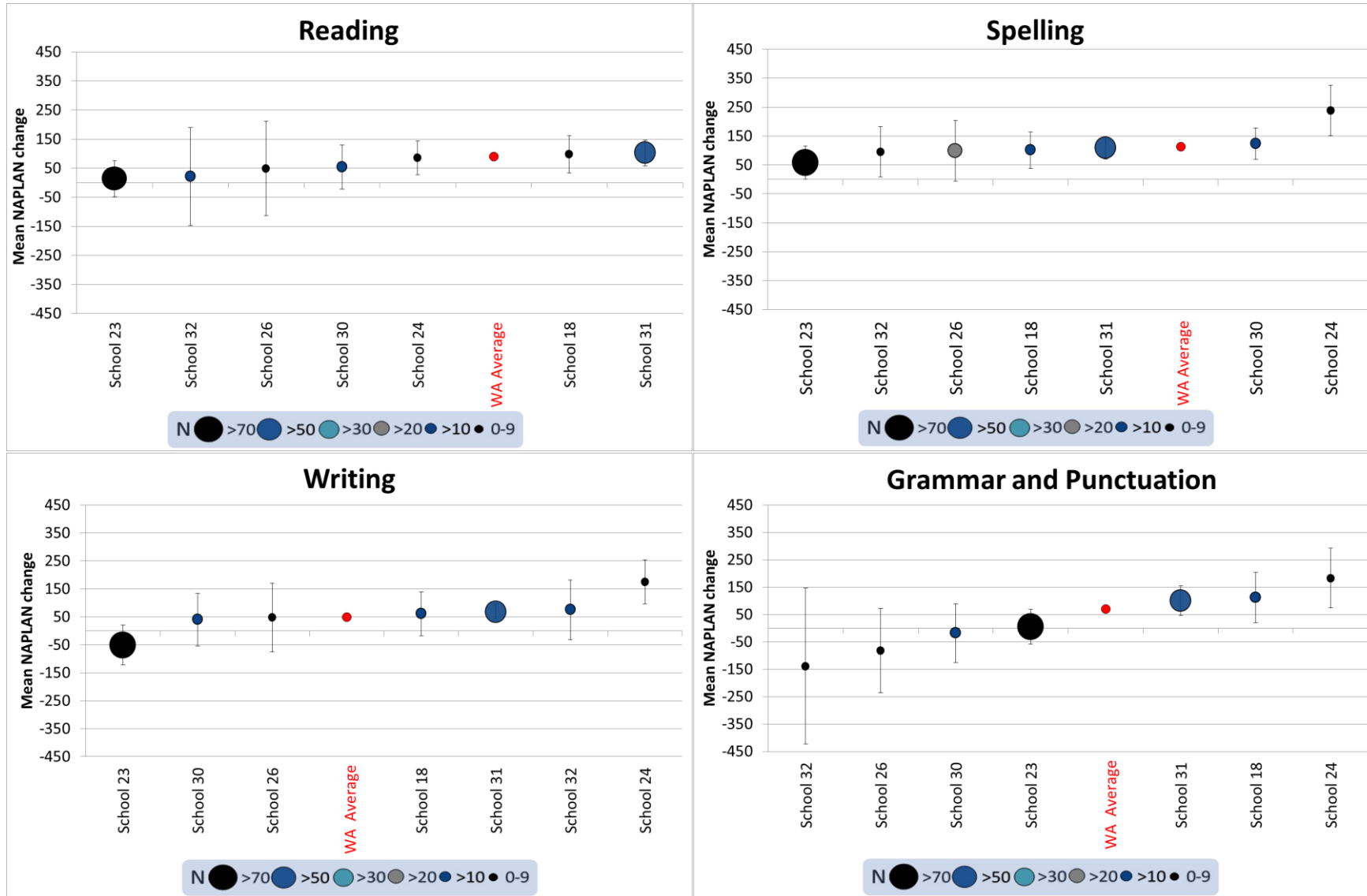
Figure 34:WA Catholic education schools mean NAPLAN change Year 3 (2015) to Year 5 (2017).

Impact: School Staff Perceptions

School staff surveys

Overall, 21 school staff from WA Catholic education program schools participated in the survey. The staff surveyed comprised of 12 teachers, four teaching assistants, four principals, and one other member of school staff. They were asked to provide their perspectives of and experiences with the FLFRPSP.

As shown in *Figure 35*, the majority of teaching staff surveyed agreed or strongly agreed that the program had improved students' literacy (85%) and made a difference for the students (85%). However, questions regarding the program's impact on wellbeing and student attendance generated more mixed responses.

Please indicate the extent to which you agree with the following statements regarding the flexible literacy program

Statement	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
The program has improved student wellbeing	20%	20%	20%	15%	10%	15%	
The program has improved student attendance		5%	30%	35%	10%	20%	
The program has made a difference for the students	60%	25%			5%	5%	5%
Students have engaged with the program	40%	35%	15%			5%	5%
The program has helped improve the students' literacy	55%	30%			5%	5%	5%

Figure 35: Teaching staff attitudes to program's impact on students (n=20).

Within the long answer responses, perspectives of the program's impact on students were largely positive, with teachers often discussing positive outcomes beyond the programs primary goal of improving student literacy:

> *"Children talking more about their learning - children can express what they have learnt and often why they have learnt it."*

School staff interviews

School staff (n=7)[15] from participating WA Catholic schools were interviewed by the evaluation team to better understand the experiences and perspectives of key stakeholders directly involved in the program's implementation. Overall, WA Catholic schools reported improvements in students' literacy, particularly in relation to the use of standard English and spelling.

> *"Achievement in literacy has skyrocketed in comparison to prior to EDI."*

> *"The data doesn't lie—our literacy levels have gone through the roof."*

A number of participants credited these learning gains to the programs structure, with one principal noting the program allows them to "focus more on content than process" because the "students know the process".

Similarly, one teacher felt that the programs structure acted as a protective factor from disruptions in their learning environment, such as teacher turnover:

> *"DI creates some familiarity. The kids are able to cope with the changes."*

---

[15] 4 principals and 3 teachers.

As with other jurisdictions, teaching staff spoke about how improvements in student literacy led to other academic gains and increased student confidence:

> "We have children say, 'I want to do some work now' or 'I am spelling now'."

3.8.3. Teachers

For the WA Catholic education program schools, 33% of the surveyed staff (n=21) had over 15 years of teaching experience and 20% had 5-10 years of teaching experience. However, of these 21 staff surveyed, 33% and 23% had only been at their particular school for less than 2 years and between 2-5 years respectively.

Implementation

As part of the program implementation, teachers and school staff were provided with program support. Program support refers to implementation and monitoring assistance provided by GGSA, in school-coaches and program developers support (NIFDI, DataWorks).

The average number of observations in WA Catholic schools was at its highest in Term 1, 2016, with the most observations typically occurring at the start of each year (see Figure 36). The highest number of teacher training sessions was in 2016 and then decreasing in 2017.
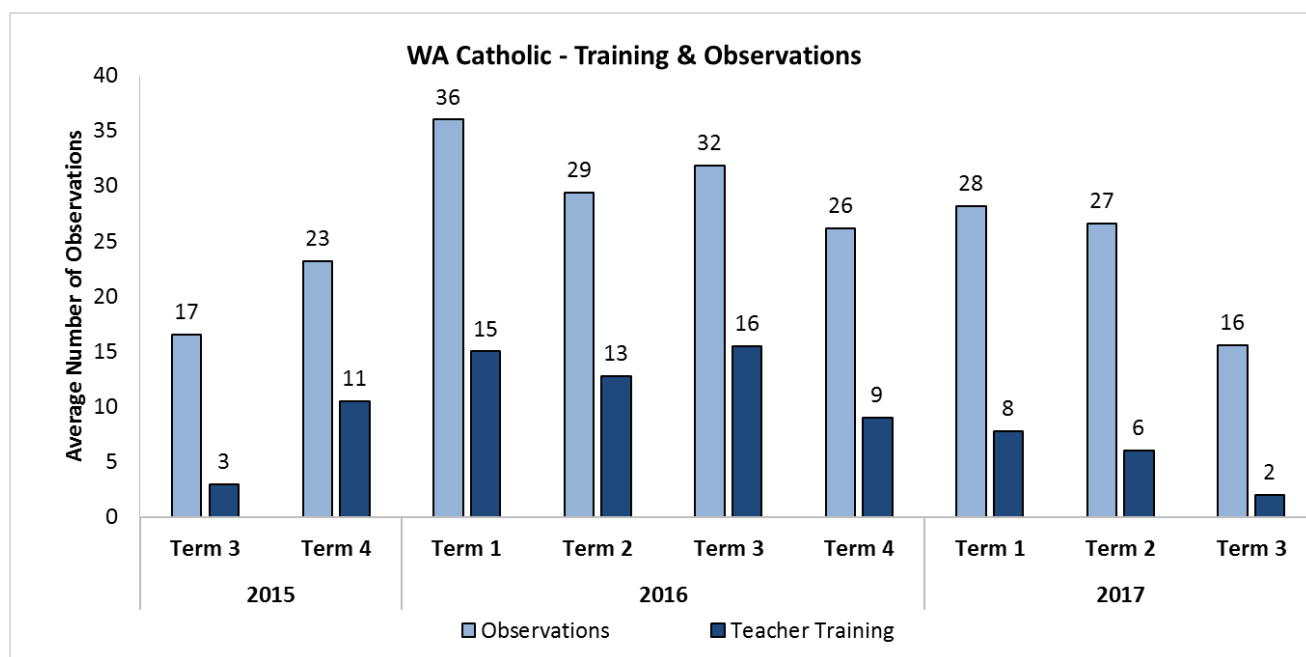
Figure 36: Average number of training sessions attended and coaching observations by term for WA Catholic education program schools.

Impact: Program Data

The program data provides a measure of teacher effectiveness across three dimensions, which include classroom organisation, instructional delivery, and behaviour management and engagement. Classroom organisation entails recording of data, material use, and time allocation. Instructional delivery entails the use of pedagogical activities, including assessments, corrections, use of repetition, pacing, and use of signals. Behaviour management entails task transitions, teacher engagement during independent and group work, maintaining motivation, and positive behavioural support.

Figure 37 shows data for teacher effectiveness in WA Catholic schools, which was only collected starting Term 2, 2016. Within this data, scores of teacher effectiveness rose initially in Term 3 and Term 4, 2016, and while these scores have fluctuated across all three measure of classroom organisation, instructional delivery, and behaviour management and engagement, effectiveness scores have remained higher than their initial measurement in Term 2, 2016. Classroom organisation was at its highest, around 85% in Term 2, 2017, dropping again to around 70% in Term 3. Instructional delivery has remained at or above 70% since Term 3, 2016. Behaviour management and engagement has remained consistently high, with scores around 80% since Term 3, 2017.
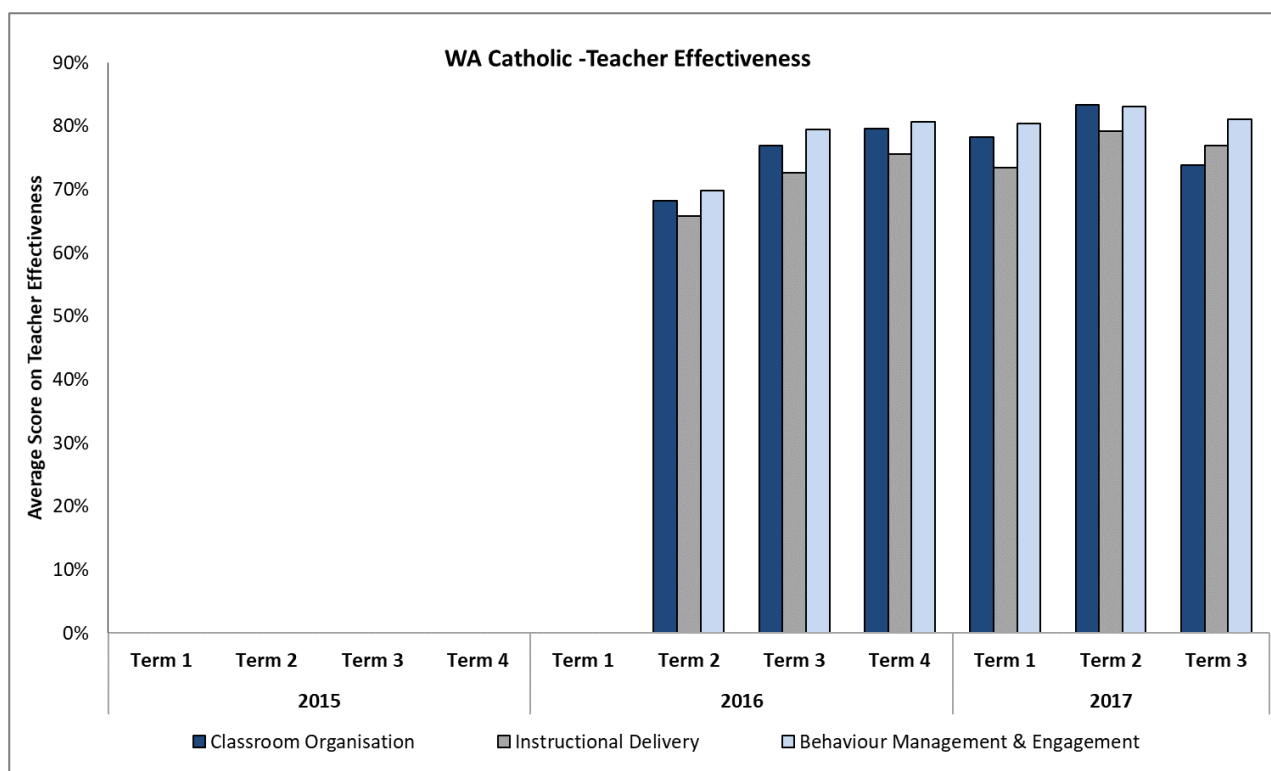
Figure 37: Teacher effectiveness measures for WA Catholic education program schools.

Impact: School Staff Perceptions

School staff survey

School staff were also asked to reflect on how the program had impacted their pedagogy and work as teachers more generally. Overwhelmingly, the staff surveyed believed that the program benefited them as teachers (78%) and improved their ability to teach literacy (71%) (see *Figure 38).*

In general, I think that the flexible literacy program...

Statement	Strongly Agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
Improved my ability to teach literacy	50%	21%	7%	7%		14%	
Helped me to develop better teaching skills	64%	14%	7%	7%		7%	
Is a good match for our school	71%	21%				7%	
Is important for supporting the teaching staff	57%	21%	14%			7%	
Guides my teaching practice	50%	29%	14%			7%	
Creates extra work for me	7%	14%		21%	7%	14%	36%
Benefits me as a teacher	57%	21%	7%	7%		7%	

Figure 38: School staff perception of the program's impact on their pedagogy (n=14).

Surveyed staff repeatedly mentioned the importance staff engagement and commitment to the program:

> *"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."*

Surveyed staff also reported having a solid understanding of the program and its associated teaching practices, as shown in *Figure 39* on the following page. They also emphasised that the program supported their teaching overall, and that the structure meant they spent less time on lesson preparation and planning.

> *"The program has been written for you and it ensures the curriculum is covered. I don't have to worry about planning, just teaching!"*

I feel as though I have a good understanding of...

Statement	Strongly Agree	Agree	Somewhat agree	Neither agree nor disagree
The specific tasks and activities students are required to do	47%	47%	6%	
My role within the EDI/DI program	56%	38%	6%	
The pedagogical approach specific to EDI/DI	47%	35%	12%	6%
How to teach literacy	50%	44%	6%	

Legend: ■ Strongly Agree ■ Agree ■ Somewhat agree ■ Neither agree nor disagree ■ Somewhat disagree ■ Disagree ■ Strongly disagree

Figure 39: Teacher perception of the program's impact on their role and work as teachers (n=16-17).

School staff interviews

Many interview participants felt that the program had changed the approaches to teaching and learning within their classrooms and schools. Some interviewees claimed that the pedagogical principles embedded within the program had become a part of their school's approach to teaching and learning more generally.

> *"[We have integrated these strategies] in pretty much every other area. We start every lesson with a learning objective, we use those engagement norms. We do that throughout every lesson of the day."*

One principal mentioned that there was now more consistency in the approaches of teachers within their school:

> *"Previously teachers were doing different things, not entirely different pedagogy, but often there was discrepancy between them- and now there is consistency between teachers and across year levels, making it easier for students to transition between different grades."*

Teaching staff also explained how they were learning as part of the program and that implementing the program with skill took time:

> *"We are talking about habits that have been developed over many, many years. So, that can take a little while to change and when you are having to constantly think about what you are doing, that can be quite taxing, but once it becomes automated it becomes very rewarding too."*

3.8.4.   School Level Data

At the school level, understanding of implementation is measured with DI program data. Program data includes measure of school fidelity as well as program support through training and observations (for school staff). Perception of impact from school staff after training was included in the previous section.

Staff Retention

As part of the program implementation, staff numbers were tracked using school reporting tools. Each semester a list of staff who were present for that semester was provided. The data was used to create estimates of staff retention across 2016-2017. Teacher and principal retention values were created for each school, based on the number of semesters each staff member was listed across 2016 and 2017. Higher values represent higher staff retention.

The results indicated that the majority of school (71%, n = 5) retained the same principal for both 2016 and 2017, however, two schools (29%) changing principals at least once in two years. Teacher retention, as seen in *Figure 40*, varied across schools with retention rates varying between 58% and 28%.
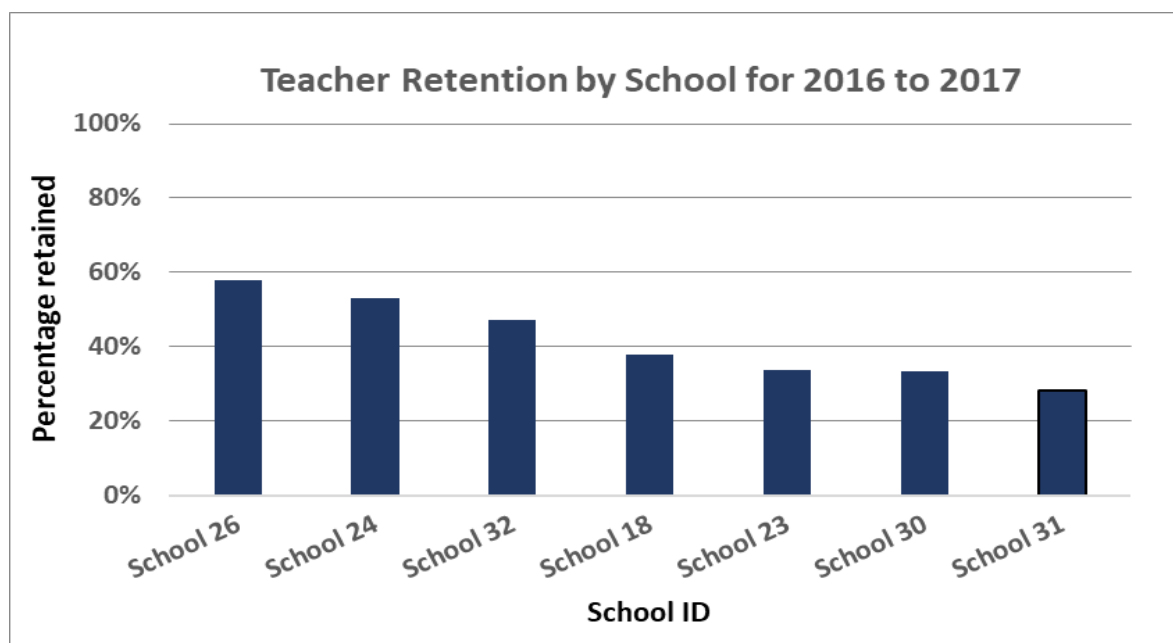
Figure 40: Staff retention for WA Catholic education school for 2016 to 2017.

Implementation

School fidelity measures comprise teacher readiness, instructional leadership, classroom readiness, resource readiness and GGSA readiness that are combined to give an overall fidelity score. A summary of each fidelity measure can be found in Appendix 6. Figure 41 shows scores for all factors and overall school fidelity scores across 2015 (Terms 3, 4), 2016 (Terms 1, 2, 3, 4) and 2017 (Terms 1, 2, 3). There is variability across all four factors over time and no obvious trend for each factor can be identified. However, the overall school fidelity showed a spike in Term 4, 2016 before declining and becoming relatively stable in 2017.
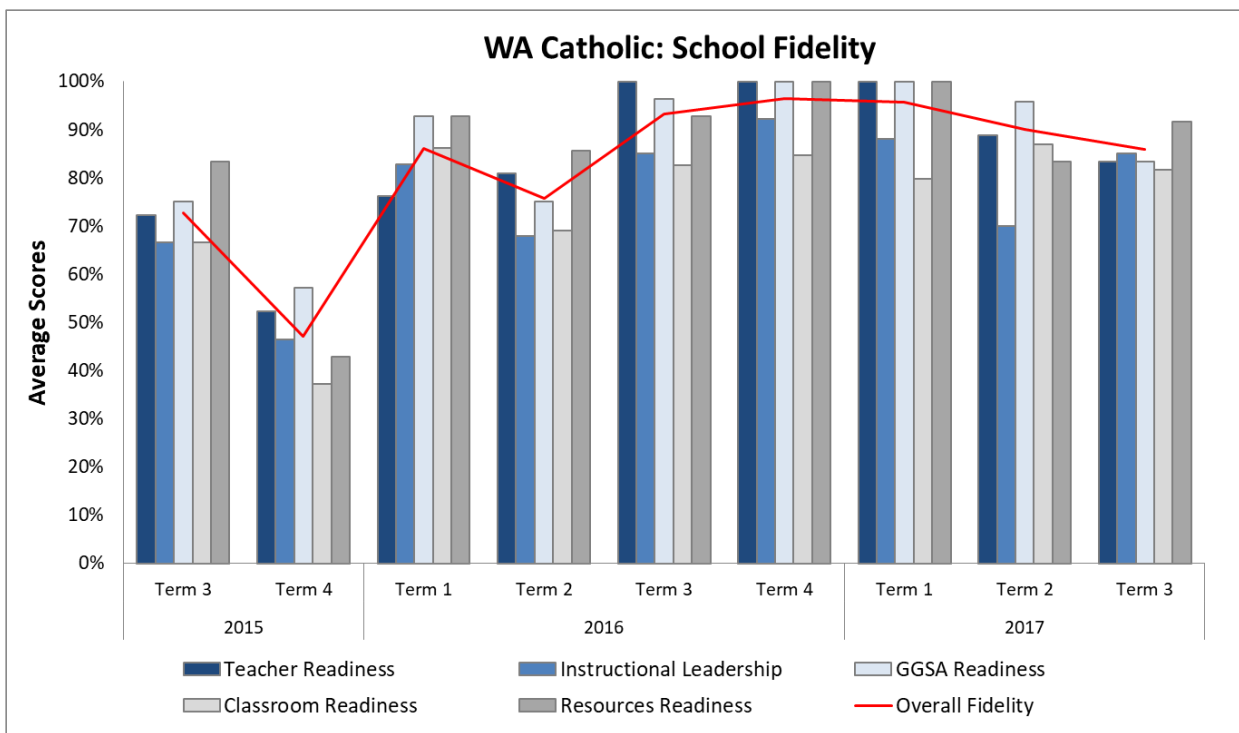
Figure 41: School fidelity measures for WA Catholic education program schools by term, 2015-2017.

Impact: School Staff Perceptions

School staff surveys

In addition to their perspective of the program on student outcome, teacher pedagogy, implementation training and understanding of the program, school staff were also asked whether they would continue with the DI/EDI program, given the opportunity. For the WA Catholic education program schools, almost all staff surveyed (93%) indicated that they would like to continue implementing the program with no negative responses (7% no response).

Some staff explained their reason for continuing the program as follows:

"I would definitely continue with the DI program in our school if given the opportunity. As I have already stated it has produced by far the best results of any program that has been used in our school. I believe it would be a tragedy if it were not continued and supported. I have seen students at our school go from making virtually no progress for years improve their literacy skills markedly."

"100% yes if we were still provided the power points because I can focus on delivering quality lessons which is the most important thing. If I had to plan my own EDI lessons for literacy no because I would have to spend hours developing powerpoints and practise activities - I would have the time to plan how I was going to teach it."

"Yes, we can see the results and improvement in the data collected throughout the primary school and each individual child's progress."

Overall, the DI/EDI program appears to be well received by staff in WA Catholic education schools, and most staff have seen a positive impact on student outcome through the program.

From *Figure 42* below, the majority of teachers who underwent the initial training felt that they were provided with sufficient resources to implement the program (94% agree and strongly agree). Overall, surveyed staff indicated a positive experience with the initial training, and a majority (65% agree and strongly agree) felt that they were more prepared to implement the program in their classrooms post-training.

After the training...

Statement	Strongly Agree	Agree	Somewhat agree	Neither agree nor disagree	Disagree	Strongly disagree
I felt supported in implementing the EDI/DI program	53%	47%				
I was given enough resources to implement the EDI/DI program	59%	35%	6%			
The potential benefits of the EDI/DI program were clear to me	41%	35%	12%	6%		6%
The goals of the EDI/DI program were clear to me	24%	65%		6%		6%
I felt ready to implement the EDI/DI program in my classrooms	18%	47%	35%			

Legend: ■ Strongly Agree ■ Agree ■ Somewhat agree □ Neither agree nor disagree □ Somewhat disagree ■ Disagree ■ Strongly disagree

Figure 42: Staff perceptions of their understanding of the program following initial training (n=17).

These positive experiences were mirrored in some of the long answer survey questions:

> *"Knowing that all the staff are supportive of the program and there are multiple chances to improve implementation with regular staff DI meetings."*

> *"A great team of committed professional teachers whose only goal is to improve student outcomes. The support from GGSA has been tremendous as well.*

School staff interviews

WA Catholic school teaching staff were the most positive about the program and its impact on their school community, but echoed the same barriers and enablers to the program's success that were identified in the other jurisdictions. Interviewees also spoke of the of the ongoing support they had received and how it enabled them to succeed with the program:

> *"We didn't get trained and left to deliver it. We have had mentor support ever since."*

3.8.5. Overall Summary of Collective Results for WA Catholic Schools

Students

- Control and program schools progressed at a comparable rate for students moving from Year 3 in 2015 to Year 5 in 2017.

- Schools demonstrated a medium to large effect size for change over two years (NAPLAN Reading: Hedge's g = 0.85; NAPLAN Writing: Hedge's g = .74; NAPLAN Spelling: Hedge's g = 1.72; NAPLAN Grammar and Punctuation: Hedge's g = 0.42).

- The majority of teaching staff surveyed agreed or strongly agreed that the program had improved students' literacy (85%) and made a difference for the students (85%).

Teachers

- Overwhelmingly, staff believed that the program benefited them as teachers (78%) and improved their ability to teach literacy (71%)

- Most teachers (93%) reported they wanted to continue with the program, the highest rate of any jurisdiction.

- School staff reported having a very strong understanding of the program and its associated teaching practice.

- Overall, the DI/EDI program appears to be well received by staff in WA Catholic education schools and most staff have seen a positive impact on student outcomes through the program.

School

- The majority of schools retained the same principal across 2016 to 2017; however, teacher retention varied across schools, with retention rates varying between 58% and 28%.

- School fidelity was variable across all four factors with no obvious trend for each factor.

3.9. Overall Results (All Jurisdictions)

This section describes the results aggregated for all schools in the program (where data was provided). It begins by considering student-based impacts then moving to teacher and school-based impacts. The evaluation team considered it important to present the survey and interview results as a section in their own right in addition to what has already been presented so that the voice of schools was made prominent in the evaluation. The evaluation team is immensely appreciative of all those teachers, principals, and teaching assistants that took the time to complete a survey or participate in an interview.

3.9.1. Student Level Data

Impact: National NAPLAN

To explore change in NAPLAN over time at the national level, both within and between the program and control schools, two approaches were used. The first approach reported means in graphical format to demonstrate the direction of growth for program control schools compared to National and Very Remote averages. The second approach used independent t-tests to explore the differences in growth between the program and control schools using a change score.

Direction of growth

To demonstrate the direction of growth for program and control schools, mean scale scores across each NAPLAN domain for 2015 and 2017 were created and compared to the National Average and Very Remote National Average. These mean change scores are displayed in *Figure 43* below. As can be seen, both program and control schools experienced growth. However, control schools started slightly higher and ended slightly higher then program schools, with both control and program schools falling below the National and Very Remote school averages.

Growth differences between program and control

While both program and control schools experienced growth, a series of independent t-tests were conducted to determine if program schools experienced more growth compared to control schools between 2015 and 2017. This was accomplished by comparing NAPLAN data from Year 3 in 2015 and Year 5 in 2017 for the NAPLAN domains of Reading, Writing, Spelling, and Grammar and Punctuation. Using this data, a growth variable (2017 – 2015) was created for each NAPLAN domain. Growth scores were then compared between program and control schools.

Preliminary assumption testing did detect some assumption violations, with the presence of minor outliers; however, normality was within acceptable limits, with skewness/kurtosis between with ±2. 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 was 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[16].

An independent-samples *t*-test was run to determine if there were differences in NAPLAN change score between program and control schools. Overall, no differences were detected in overall growth on NAPLAN between control and program schools at the national level between 2015 and 2017 on NAPLAN Reading $t(272) = 0.43$, p = .67; NAPLAN Writing $t(281) = -1.30$, p = .19; or NAPLAN Grammar and Punctuation $t(284) = 0.35$, p = .73. However, significant differences were detected for NAPLAN Spelling $t(284) = -1.25$, p < .01.[17]

---

[16] Where results were the same between original and imputed data, original results were reported.
[17] Sensitivity testing confirms results.
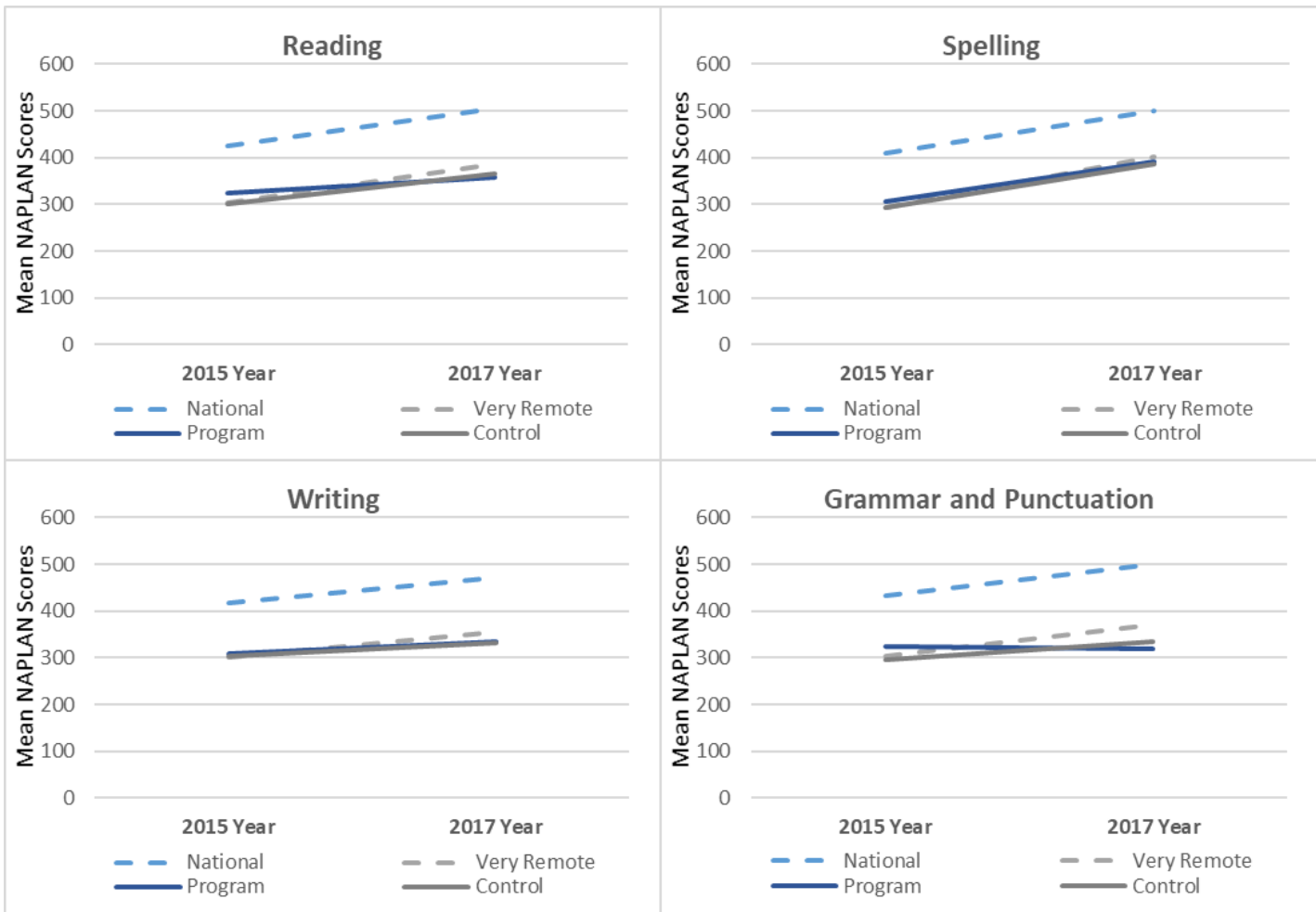
Figure 43: Change in mean NAPLAN scores for all program schools alongside control, National Average, and Very Remote Average

In addition to mean change scores, effect sizes were calculated for each NAPLAN assessment at each school. The distribution of effect sizes across schools are shown in *Figure 44*. Average effect sizes are indicated on each graph. Please note these effect sizes are calculated across two years and are reported using Unbiased Hedges' g. Hedges g is interpreted using the same method as Cohen's d; thus, Cohen's effect size interpretation guide (Cohen, 1988) will be used. As seen in Table 10, a score of .80 or above is considered large. Large average effect sizes were observed for Reading and Spelling, medium effect sizes were observed for Writing, and a small effect was observed for Grammar and Punctuation.

Table 10: Hedge's g Effect Size Interpretation

Interpretation	Effect Size
Small Effect	0.2
Medium Effect	0.5
Large Effect	0.8

Reading — Average = 1.06
Writing — Average = .95
Spelling — Average = 2.30
Grammar and Punctuation — Average = .48
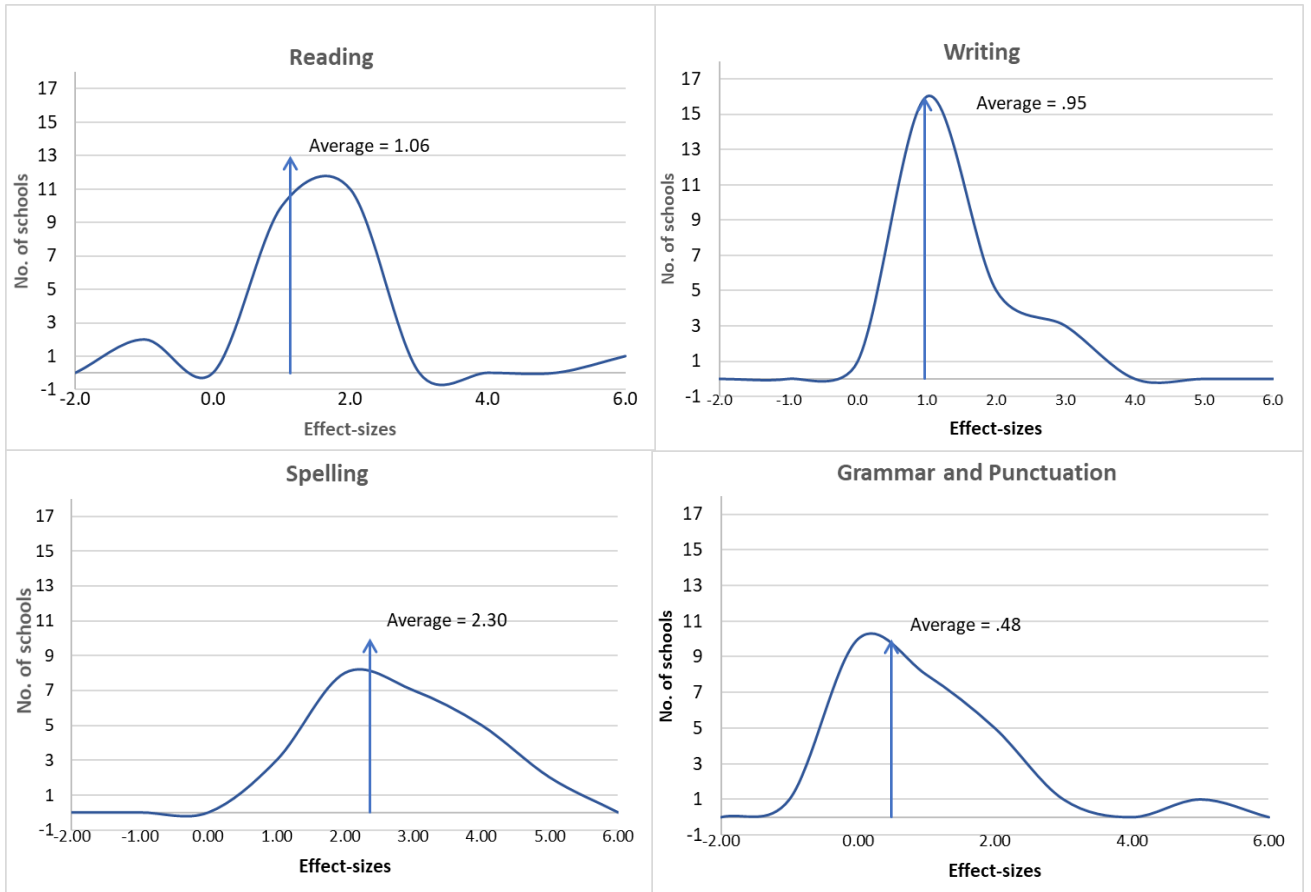
Figure 44: Changes in effect sizes, all schools over 2-year time period.

To explore effect sizes over one year, the effect sizes used above were halved to account for one year's growth. As NAPLAN data is only available in every two years, these scores are not a true reflection of one year's growth as growth may vary by school by year; however, these are the best estimate and are used as a representation. As can be seen in Table 11, Table 12, and Table 13, results are mixed for Reading and Writing, with variation between small to large; Spelling shows mostly large positive effects sizes, while Grammar and Punctuation show considerable negative effect size, although mostly small.

Table 11: Hedge's g Effect Sizes over One Year, NT Program Schools

School	Reading	Writing	Spelling	Grammar and Punctuation
School 03	No Data	Medium +	Large +	Small -
School 05	Small +	Small +	Large +	Small -
School 06	Small +	Large +	Large +	Small -
School 08	Small +	Small +	Large +	Small -
School 11	Large +	Small +	Large +	Small +
School 12	Medium +	Large +	Large +	Small -
School 13	Large +	Large +	Large +	Large +
School 14	Medium +	Small +	Large +	Large +
School 15	Large +	Small +	Large +	Medium +
School 17	Small +	Small +	Large +	Small -
School 34	Small +	Small +	Large +	Small +
School 36	Small +	Small +	Large +	Small -
School 37	Medium -	Small +	Large +	Small -

Table 12: Hedge's g Effect Sizes over One Year, WA Catholic Education Program Schools

School	Reading	Writing	Spelling	Grammar and Punctuation
School 18	Medium +	Small +	Large +	Medium +
School 23	Small +	Small -	Small +	Small +
School 24	Large +	Large +	Large +	Large +
School 26	Small +	Small +	Medium +	Small -
School 30	Small +	Small +	Large +	Small -
School 31	Medium +	Medium +	Medium +	Medium +
School 32	Small +	Small +	Large +	Medium -

Table 13: Hedge's g Effect Sizes over One Year, WA Government Program Schools

School	Reading	Writing	Spelling	Grammar and Punctuation
School 19	Large +	Small +	Large +	Small +
School 21	Large +	Large +	Large +	Large +
School 28	Medium +	Small +	Small +	Small +
School 29	Medium +	Medium +	Large +	Small +

Impact: Perspectives of School Staff (Surveys)

In addition to the survey results presented previously for each jurisdiction, following is a summary of additional survey results from teachers, teaching assistants, and teaching principals regarding their opinions of the program. This includes additional questions as well as aggregate findings for questions already presented.

Demographics

A total of 61 teachers responded to the survey either paper based or online. Most respondents were teachers and were from NT government schools. A demographic summary of respondents is shown in Table 14 below.

Table 14: Breakdown of Survey Respondents by Jurisdiction

	NT GOV	WA GOV	WA CATHOLIC ED
Teachers	27	5	12
Teaching Assistants	1	3	4
Principals	3	-	4
Others	-	1	1
Total	31	9	21

Teaching experience and school tenure

Respondents were asked how long they had been a teacher as well as the length of time they had been teaching at their current school. The data indicated that the majority of teachers who responded to the survey had been teaching for more than 15 years. However, 80% of those who responded had been employed at their school for less than five years, with half of those again employed for less than two years (see *Figure 45*).
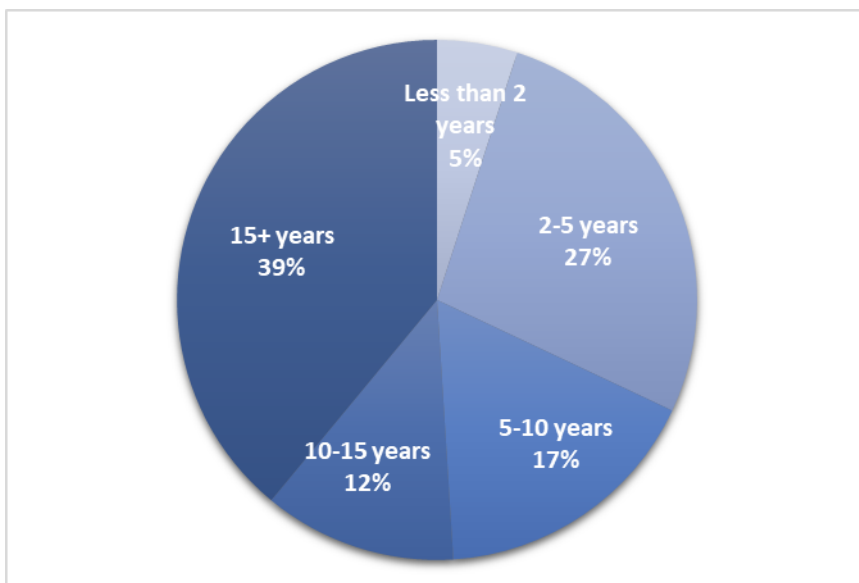
Figure 45: Length of time respondents had been teaching.
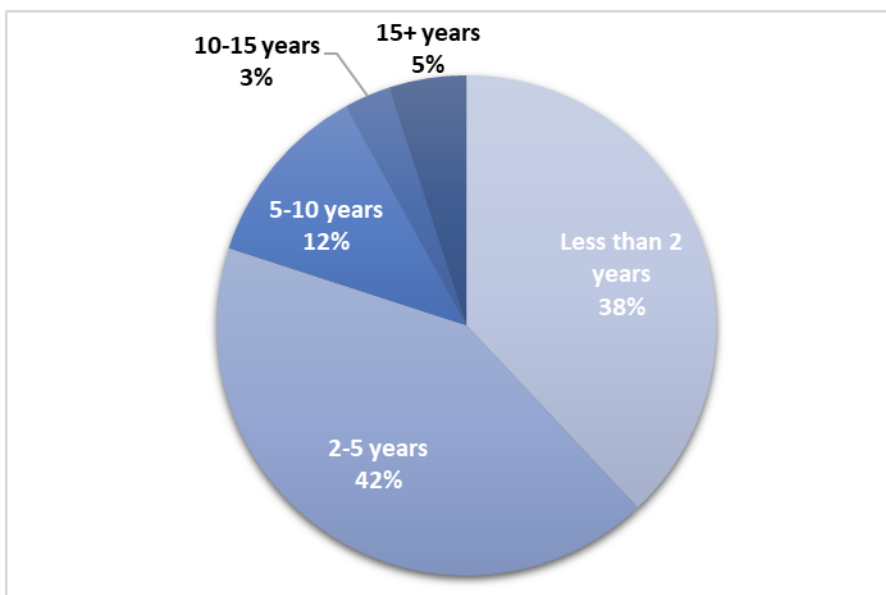
Figure 46: Tenure at current school.

Staff perspective: Impact on students

This section compares the perspectives of school staff across all three jurisdictions of program impact on student outcomes. The responses from teachers, principals and teaching assistants are shown below. The majority of the teachers surveyed are from NT program schools.

Figure 47 indicates that generally teachers either strongly agreed or agreed that students engaged with the program (52%), that it had improved student literacy outcome (57%) and made a difference for them (57%). However, when it came to student wellbeing and student attendance, some teachers were indifferent (23% and 33% neither agree nor disagree) while others did not think that the program had impacted student wellbeing or attendance.

Similar to teachers, school principals generally had a positive perspective of the program impact on improving student literacy and student engagement as shown in *Figure 48*. There were mixed responses (neutral and slightly disagree) from principals regarding the program having improved student wellbeing and attendance.

Please indicate the extent to which you agree with the following statements regarding the flexible literacy program

Statement	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
The program has improved student wellbeing	13%	13%	15%	23%	4%	13%	17%
The program has improved student attendance	8%		14%	33%	4%	20%	20%
The program has made a difference for the students	37%	20%	12%	16%	4%	8%	4%
Students have engaged with the program	19%	33%	15%	15%	4%	4%	10%
The program has helped improve the students' literacy	38%	19%	12%	15%	4%	6%	6%

■ Strongly Agree ■ Agree ■ Somewhat Agree □ Neither Agree nor Disagree ■ Somewhat Disagree ■ Disagree ■ Strongly Disagree

Figure 47: Teachers' attitudes to program's impact on students (n=51-52).

Please indicate the extent to which you agree with the following statements regarding the flexible literacy program

Statement	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
The program has improved student wellbeing		33%		33%	17%		17%
The program has improved student attendance			17%	33%	33%		17%
The program has made a difference for the students	33%		33%	17%		17%	
Students have engaged with the program	17%	50%			17%		17%
The program has helped improve the students' literacy	50%	17%	17%		17%		

■ Strongly Agree ■ Agree ■ Somewhat Agree □ Neither Agree nor Disagree ■ Somewhat Disagree ■ Disagree ■ Strongly Disagree

Figure 48: Principals' attitudes to program's impact on students (n=6).

Finally, teaching assistants were more positive about the outcome of the program, and the majority of the respondents strongly agreed or agreed that the program made a positive impact on students across all aspects of student impact, except for attendance, as seen in *Figure 49*.

Please indicate the extent to which you agree with the following statements regarding the flexible literacy program

Statement	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
The program has improved student wellbeing	23%	23%	31%	8%	15%		
The program has improved student attendance		17%	42%	8%	8%	8%	
The program has made a difference for the students	50%	25%	8%	17%			
Students have engaged with the program	38%	23%	23%	15%			
The program has helped improve the students' literacy	50%	25%	17%	8%			

Figure 49: Teaching assistants' attitudes to program's impact on students (n=12-13).

In summary, across all jurisdictions, overall staff perspectives of the program's impact on students are as follows:

- Generally, teachers and principals thought that the program had a positive impact on the students in terms of student engagement and improving literacy.

- Teachers and principals were more neutral or negative with respect to the program improving student wellbeing and attendance.

- Teaching assistants were the most positive in their responses across all factors, compared to teachers and principals.

Staff perspective: Impact of training in DI/EDI

This section compares the perspectives of school staff across all three jurisdictions on program implementation training. Responses from teachers, principals and teaching assistants are shown separately below.

As seen in *Figure 50*, teachers generally found the implementation training useful with the majority of teacher respondents either strongly agreeing or agreeing that they felt prepared to implement the program after participating in training. Additionally, teachers felt they received enough support (71% strongly agreed or agreed) and resources (78% strongly agreed or agreed) through the training sessions.
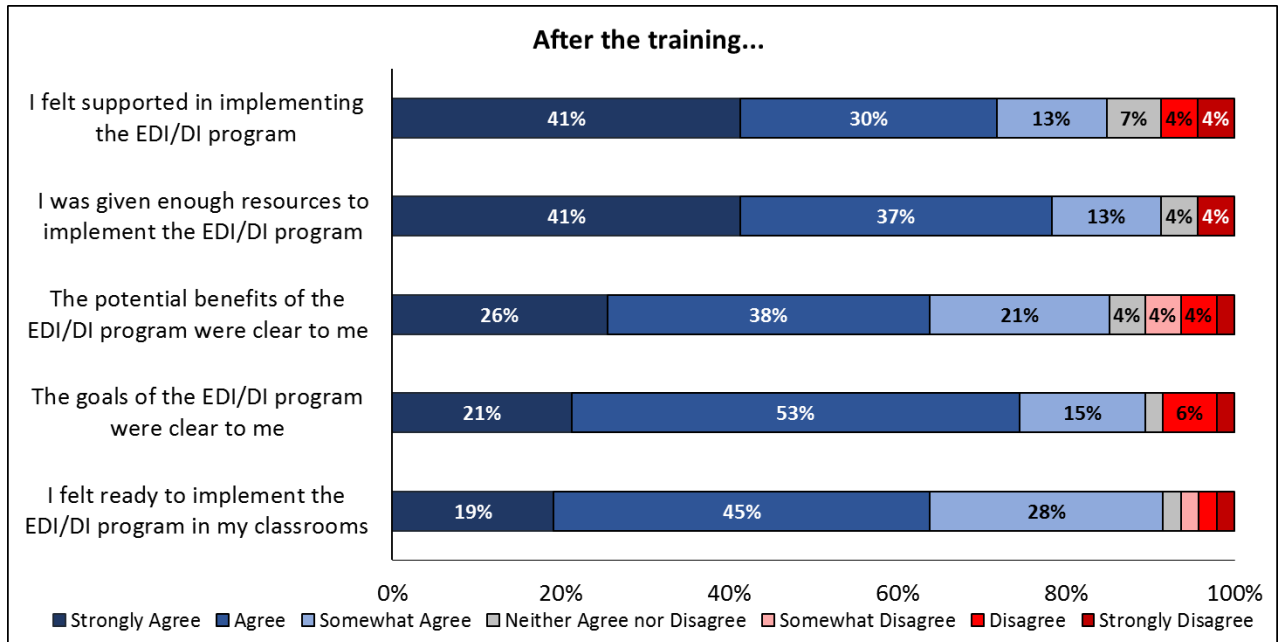
Figure 50: Teacher perceptions of their understanding of the program following initial training (n=46-47).

Principals had an even more positive perception regarding the training compared to teachers, as seen in *Figure 51*. The majority of principals surveyed (>80%) strongly agreed or agreed that the training was beneficial in terms of having enough support, resources, clear goals and benefits of the program and readiness to implement the program after the training.
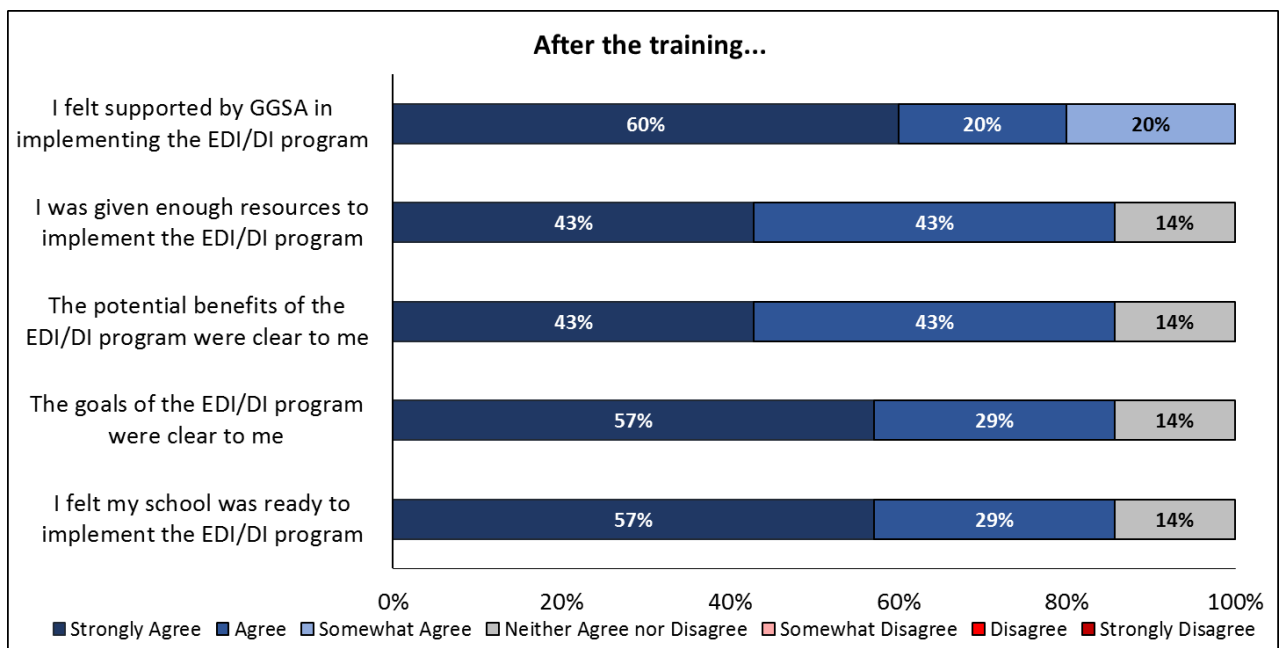
Figure 51: Principal perceptions of their understanding of the program following initial training (n=7).

Similar to principals, teaching assistants also had a positive perspective on the implementation. *Figure 52* shows that teaching assistants were the most positive about being aware of the potential benefits (80% strongly agreed or agreed) and goals of the program after the training (72% strongly agreed or agreed).
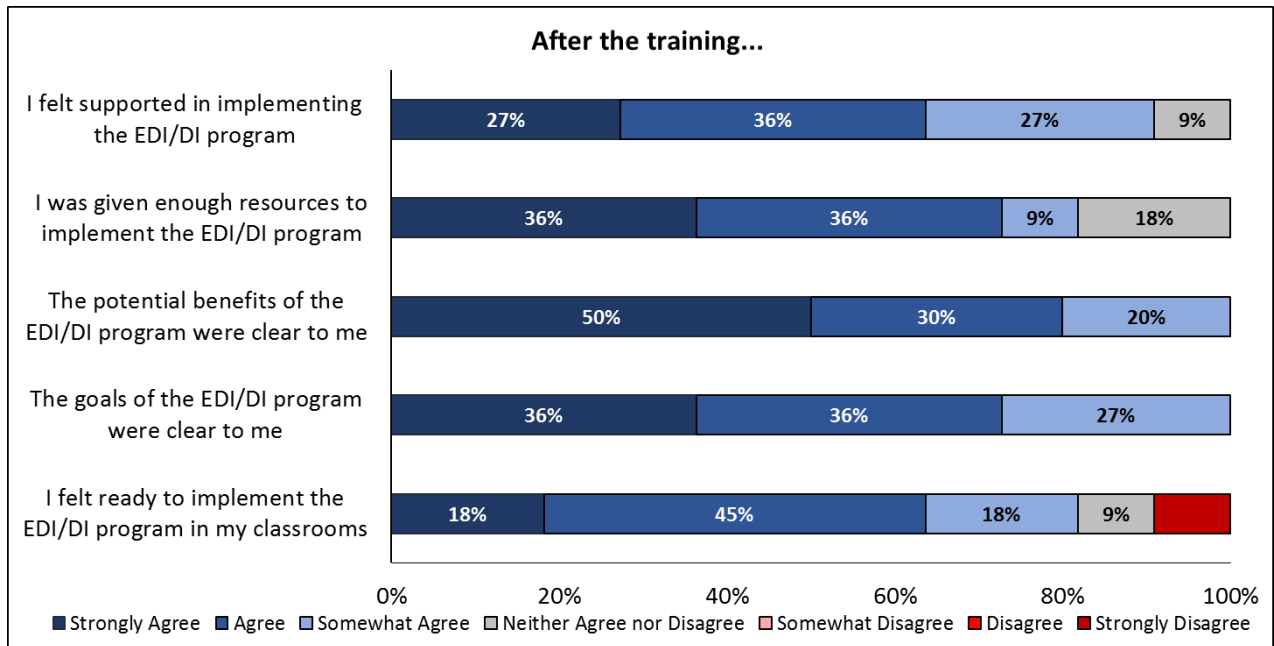
Figure 52: Teaching assistants' perceptions of their understanding of the program following initial training (n=11).

In summary, across all jurisdictions, overall staff perspectives on implementation training were as follows:

- Generally, the implementation training had a positive outcome, with the majority of staff feeling adequately supported with enough resources.

- Principals and teaching assistants had similar views that the training was beneficial in helping them understand the potential benefits and goals of the program.

- A small fraction of the teacher respondents had mixed or more negative opinions about the impact of the training on their understanding of the program.

Staff perspective: Impact on pedagogy

This section compares the perception of school staff across all three jurisdictions on how the program has impacted their pedagogy. The responses from teachers, principals and teaching assistants are shown separately below.

As seen in *Figure 53*, teachers had mixed views on program impact on their pedagogy. Most positive responses were with regards to teacher perceptions of the match of the program for their schools (53% strongly agreed or agreed), how the program guided their practice, and how the program was important for supporting teaching staff (51% strongly agreed or agreed). However, it is difficult to conclude whether or not the program has positively impacted teacher pedagogy as a whole.

In general, I think that the flexible literacy program...

	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
Improved my ability to teach literacy	19%	19%	11%	17%	6%	17%	11%
Helped me to develop better teaching skills	21%	21%	11%	21%		13%	11%
Is a good match for our school	36%	17%		11%	9%	17%	9%
Is important for supporting the teaching staff	19%	32%	17%	15%		9%	6%
Guides my teaching practice	15%	36%	19%	13%		9%	6%
Creates extra work for me	6%	13%	11%	26%	4%	23%	17%
Benefits me as a teacher	23%	19%	15%	23%		9%	9%

Figure 53: Teachers' perceptions of the program's impact on their pedagogy (n=47).

In contrast to teachers, school principals' perception of the program's impact on teacher pedagogy was more positive, as shown in *Figure 54*. Almost all principals disagreed that the program creates more work for their teachers and agreed that it helps guide the practice of teachers. In addition, principals all agreed or strongly agreed that the program was a good match for their school and beneficial to their school.

In general, I think that the flexible literacy program...

	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
Is a good match for our school	50%	50%					
Is important for supporting the teaching staff	50%	25%		25%			
Guides the practice of the teachers in my school	75%		25%				
Creates extra work for the teachers in my school					25%	75%	
Benefits my school	75%	25%					

Figure 54: Principals' perceptions of the program's impact on teacher pedagogy (n=4).

Like principals, teaching assistants' perceptions of how the program impacts their pedagogy was more positive compared to the perceptions of teachers (see *Figure 55*). The majority of teaching assistants (>50%) strongly agreed or agreed that the program improved their teaching skills and that it is a good match for their school. However, most teaching assistants agreed or strongly agreed that the program created extra work for them.

In general, I think that the flexible literacy program...

	Strongly Agree	Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Disagree	Strongly Disagree
Improved my ability to teach literacy	50%	40%		10%			
Helped me to develop better teaching skills	50%	30%	10%	10%			
Is a good match for our school	50%	30%	10%	10%			
Is important for supporting the teaching staff	44%	33%	22%				
Guides my teaching practice	50%	20%	10%	20%			
Creates extra work for me	13%	25%	25%	25%		13%	
Benefits me as a teacher	25%	50%	13%	13%			

■ Strongly Agree ■ Agree □ Somewhat Agree □ Neither Agree nor Disagree □ Somewhat Disagree ■ Disagree ■ Strongly Disagree

Figure 55: Teaching assistants' perceptions of the program's impact on their pedagogy (n=8-10).

In summary, across all jurisdictions, general staff perspective of the program's impact on their pedagogy was as follows:

- Generally, teachers' perceptions of the program's impact on their pedagogy was mixed, leaning towards positivity. However, many respondents felt that the program created extra work for them.

- As with responses for implementation training, principals and teaching assistants had similar views that the program had positively impacted teacher pedagogy.

- Principals generally did not think that the program created more work for their teachers, while teachers and teaching assistants had mixed views about the program creating more work for them.

Staff perspective: Understanding of the program and role within the program

This section compares the perception of school staff across all three jurisdictions on their understanding of the program as a whole. The responses from teachers and principals are shown below. Note that teaching assistants were not asked these questions in the survey,

Overall, most teachers seemed to have a good understanding of the program as a whole (*Figure 56*). Almost all teachers surveyed either strongly agreed or agreed that they understood the tasks and activities of the program and how to teach literacy using the EDI/DI pedagogical approach.

I feel as though I have a good understanding of...

The specific tasks and activities students are required to do	46% · 46% · 4%
My role within the EDI/DI program	47% · 39% · 10%
The pedagogical approach specific to EDI/DI	38% · 42% · 12% · 4%
How to teach literacy	41% · 47% · 6% · 4%

Legend: ■ Strongly Agree ■ Agree □ Somewhat Agree □ Neither Agree nor Disagree □ Somewhat Disagree ■ Disagree ■ Strongly Disagree
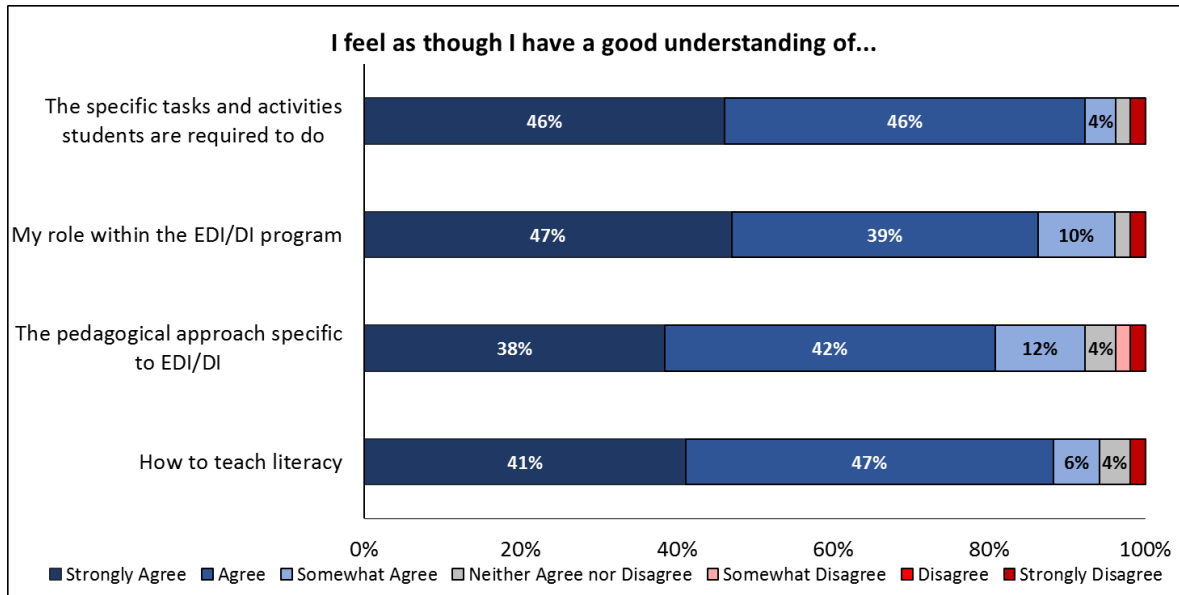
Figure 56: Teachers' perceptions of their understanding of the program (n=51-52).

As shown in *Figure 57*, principals also showed that they had a good understanding of the program and indicated that they were aware of their role within the EDI/DI program structure.

I feel as though I have a good understanding of...

The specific tasks and activities students are required to do	43% · 43% · 14%
My role within the EDI/DI program	43% · 43% · 14%
The pedagogical approach specific to EDI/DI	57% · 29% · 14%
How to teach literacy	57% · 29% · 14%

Legend: ■ Strongly Agree ■ Agree □ Somewhat Agree □ Neither Agree nor Disagree □ Somewhat Disagree ■ Disagree ■ Strongly Disagree
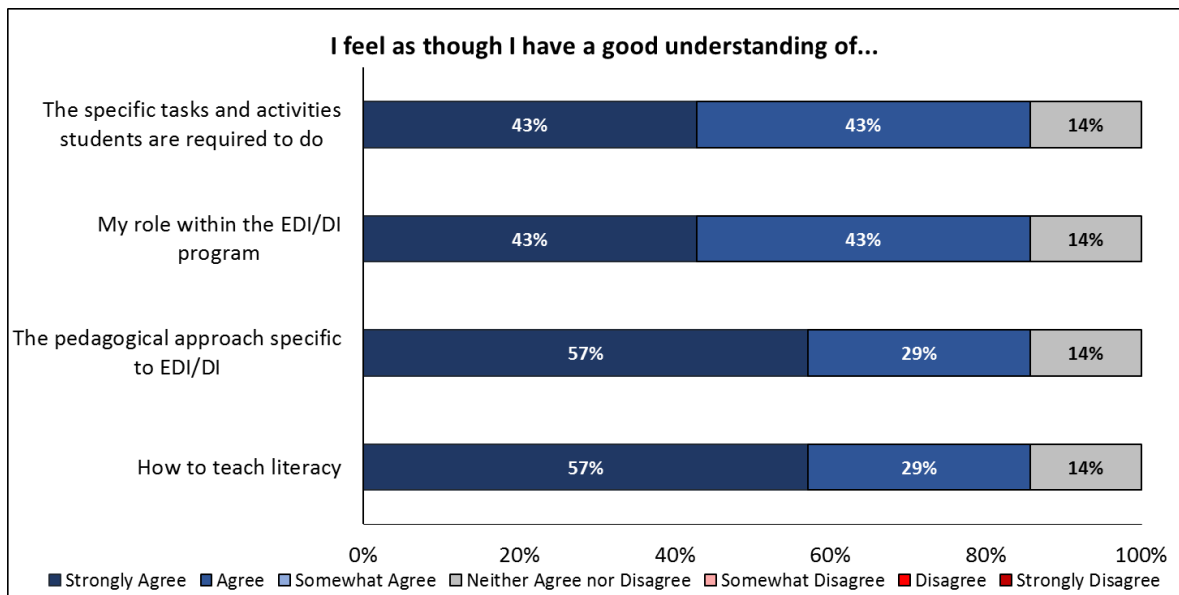
Figure 57: Principals' perceptions of their understanding of the program (n=7).

In summary, across all jurisdictions, general staff perspectives on their understanding of the program were as follows:

- A small fraction of the teacher respondents had mixed or more negative opinions about their understanding of the program on each factor as compared to principals.

- Overall, most staff (teachers and principals) indicated that they had a strong understanding of the EDI/DI pedagogical approach and were aware of how to teach literacy through the program.

Teacher-driven modifications to program

Respondents were also asked whether they made modifications to the program in the course of teaching DI or EDI. The majority of respondents replied that they had. These modifications were then coded into specific categories in the analysis. Results indicated that most modifications were content adaptations either to culturally align the program to student needs or to simplify information to better suit student abilities. *Table 15* provides the relevant categories and examples of responses from the surveys. *Figure 58* displays the percentage breakdown of each modification category.

Table 15: Modification Categories and Quotes from Teachers

Category	Quote
Content adaptation	I have simplified slides to make them more appropriate for student understanding and often replaced examples with more culturally relevant texts
Content enhancement	Added incidental teaching moments and own lesson plans to support the topics.
Additional practice	We have focused on 'Value Adding.' We stick to the consistent language provided but will add additional practice on difficult items. For example, when learning about the parts of a car, we will go outside and touch the parts of the school car.
Ability grouping	We stream our students so that they are working at a level that allows them to be successful.
Activity removal	Have not included the Thermometer reward chart system.
Time allocated	Timing. Sometimes it is just not appropriate to spend the required time. However other days more time was spent on DI.
Pacing	Pace, slowed down and no groupings.
Differentiation	Made easier for weaker kids – eg using independent practise from a lower grade level instead of the independent practise for their year. Harder for more able student eg think of your own example instead of using the books example.
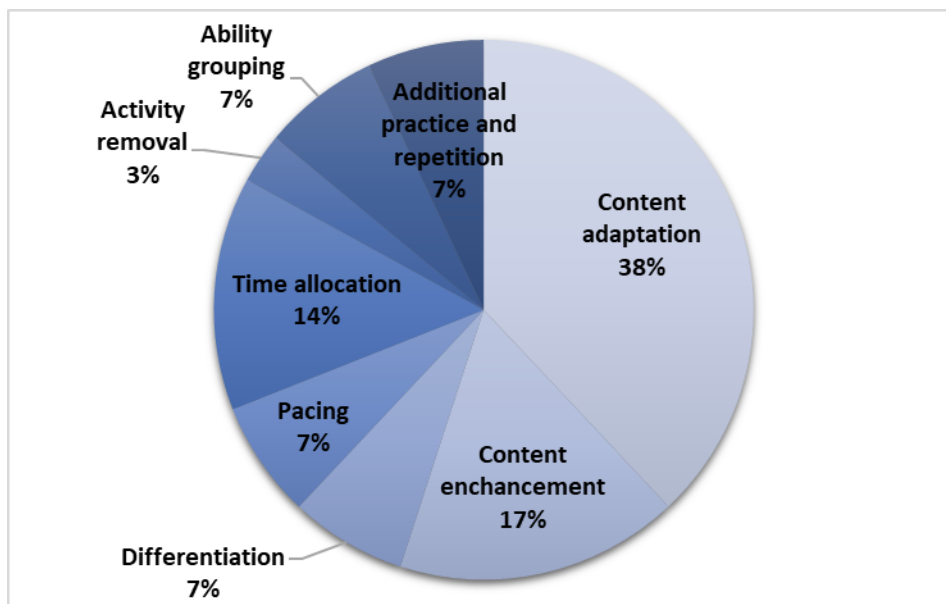
Figure 58: Breakdown of modification categories.

Time spent teaching the program and covering content

The DI program is scheduled to be taught for 2.5 hours per school day. Respondents were asked how long they taught the program each day with the overall average amount of time calculated as 2.5 hours. Nonetheless, many teachers reported allocating more time to the program (e.g. 3+ hours) and some reported time reductions (2 hours). The minimum allocated time was one hour and the longest allocated time was five hours per day.

In addition to the allocated hours per day for the program, respondents were also asked if they were able to cover the allocated content in the time available. The majority of respondents (52%) indicated that they were able to cover most of the content in the allocated time. Of the 38% that said they were not able to cover all the content, reasons given for this included: student attendance, behaviour and their understanding of concepts, lack of staffing, and multi-age classes. 10% of respondents did not provide commentary for this question.

Overall opinion on the success of the program

Respondents were asked their overall opinion of the program. This question was slightly different between the paper based and the online survey. For the online survey, free text responses were coded with results indicating strong positive support for the program overall (n=23) (see *Figure 59*). In the paper-based survey, respondents were asked if they felt the program had been successful overall. Responses were evenly divided between yes and no (overall n=15).

> *"I enjoy seeing the education of students moving forward and that they can make good progress quickly. I liked the student /teacher game because it resolved behavioural issues with the minimum of effort. I liked the organisation of the lessons. They fitted smoothly together. I liked the scientific approach, it was based on evidence, it has been observed over time."*
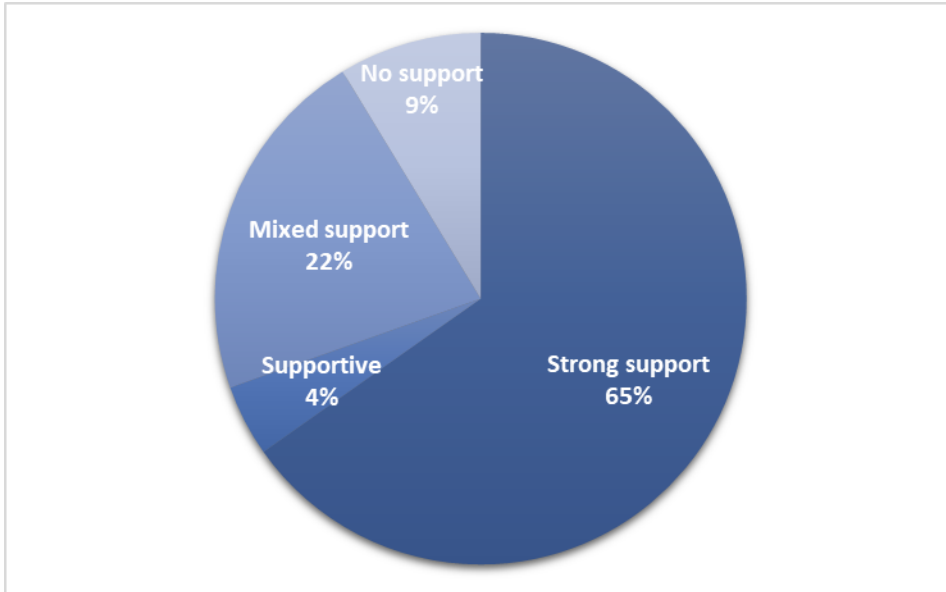
Figure 59: Breakdown of responses on overall support for the program (n=23).

3.10. Schools that Have Exited

In 2017 following the release of the Main Report, funding for the FLFRPSP was reviewed and in consultation with all stakeholders, eight schools exited the program. The following analysis tables the performance of these schools against the average of the remaining schools (active) using NAPLAN as well as feedback from exited schools on their perceptions of the program. While the NAPLAN results were in the main quite positive for exiting schools, it became clear from the discussions with school staff interviewed that the program was not a good fit for their context. Staff reported being forced into taking on the program by their jurisdictional department, and reported that a lack of community consultation and acceptance as well as staff resistance contributed to the problems with the program.
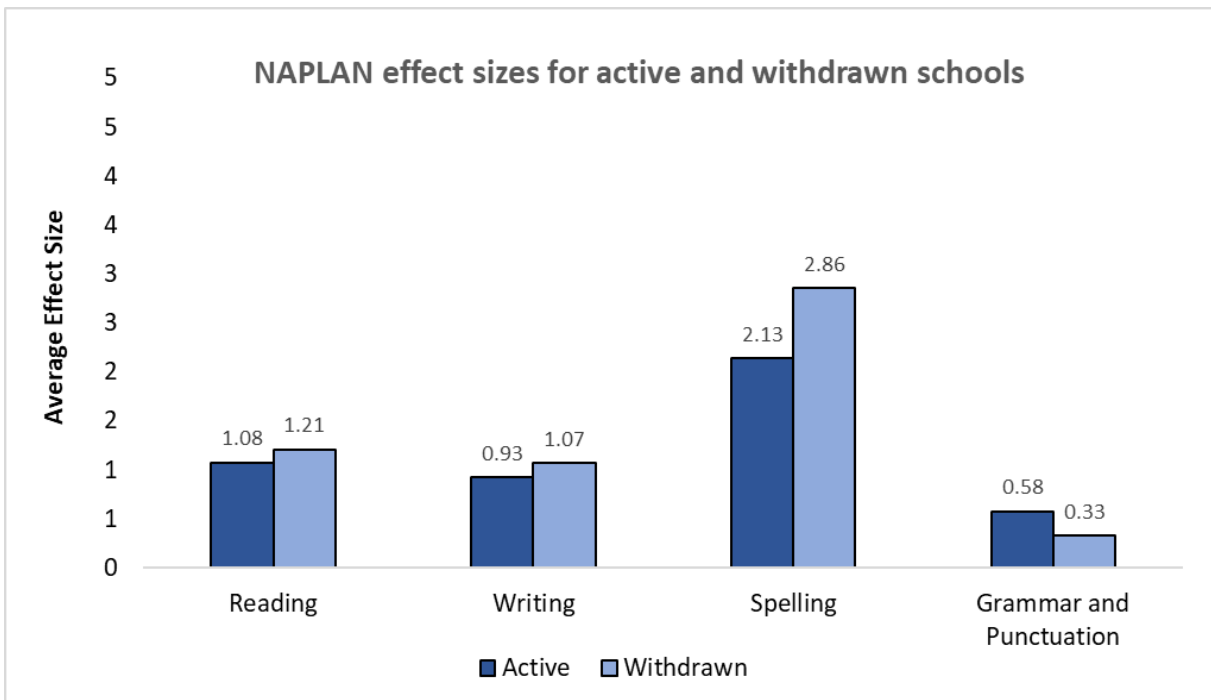
Figure 58: Average overall effect sizes for NAPLAN Years 3-5 2015-2017 for active and withdrawn schools.

Table 12: Effect Sizes for Withdrawn Schools across NAPLAN Years 2015-2017

	Reading	Writing	Spelling	Grammar & Punctuation
School 5	0.81	0.78	2.44	-0.54
School 6	0.37	2.37	4.17	-0.11
School 28	1.25	0.23	0.57	0.27
School 36	0.42	0.47	2.47	-0.78
School 13	5.61	2.49	3.57	1.66
School 34	0.70	0.70	4.02	0.61
School 15	1.70	0.94	1.80	1.54
School 37	-1.19	0.54	3.81	-0.02

Staff from three schools that exited the program in 2018 were interviewed by the evaluation team regarding their experiences with the program, as well as their school's approach to teaching literacy after having left the program. The results of these discussions are presented below.

All interviewees felt that the program was inappropriate for their community, school, teachers and students. Interviewees emphasised that the program's structure and prescriptiveness did not match the learning needs of their students.

> *"It did not fit in the context of this school. The literacy levels of the children here is exceptionally low here even for remote communities."*

> *"We felt it was completely inappropriate for our students needs and learning styles. The transience meant there was not continuity. The kids felt like they had no idea what was going on if they missed a week here and a week there. Our program needs to be very individualised and based on the kids' strengths of where the kids are at—not just a scripted program."*

Two of the three interviewees felt that the program was having a negative impact on student literacy and other academic outcomes:

> *"[Left] because the program was not meeting the needs of our students and they were basically voting with their feet and not coming to school. Because it is such a scripted program, the teachers could not build on the students' strengths as much as they wanted. At the end of the day, management in the classroom became a bit of an issue—because the teachers are required to go through the program in a very detailed way, which leaves the other kids unattended."*

All participants felt that there was insufficient consultation with both the schools and community before the program's initial implementation, and that this negatively impacted its uptake:

> *"I do not know why DI was imposed upon us. I have only been here since the start of fourth term, but it was pretty clear to me from the get go that this program was inappropriate… It is a discovery and play based learning that we have got at the moment."*

> *"Began in 2015, it was not consulted with the community. I believe there was no chance of the community to have say, and the principal at the time was not very enthusiastic but told he did not have a choice… I had no prior experience with DI before and came in very open-minded [as the new principal]."*

A reoccurring theme throughout all interviews was that teaching staff did not like the program, and that teacher 'buy-in' was critical to the program being implemented effectively and with fidelity:

> *"It caused the teachers a lot of stress; quite a lot of teachers left at the end of 2016. It was too much for them. A couple of them still remained into 2017. The amount of time that was meant to be spent on DI left no room in the curriculum for any other subjects."*

> *"The staff must be 100 percent behind it. We had staff at this school who were very against it… I was not here from the beginning. They did not like to read from the textbook. They had failed the training… We are talking about [the need for] quality staff and not just the number of staff, people who are doing the right thing. This could be walking talking text, accelerated literacy—it doesn't really matter which program we are talking about, it has got to be the whole school on board, everybody following the rules."*

All interviewees reported being more satisfied with the approaches to teaching and learning literacy that were implemented after leaving the program. This was in terms of both student learning and progression, as well as broader community and cultural engagement:

> *"Using the new phonics program, we have children how to decode… The whole time last year, they did not even know their letters, and now the teacher has got some of them decoding some simple words like cat. It is like it has clicked in a few weeks. It goes to show we made the right choice. The attendance has been up this term as well [since leaving DI]."*

> *"The families are engaged more too [since DI]… I do not think you can understand the extent of the damage of the school-community relationship over the last through years because of DI.*

4. Discussion

4.1.   Summary of Results against program aims

The foundational aims of the FLFRPSP were to:

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

The intent of this report is to provide a summative assessment of the impact of the FLFRPSP for the school years 2015-2017 against these aims. Since the last report additional data has been collected in the following domains:

- Literacy data for 2017 school year (WA Catholic EYLND data only)
- NAPLAN 2017
- Program based data for Terms 1, 2 &3 2017
- Additional perceptions of the program from teacher aides, teachers and school leadership

With the inclusion of the additional data sources the following results are observed for the program.

4.1.1.   Aim 1: Improving students' literacy abilities and results

According to the CIPP model (Stufflebeam, 2000) utilised within this evaluation, Product (or Impact) for students was defined changes in literacy skills, engagement and behaviour and wellbeing. The measurement model identified relevant data collection instruments to capture these outcomes within the scoping brief from the Australian Government. Literacy skills have been assessed using all possible and available data depending on the jurisdiction. Specifically, all jurisdictions provided NAPLAN assessment data for the majority of schools. Catholic Education WA provided literacy data in the form of EYLND results for students in years 1-3. There is still a need to acquire more extensive literacy data and this is discussed further in the following section.

How should progress be measured?

In the original logic model, progress for students was cited as follows:

- 50% of students at or above grade level
- 50% of students at or above the 50th percentile on PAT-R, EYLND and PM Benchmark Reading measures

Given that NAPLAN is an assessment of general literacy skills, it was initially considered too broad to examine the effectiveness of the program. Unfortunately, as the evaluation has progressed, NAPLAN has become prominent in the analysis of program impact due to a lack of access to other literacy data.   It should be clearly acknowledged that NAPLAN falls short as a single outcome measure of literacy progress on which to assess the FLFRPSP because of its measurement properties and its content. NAPLAN is inadequate on its own because:

- Not all students in the program are represented at any one time – only those in Years 3 and 5 of the relevant evaluation year;
- As mentioned it is a relevant but more distal measure of skills taught in the program but does not measure directly whether students have mastered component reading skills such as Phonics or phonemic awareness

Literacy progress

NAPLAN: Impact by literacy domain.

With the addition of 2017 NAPLAN results, steady improvements in NAPLAN were observed, particularly for reading, writing and spelling. Although not statistically significant, program schools demonstrated stronger progress compared to control schools in Writing and Spelling. The distribution of effect sizes for NAPLAN change between 2015 and 2017 demonstrated an average Cohen's d of 2.37 for Spelling (2 years), and 1.12 for Reading (2 years) indicating substantial gains on these component academic skills. Effect sizes for Writing and Grammar and Punctuation were more moderate. Strong gains in Spelling across most schools. Twenty of the schools with adequate NAPLAN data demonstrated large, positive effect sizes on Spelling over one year.

Literacy measures – The preceding report demonstrated cautious pockets of student literacy growth predicting future positive impact of the program on literacy skills. These pockets were particularly evident in WA Catholic Education schools. WA Catholic Education schools provided specific literacy data (EYLND) for their students. The results of the analysis of this data showed a strong trend towards greater gains in program versus control schools. Significant growth was demonstrated by program schools at each analysed time point and based on the current data, program schools appear to be 'breaking away' in terms of performance from their control school counterparts. Nonetheless, more time is needed to understand if this trend is sustained, although current indications are positive. Further analysis of sub-tests in the EYLND assessment indicated significant differences between program and control schools with program schools performing better on 5 subtests. This is an increase on the outcomes in the last report where program schools performed significantly better on only one subtest. This analysis is considered more relevant than NAPLAN in terms of understanding program impact as the individual sub-tests within the EYLND assessment are more closely aligned to the skills taught within the program. In addition, this analysis shows that literacy gains by program schools in their own right and against control schools have been consistent over time. This supports the idea presented in the last report that the program should be afforded the provision of time so that it can embed within each context and gains can be realised.

How much progress should we expect to see for students?

Given the complexity of contexts across which the program has been implemented, what levels of progress and growth should be expected in students participating in the program? Consideration of progress in other similar student groups and with similar programs provides a starting point to understand the levels of progress that should be expected but first, the structure of the reading programs in FLFRPSP should be unpacked to provide useful comparative points. DI Reading Mastery (RMSE DI) taught in the majority of schools in the FLFRPSP is a reading program designed to teach students decoding and comprehension skills. According to the NIFDI website and from observations of the evaluation team at DI training in 2018, a typical DI reading lesson contains a set of activities that cover the following reading component skills:

- Phonemic awareness
- Letter-sound correspondence (phonics)
- Sounding out of words
- Word recognition
- Vocabulary
- Oral reading fluency
- Comprehension

Given these skills, other reading pedagogies and programs that include these component activities and development of these skills may offer a basis on which progress could be benchmarked. The literature review presented in the results section of this report provided indications of success in other programs. For example, the MULTILIT suite of programs for low progress readers are similar to the RMSE DI program in terms of content delivery and component skills taught. MULTILIT has been implemented in at-risk and low SES populations with moderate to large effect sizes between program and control groups and larger effect sizes for within program groups (up to 3.80) (Wheldall, 2009; Buckingham, Beaman-Wheldall & Wheldall, 2014). In addition, evidence from the implementation of DI in Cape York schools has demonstrated significant gains in literacy outcomes with large effect sizes (2.1 NAPLAN Reading, 1.1 NAPLAN Writing) (Hattie, 2016). Given these evidence-based benchmarks, the FLFRPSP is achieving similar magnitudes in its NAPLAN outcomes. A task of the evaluation in the forthcoming year is to work closely with jurisdictions to further develop benchmarks based on additional contextual factors.

Perceptions of program impacts on students

In addition to literacy outcomes, the perceptions of school staff regarding program impacts on students across several domains were assessed through interview and a voluntary survey sent to schools. This is a critical aspect of the evaluation in terms of capturing the voice of the staff and understanding their perceptions of student impacts. Literacy and NAPLAN data can show whether students are improving however, if staff feel there are other impacts on students (positive or negative) then these need to be captured as equally important to student literacy gains.

Specific questions in the survey and interview addressed the extent to which the program had improved student wellbeing, attendance and engagement as well as literacy. These responses were captured as ratings and commentary from school staff. Across the jurisdictions there was overwhelmingly positive support for the idea that the program had improved students' literacy and that students had engaged with the program. This indicates, even in the absence of complete and comprehensive assessment data that improvement in student literacy has been observed. This observation of improvement was often seen in spite of staff feeling negative about the program and its approach, and the impacts on their teaching practice. In essence, student progress was acknowledged even when staff did not support the program. Nonetheless, a relatively small proportion of staff (16% of survey respondents) did not agree that the program had improved students' literacy and were largely negative across all aspects of the program. It is important that these views are acknowledged in this report.

There was more mixed support for the ideas that the program had improved attendance or student wellbeing. In WA Government schools, just over half of those school staff surveyed felt that the program had effected improvements on student wellbeing and attendance however, a quarter of staff strongly disagreed that improvements in attendance had been made as a result of the program. Similarly, nearly half of staff in NT Government schools indicated that they disagreed that the program had improved student attendance. In Catholic Education WA schools, one third of staff felt there had been no impact on attendance. The discussion below addresses the issue of attendance in more detail.

> *To conclude, the evaluation findings provide general support that the FLFRPSP has improved student literacy outcomes. Large effect sizes have been observed in NAPLAN Spelling across a majority of schools as well as Small through to Large effect sizes in NAPLAN Reading and Writing. Particular gains are demonstrated in Catholic Education WA schools based on more specific literacy data. Overall, observed gains have demonstrated similar levels of effect as previous compatible interventions in compatible student populations. Furthermore, teacher and Principal perceptions of literacy gains were overwhelmingly positive, as was support for increased student engagement. Less positive perceptions were recorded by school staff regarding improvements in wellbeing and attendance suggesting that these factors remain complex issues in the context of remote and rural primary school education.*

4.1.2. Aim 2: Increase teacher pedagogical skills in teaching literacy through the use of alphabetic teaching approaches, in particular DI or EDI

Implementation of any new pedagogical approach alters and shifts the practice of the classroom teacher. In the FLFRPSP, markers of teacher skills and practice included measures of teacher effectiveness, derived from the program itself and assessed against 3 indices: Classroom Organisation, Instructional Delivery and Behaviour Management and Engagement. Whilst there are pockets of missing data on these measures, available results indicated steady levels of teacher effectiveness across all jurisdictions. To also note, teacher perceptions of their preparedness to teach the DI or EDI programs after receiving training indicated a large majority of staff felt supported, adequately resourced and ready within themselves to implement and teach the program in their classroom.

The evaluation also sought to understand teacher perceptions of how the FLRPSP impacted on their practice. This was captured through interviews and survey questions. Teachers in WA Catholic Education schools were generally positive about the impact of the program on their teaching practice indicating that the program had

enhanced their ability to teach literacy as well as their general teaching skills. This was also the case for teachers in WA Government schools. Teachers in the NT were slightly less positive regarding whether the program improved their literacy and general teaching skills although more than 50% of responses were positive. However, NT teachers were less positive about the program being a good match for their school compared with WA Government and Catholic Education teachers.

Interviews with Principals and teachers provided a more nuanced understanding of the formative nature of the program on teacher practice. There were several comments regarding the prescriptive nature of the program and that this had a negative effect on teacher motivation with one Principal noting that his school had *'quite a lethargic approach to teaching it.'* In contrast, others noted that the mastery nature of the program and the high visibility of progress and results provided opportunity for staff to buy into the students' success and this was seen as professionally motivating.

> *"It has made me extremely excited to see results with children who might not have ever seen the joys of literacy! I feel happy to see all levels of literacy being taught effectively."*

In summary, one of the overall aims of the FLFRPSP was to increase teacher pedagogical skills in literacy through the use of alphabetic teaching approaches. According to the data collected across the evaluation (2014 to 2017), both program-based data and teacher/Principal perceptions, this aim has been achieved. GGSA in conjunction with NIFDI has successfully delivered training that provided teachers with the necessary understanding of the program and its goals for teachers to feel 'classroom ready' to teach it. Furthermore, teachers almost unequivocally (with some minor exceptions) considered that the program had improved their ability to teach literacy as well as improving their more general teaching skills.

4.1.3. Stage of implementation

Whilst understanding the degree and nature of impact on students and teachers is of primary importance, this knowledge must be contextualised with time based on the stage of program implementation and the results demonstrated to date. Figure 60 illustrates the proposed progression of impact for the program. The last report in June 2017 indicated that Clusters of Impact were apparent, particularly for Catholic Education WA schools. One year on and at this stage of implementation, more diffuse observable impact should be apparent. This is because the program has had time to embed and more students have been exposed to the pedagogy. Diffused, observable impact is demonstrated in Catholic Education WA schools where it has had a large and sustained effect on student literacy outcomes. NAPLAN results also indicate steady growth and differences between control and intervention groups in Writing and Spelling. Furthermore, it could be reliably predicted based on current trends that the intervention groups in some areas are set to break away from control groups in their progression.

Unfortunately, it is not possible to make detailed or firm conclusions about the impact on student literacy in NT Government schools as the evaluation team was not afforded access to the literacy data, only NAPLAN was made available for the evaluation. Based on NAPLAN data it appears that the program is having large effects on Spelling skills and smaller effects on Reading and Writing. In aggregate, these effects are not significantly different from control schools although individual program schools are showing significant growth compared to control schools. In summary, whilst we would expect to see diffuse observable impact in NT and WA Government schools, the evaluation does not have adequate data to support this claim.
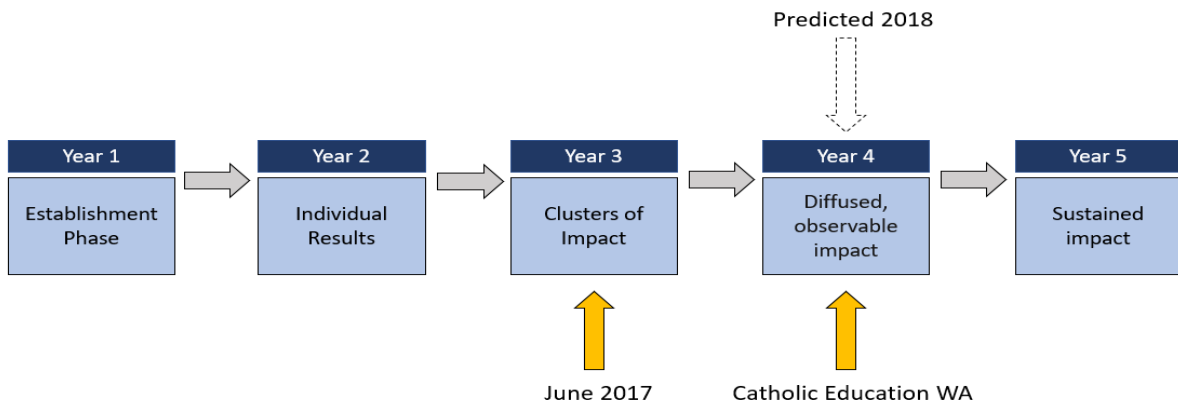
Figure 60: Progress over a 5-year period

4.2.   The theory of change

When asking if the program has achieved its aims, discussion of impact at this point has centred on an isolated view of the theory of change which forms a subset of the overall program logic model. Program logic models are like "road maps" that illustrate how the program will work and what the underlying assumptions and external factors are that influence how the results will be achieved. The main aim of program logic models, is to "clarify and test" the rationale underpinning what can be achieved given available resources and how the design and implementation have been conducted. As stated in the last report, logic models or theories of change for programs evolve over time as the context in which the program is implemented becomes more salient and nuanced and the outcomes within direct influence of the program more obvious. The theory of change is really a subset of the overall logic model, a statement of what is expected to happen if we hold all other factors constant.

Nonetheless, consideration of impact is only partially complete without factoring in other aspects of context and implementation. To broaden the understanding of impact and implementation, the isolated theory of change must be placed into context (Figure 61.). The evaluation has identified a set of factors considered important in differing degrees across participating schools categorised as follows: students, program, school and system. A fourth category, considered separately is evaluation readiness.
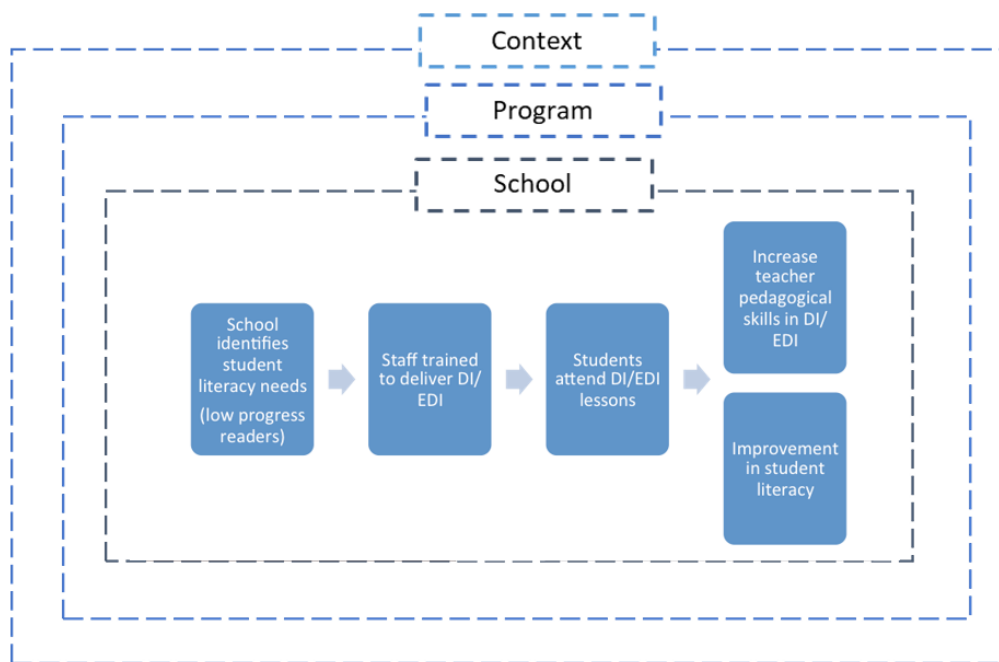
Figure 61 Contextualised Theory of Change

4.2.1. Factors affecting the theory of change

Literature summarised by Fixsen, Blase, Metz, & Van Dyke (2013) suggest three categories of implementation that must be considered: diffusion, dissemination and implementation. Diffusion refers to the spread of an innovation which is unplanned and informal; dissemination reflects a more planned and centralised adoption of innovation and full implementation refers to adoption and *integration* of new practices within a context. The FLFRPSP has disseminated direct instruction teaching approaches into rural and remote schools but has not fully implemented.

There are a number factors that contribute to full implementation, including practitioner and organisation readiness, system influences, guidelines, training, and other forms of communication that urge understanding and use of innovations (Brownson, Colditz, & Proctor, 2012). With full implementation, the goal of using an innovation is pursued until the core elements of that innovation are being used and are delivering promised results in practice (Fixsen et al., 2005). In the evaluation of the FLFRPSP it has become evident that the theory and rationale for the intervention is strong and assumptions are correct, however it is the theory of implementation where issues arise. Implementation goes beyond just 'doing the program'. For full implementation of the FLFRPSP to occur, these factors must be fully understood. Whilst many of them are contributing positively, others are more problematic. There is little doubt that schools and perhaps sectors can adapt the theory of change underpinning the FLFRPSP and achieve success, however there are conditions that are necessary to ensure adaptation can have a more consistent impact.

Using mixed methods the evaluation was able to collate and triangulate information to understand the influence of specific factors in the logic model that acted either as a push (promotion) or pull (demotion) influence on the program theory of change and the life course of implementation (see Figure 3).
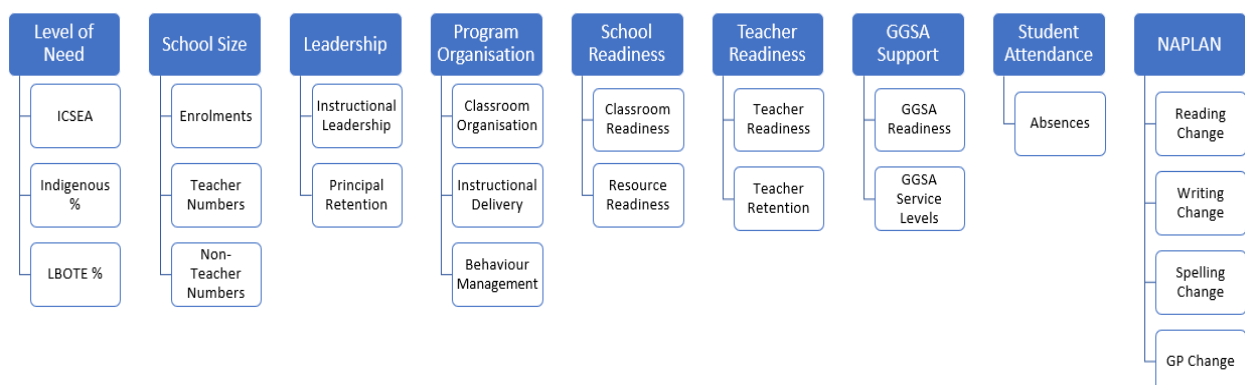
Figure 62 Factors associated with the intervention and implementation

The logic model developed in the early phase of the evaluation is presented as Figure 63. The discussion to follow highlights those aspects of the logic model considered particularly critical to the program through our collection of data and perceptions of the program's success. More specifically, those factors that have supported the theory of change and those factors that have pressured it which are now discussed.

Students

Characteristics of students in rural and remote schools

One aspect of the model that warrants attention is the situational analysis of students in terms of defining the population of interest for this program and their need. It is critical to highlight that most students entering the program were low progress readers. Low progress readers are those that are assessed as reading below their chronological age (Wheldall, 2009). In the FLFRPSP, this is evidenced by several sources including: NAPLAN scores well below the national average, initial placement testing in the program and other school-based indicators of student attendance and language backgrounds other than English that would contribute to poor reading progress. In setting the context for the FLFRPSP, the Government acknowledged that specific learning barriers

exist for students in remote locations 'inside and outside the school gate.' Those outside the school gate include housing uncertainty, health and employment concerns and high mobility of families following employment. Within the school gates, teacher turnover and LBOTE for most students make traditional learning a challenge. For some students, English is their eighth language. Therefore, consistent and effective reading instruction, particularly in the early years becomes difficult and students fail to grasp the fundamentals of literacy and learn at a rate compatible to their metropolitan peers.

Low progress readers are qualitatively different to normal readers in terms of their skills and hence, require specific instructional techniques and content to remediate their reading deficits. In particular, they often lack awareness of sounds in words and the ability to manipulate those sounds through segmentation or blending. These skills are seen as the basic building blocks of reading (Wheldall, 2009).

Conceptually, the FLFRPSP, or more specifically DI (and EDI) based literacy programs address the needs of struggling readers through their instructional method: direct and explicit instruction of material, repetition and progression only through mastery of preceding concepts. In addition, the program content is based around factors of effective reading instruction (phonics, phonemic awareness, fluency, vocabulary and comprehension) which have an extensive evidence base for their effectiveness with low progress readers (see Rose, 2006; Department of Education, Science and Training, 2005; National Institute of Child Health and Human Development, 2000). Hence, given the characteristics of the students within the program, theoretically the selection of DI and EDI based literacy approaches are appropriate to address the problem of low progress reading. This is true, not just because of the instructional approach (which is hotly debated in education circles) but also because of the tenets of effective reading instruction on which the programs are based (a point often lost in the debates on direct instruction).

The cultural appropriateness of DI for rural and remote students

When context is then brought into the frame, a recurring criticism or concern raised by school staff and others is the cultural appropriateness of the program content for indigenous students. Specific criticisms include the use of American based examples (ie 'trashcan' in place of 'rubbish bin') and the stories used as the reading materials. The scope of this concern needs further clarification as it was raised by a small proportion of the schools who responded to requests for interview and survey responses. Furthermore, in capturing perspectives from school staff, natural modifications and adaptations to better align the program from a cultural perspective are occurring in spite of this criticism. Hence, cultural misalignment of the program is acknowledged, with the impact of this misalignment variable according to the context. Some schools were vehement on this issue, others recognised it and adapted the program themselves, others were not concerned by the misalignment or had a student cultural profile that was not predominantly indigenous. For example, one school has a student cohort that is largely Vietnamese.

In addition, ICSEA values for program schools indicate the majority of families come from low SES which has shown to contribute to literacy difficulty. Home learning environment, particularly shared early reading are strongly associated with reading ability and general academic achievement in primary school.

Student attendance

The majority of schools in the program worked with the issue of low school attendance, particularly for Indigenous students. There was a significant, negative correlation between the percentage of Indigenous students and attendance rates suggesting that the problem is more associated with Indigenous students than non-Indigenous students. Gillan, Mellor and Krakouer (2017) suggest that the reasons for low Indigenous school attendance are complex and varied, ranging from disengagement from school to cultural and community commitments. They also suggest that increasing attendance rates would be difficult in the absence of specific remediating interventions. In the context of the FLFRPSP, our preceding report suggested that a strength of the program's rigid structure was that students with poor attendance could more seamlessly fit back into the curriculum and this was reiterated through the teacher survey.

Nonetheless, the issue of attendance remains a challenge for the program moving forward and contributes to issues associated with and beyond student learning. For example, every Principal interviewed in the NT cited student attendance as a barrier to student achievement, which was echoed by teachers. In addition, in the

survey to teachers, respondents were asked if they found it difficult to maintain record keeping for the program. Whilst the majority of respondents said 'no', those that did find record keeping difficult often commented that the difficulty lay in the fact that students did not attend regularly so it was harder to keep track of their progress. Hence problems with record keeping were not a function of the program but a function of student attendance. Given the pervasive nature of the issue, our recommendation is that the program implementers consider expectations of progress within the program that takes into account the specific attendance issues for each school. This has been evidenced in practice with one school reporting that they have a separate grouping for low attending students who work with Teaching Assistants on an almost 1:1 rotational basis. Schools with a high proportion of indigenous students should consider the possible available teaching days with consideration to community, culture and family and plan program progress based on these days rather than total school days per year.

In the original logic model developed with GGSA in 2015, student attendance was categorised as an outcome of the program. With a view of the program across three years, it is evident that poor attendance is a contextual feature and should be re-categorised as an external factor influencing the theory of change. Having gained a better understanding through evaluation of the program and other literature on indigenous contexts regarding the nature of student attendance in remote schools, the program would not be expected to have a significant impact on the improvement of student attendance. A small portion of improved student attendance could be attributed to aspects of the program delivery. For example, anecdotal evidence from schools that the program has provided some students with structure through which they can achieve literacy gains and the flow-on effect of this is greater engagement in the classroom and overall school attendance. Nonetheless, the profile of low student attendance in the current contexts is outside the remit of the program.

Summary

Earlier in the discussion the evidence base for reading programs using direct instruction and effective reading instruction practices was presented and showed moderate to large effect sizes on a range of reading measures. Therefore, the theory of change for the FLFRPSP is supported with extensive evidence for effectiveness in similar populations with similar social and academic issues. Despite this, the program has not worked universally suggesting that the reality of implementation has surfaced factors that have influenced positively or negatively the overall results. These are now discussed in detail.

Program

Resourcing and support (GGSA, Programmatic and System support)

One theme that surfaced through interviews and surveys with school staff was a lack of resourcing in some program schools. Several Principals noted that they were promised additional staff to implement and run the program but that these additional staff did not come to fruition. The impact of this appears to be multi-faceted with some school staff indicating that they could not run the program to fidelity because the teacher could only work directly with one group at a time hence other student groups did not make as much progress. Furthermore, school staff suggested that lack of resourcing added to staff stress around implementing the program and that student classroom behaviour was impacted due to lack of direct oversight. In addition to Principals, teachers, via the surveys also expressed some frustration at a lack of resourcing to run the program effectively. One teacher commented that a significant barrier to the program's success was that '2/3 of the class can't be supervised.' This was also a concern raised by the program providers who suggested that having enough staff members within the schools to deliver the program was crucial. Given the high teaching intensity of the program and its remedial nature, it is recommended that a resourcing needs analysis be conducted for schools remaining in the program to determine if funding can be directed to additional staffing requirements.

Programmatic support

GGSA provides support to schools implementing the FLFRPSP and acts as a conduit between Dataworks/NIFDI and the schools. School staff were specifically asked about factors they felt had contributed to the success of the program and the support from GGSA, NIFDI and Dataworks featured heavily in their responses. It seems the

role of these organisations is to ensure fidelity of implementation, troubleshoot issues and encourage ongoing commitment to the program. As one Principal commented,

> *"There has been a lot of training, and having the implementation manager come out a few times, she is always available on email- in that respect the support has been good."*

Nonetheless, in discussion with both representatives from the program developers and GGSA, several concerns were raised regarding implementation in the last 12 months. For example, GGSA have indicated that uncertainty around funding affected their resourcing and subsequent ability to support the schools in the second half of 2017. This is discussed in more detail below with regards to gaps in data for 2017. In addition, staff turnover at GGSA was a concern raised by the program providers and the constant attrition of institutional knowledge as a result of this turnover. As one participant commented, 'we are basically starting all over again this year."

In summary, the general trend with respect to programmatic support suggests it has been consistent from program developers and considered important to success by the schools. GGSA has also been able to provide consistent support through the early years of the program however, in the last 6 months, support appears to have waned slightly and this could be related to funding changes for the program.

School

As was discussed in the previous report, the FLFRPSP is a program implemented across diverse and complex contexts. This is all too evident in the data gathered in this evaluation regarding school demographics, student characteristics, teacher characteristics, all layered with a set of cultural, political and social needs specific to each school or jurisdiction. For example, assessment of the impact of the program, using NAPLAN revealed few relationships between other variables when considered at an aggregate level across the program. We can only surmise that complexity is at play here, diminishing any possible relationships between these variables. The result is an evaluation that uncovers pockets of success and failure but no consistent message as to why the success (or failure) has occurred. Factors that negatively impact one school, did not impact another school in the same way. Further, the collection of teacher and leadership voice indicated there was considerable variability in engagement with implementing the program.

Teacher retention

In the preceding report, the turnover of teachers in schools was highlighted as a major concern and that 'the program is being delivered by a perpetually learning workforce of teachers.' (Clinton, Dawson, McLaren & Koelle, 2017, pg. 116). The average percentage of teachers retained within a school is 42%, less than half the teaching staff over a school year. The impacts of teacher turnover extend beyond disruption to students in the classroom through to administrative time spent recruiting replacement staff as well as the induction and skilling up period required to have teachers fully functioning in the classroom. Other teacher turnover based issues raised by schools included complicated registration processes that delayed the ability for teachers to begin teaching. In addition, NIFDI representatives remarked on the issue of teacher turnover and commented on examples of schools where there was not the same teacher for three years running.

The Halsey report (2018) considered the issue of teacher retention in remote and country schools "one of the most persistent challenges on the education agenda." (Halsey, 2018, pg. 38). This is a significant workforce issue for the program (and any subsequent program) and one that needs to be addressed at a policy level. This aspect of the logic model places considerable pressure on the theory of change for the program. If the teaching staff continually change, training and upskilling needs to maintain the pace of turnover which is not always feasible or possible. Teacher training is run annually, hence any new teachers commencing within a given school year, must wait until the beginning of the next year for formal training and are consequently teaching DI/EDI as 'self-taught' until that training is conducted. GGSA and the program developers must consider this issue moving forward and consider additional training opportunities throughout the school year.

Leadership retention

In contrast to low teacher retention, Principal retention was relatively high across most schools. Ten of the 39 schools in the program during 2017 replaced a Principal with the remainder having no turnover of leadership. Survey responses reiterated the importance of leadership in terms of support for the program. Furthermore, program developers were strong in their advocacy for good supportive leadership in ensuring program success.

System

Community support and context

The levels of community support and its importance to the program was captured through the school staff interviews and surveys. There were several instances where community support was considered in a negative light in terms of its relationship to the program. Several Principals commented that the program does not fit with the school community and that school had been perceived as 'not fun anymore.' In addition, schools mentioned adapting the program to meet other community needs such as cutting back the time each day for the program to focus on learning on country. Others mentioned that they were working hard with the community to re-build initial negativity about the program in circumstances where literacy results were positive.

Government

Our federalised system of government in Australia provides the political environment for the FLFRPSP. The program was part of a 2013 election promise and formed a component of the Coalition's policy for schools: Students First aimed at improving amongst other things, literacy outcomes for students in remote areas. Nonetheless, the program's implementation is State and jurisdictionally based in schools with critical data to assess implementation and impact the responsibility of the schools, State governments and Catholic Education WA. This is a relatively unique but not unprecedented situation, which adds a layer of complexity to the implementation and the evaluation of the program.

Evaluation Readiness: Capacity and willingness to engage

Evaluations have an impact on programs through their data collection activities, capacity building, and relationships with key stakeholders and this also needs to be taken into account in the final review of impact.

One persistent challenge for the evaluation has been obtaining relevant data to make a full evaluative judgment on the worth of the FLFRPSP particularly with regards to student literacy outcomes. This dearth of data has come in many forms including:

- The data has been collected within the school but not shared due to jurisdictional preference
- The data has been collected within the school to some extent but access arrangements for the evaluation team have been cumbersome and prohibitive
- The data has not been collected at all.

In the original scope and briefing of the program, a key tenet was that schools enrolled in FLFRPSP would not have additional data collection burdens imposed. Whilst well intended, it is the view of the evaluation team that this tenet has worked against the success of the program due to the insufficiency of data available to understand the full extent of program impacts. Engagement and capacity to engage in evaluative processes are critical to support ongoing progress towards outcomes and sustainability of programs. They should not be considered mutually exclusive to program outcomes. A recent study by Clinton, (2014), examined over 300 evaluation cases which demonstrated that engagement in evaluation i.e., willingness and capacity, relates significantly to the sustainability of a program as well as program outcomes.

Literacy data.

The evaluation team were informed by select schools that significant literacy progress data (PAT-R) had been collected by teachers and was available for analysis. Furthermore, some schools indicated that they had conducted their own analysis on their literacy data and seen positive results for students. However, this data could not be provided directly to the evaluation team due to jurisdictional control, despite the schools asking for assistance and advice from the evaluation team on tracking student progress. This seems to be a missed

opportunity to capacity build around understanding impact within the schools, to bring the teachers into the evaluation in a meaningful way and to better understand the impact of the program overall. The evaluation team seeks to continue working with relevant jurisdictions to access and analyse literacy data more closely aligned to the program to fully understand the overall impact. This is particularly important for continuing schools and a key principle of the evaluation going forward will be to work with jurisdictions on data requirements.

Program based data.

In comparison to the completeness of data from previous years, data provided from GGSA did not include all four school terms with Term 4 2017 data unavailable for analysis. Further investigation with GGSA revealed several reasons for the reduced program data predominantly linked to funding concerns.

GGSA advised:

- A number of schools had received their allocated teaching amounts in semester one so had exhausted their support days. This suggests that schools potentially need more support than is currently funded.
- Teacher coaches were uncertain of future work and so two returned to their respective education departments that they had been on leave from
- The additional funding was less than previously allocated which resulted in reduced support largely being provided remotely in term 4. The schools subsequently provided less data to GGSA and without the regular oversight GGSA was unable to secure. The weekly data reviews continued but a large segment of this data was not forwarded to GGSA.

> *In summary, **implementation** was influenced by the dimensions of student, program, school and system but this influence was idiosyncratic for each school where a negative impact from one factor in one school may not be negative in another. In addition, the availability of data and engagement in evaluation has influenced the ability to determine the true impact of the program. The result is a paucity of observable diffuse impact for the FLFRPSP.*

The life course model: Taking all of these factors into account

The factors just presented and discussed determine the trajectory of the program's life course. The life course model presented here illustrates the flow of the program since its development several years ago (see Figure 61). In the previous evaluation it was suggested that if the program held steady on implementation and engagement then a positive progressive trajectory was predicted. Given the influence of factors in the past 12 months, this trajectory appears to have flat-lined. The diagram illustrates the previous path of the life course suggesting a promising trajectory, however when considered in aggregate, the program has struggled under contextual and resource pressures hence a projected decline in impact. Given the decrease in evaluation engagement and capacity it is difficult to determine the directional path of the program. However, it is suggested that a combination of program adaptation to context, and greater programmatic support is necessary to realign the trajectory in a more positive direction.

Situation

Students:
- Low performance in literacy
- Poor attendance
- Issues with behaviour

Teachers
- All trained
- Pedagogy and attendance variable
- Often not teaching for the allocated time

School leadership and communities
- Excuses made by schools, families and communities often affect lesson progress

Data
- Missing student and teacher fidelity data
- Lack of teacher fidelity

Inputs

Research

Invested time and money by participants and program facilitators

Materials and resources for program implementation

Pre-service training

In-service training

Coaching and mentoring

Peer practice

Fidelity observation

Instructional leadership

Community and school inputs

Outputs

Activities

Training delivery
- Pre-implementation training
- Pre-service training
- In-service training

Teaching
- Classroom organisation
- Teaching DI/EDI 2.5 hours
- Testing
- Data recording

Student grouping
- Placement testing
- Design of instructional groupings
- Scheduling of instruction

Teacher coaching
- Two and Five-minute observations
- Technical assistance observations
- Extended observations
- Demonstrations

Advisory group
- Reviews (with emphasis on scalability and systems improvement policy)
- Evaluation meetings and facilitation

Participation

Teachers and school leaders involved in DI/EDI training

Teachers implementing DI/EDI in classrooms on a regular basis and collecting data

Students attending school

Students attending DI/EDI lessons conducted at school

Schools and community supporting the implementation

State Departments providing system level data

Evaluators supporting GGSA to develop effective implementation practices and procedures

Outcome - Impact

Short
- Database development
- Identification of data sources
- Collection of data
- Set up of ethical data transfer and storing procedures
- Set up of communication pathways
- Identification of additional data needs
- Mutual development of program logic and success criteria
- Establish baseline statistics for the evaluation of the LRS program
- Analyse and interpret Year 1 data: student, system and teacher fidelity
- Report on Year 1 data

Medium
- 100% lesson completion (with support from community)
- 100% recoded absence data and disciplinary incident data recorded
- 100% of teachers submitting required data (fidelity)
- 100% teachers able to implement DI/EDI effectively
- 50% students at or above grade level (based on National benchmarks)
- - 50% of students at or above the 50th percentile on PAT-R, EYLND and PM Reading Benchmark measures.
- Improved student attendance (85%+)
- Less than 10% reported behaviour incidents

Long
- Full implementation of DI/EDI programs by participating schools, leaders and teachers
- 100% of teachers trained to standard
- Enhanced teacher pedagogical skills in Direct Instruction or Explicit Instruction.
- (improved teacher effectiveness)
- Improved literacy results for students in participating schools
- Improved student attendance
- Improved student behaviour

Assumptions

That the teachers' and school leaders participating in the program have the capability to implement DI or EDI with students in their classrooms and schools respectively.

External Factors

Positive influences: funding to support facilitation; familial support for the program.
Negative influences: irregular student attendance; movement of students between schools; attrition.
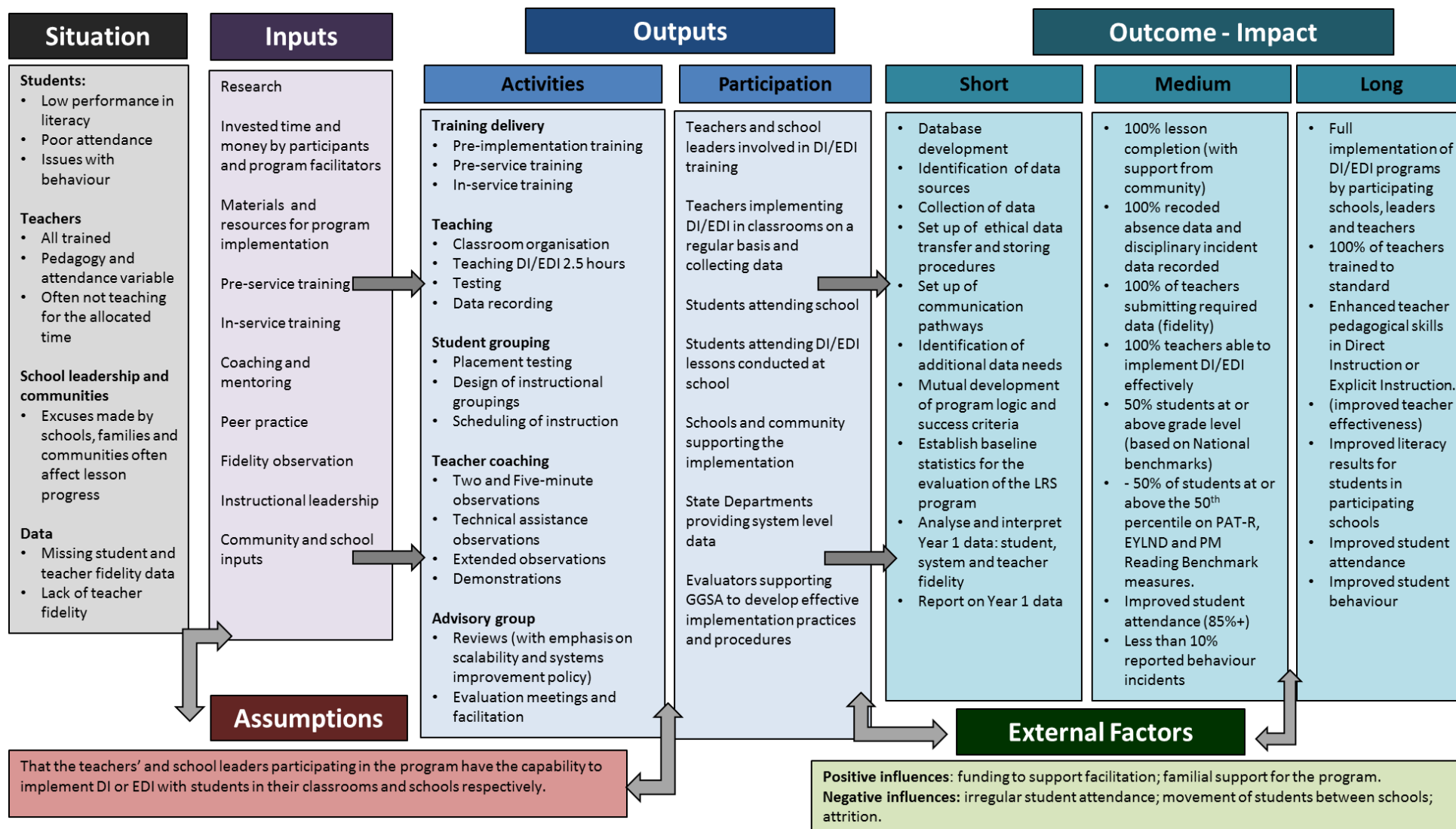
Figure 63: Original logic model developed with GGSA in October 2015.
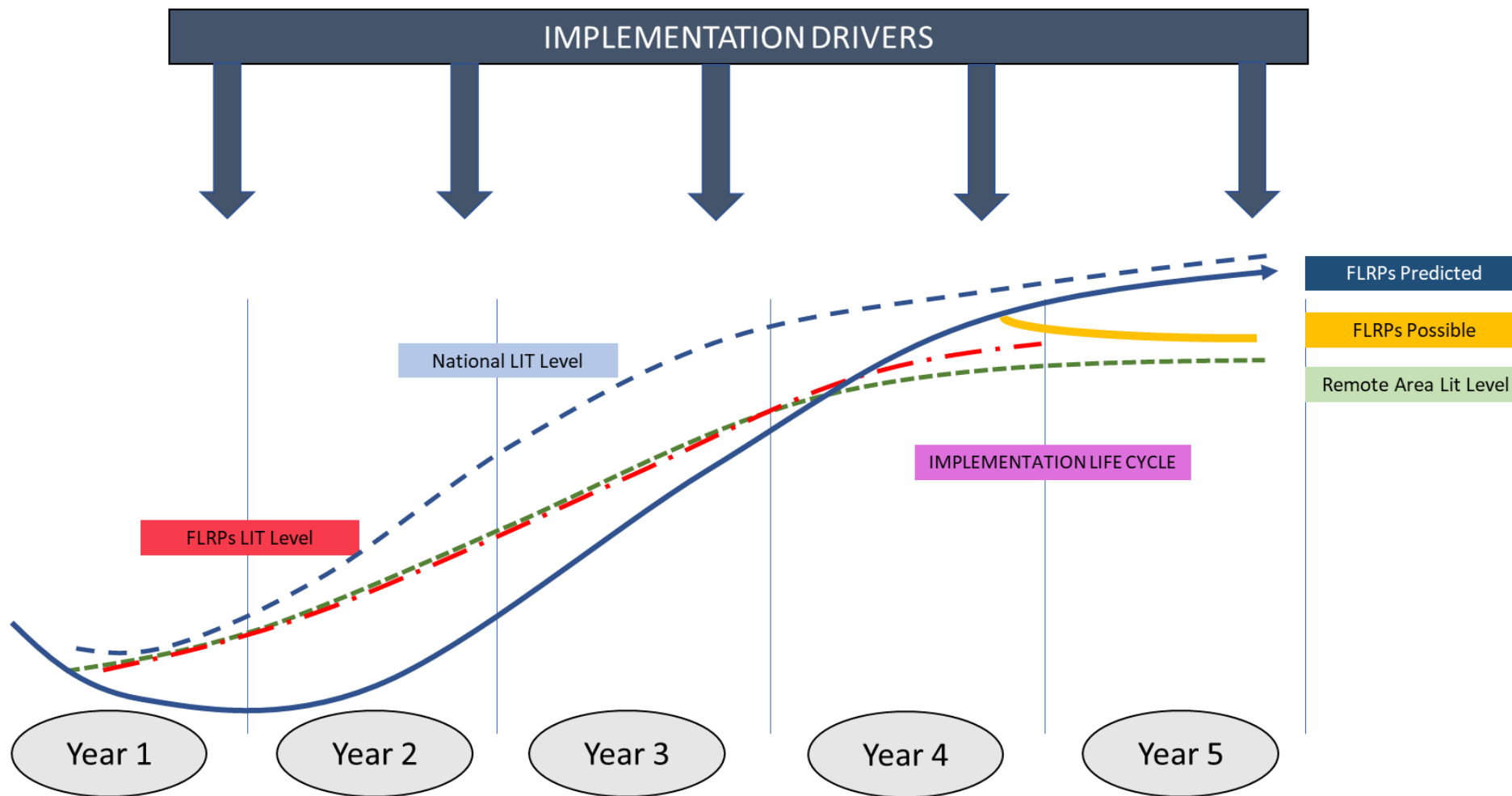
Figure 64: FLRPs life course model.

Understanding scale

Beyond understanding impact in the current implementation, an additional question at hand is whether there is enough substantive evidence of success to justify replication. What constitutes substantive will vary according to the nature of the program and the potential scope of the scale. At a minimum we would expect that a program has a strong theory of change, has demonstrated success in a number of localities, and robust monitoring and evaluation systems are in place. Scale also requires good program fidelity and an appropriate balance between standardisation and adaptability in context which can be challenging.

The question of scaling up is a vexing one for education, in fact scaled implementation of many innovative programs do not deliver the expected results. The question is why most education interventions do not scale, often beyond a single class or school; why system solutions rarely impact on the learning lives of students; why educational ideas rarely change whole systems? Elmore (1996) pinpointed fundamental tensions between district-based efforts for systemic improvement and the performance-based accountability demands.

> "These issues don't actually compromise or demean performance-base accountability," states Elmore. "They just pose problems for it –"

Elmore cautions districts (or in this case jurisdictions) as they wrestle with creating instructional change,

> "If you don't have a way of connecting instruction to management, organization, and accountability, you're behaving irresponsibly."

Elmore advocates focusing on the instructional core in schools. He cautions districts and policy makers,

> "If you push on an organization and you don't have a theory about how it shows up in teaching and learning, you're basically causing people to do rain dances." (p. 4).

In this case we have determined that the program has demonstrated success in several localities and it has a strong theory of change. Levels of program fidelity are increasing and there is recognition of evaluation readiness however the monitoring and evaluation systems are still variable which has resulted in a distinct lack of appropriate data. Over the past two years implementation and organisational structures have emerged however there are still concerns raised by the schools and providers that suggests the implementation system for FLFRPSP still needs consolidation. Furthermore, the evaluation has highlighted that resourcing in relation to funding and staff turnover has put pressure on the implementation system.

As suggested, scale is often demonstrated by an abundance of evidence that demonstrates a program will be effective and efficient within a general population. More often than not randomised controlled studies are put in place to determine scale, however a randomised controlled trial in the context of this program would not only be a poor fit but inappropriate. However, if we adopt the principle that randomised controlled trials purport we suggest that ultimately a program that is ready for scale can withstand some level resistance of a social, political, economic and cultural influence. In withstanding this resistance, the program must demonstrate fidelity of implementation allowing for some contextual adaptation all while maintaining a good level of success. The evaluation of the FLFRPSP has demonstrated in 2017/18 that the stability of the program has quavered in the face of resistance from economically, contextually and in some areas, politically driven variables. Hence, we would argue while the program has demonstrated pockets of significant success under certain conditions, it is not ready for scale. Given the possibility of success we would suggest that the program move towards developing the conditions necessary for scale by consolidating its implementation and addressing the implementation pressures around resourcing, programmatic support and community engagement.

4.3. Overall Statement and Recommendations

In summary, the evaluation team reiterates the findings of this report below and re-presents a revised logic model (see Figure 62) adjusted to reflect the evidence from the evaluation of the past 3 years. Finally, recommendations for the project going forward are presented as they relate to the FLFRPSP overall, the program characteristics and to implementation.

Impact Students

The evaluation of the FLFRPSP cannot claim that the program has produced substantive change in student literacy outcomes in aggregate. The reasons for this are:

- External program factors have created instability in the program's implementation and subsequent impacts. These factors include community support and resourcing as well as waning of jurisdictional support in some areas where the program is implemented.

- Engagement in evaluation has also waned in the last 12 months

- Data has not been available for all schools within the program for the evaluation

The evaluation can report that the program is showing sustained impact in Catholic Education WA as measured by specific literacy assessments and based on additive data collected since the last evaluation report in June 2017.

Impact teachers

Across all jurisdictions in the program, teacher effectiveness skills in direct instruction pedagogies is high, as measured by the program. However, there are mixed *perceptions* regarding the impact of the program on teachers instructional practice and teaching approaches more generally.

Implementation and context

While, direct instruction pedagogies are not suitable for all aspects of the learning lives of students, there is a strong evidence base for direct instruction in the teaching of mainstream skills such as literacy and numeracy. This is especially evident where students have made slow progress in mainstream schooling, a feature of many of the students in the schools implementing FLFRPSP. The theory of change for the program is strong but implementation reality is subject to differences between contexts in terms of engagement with the program, support and factors such as teacher turnover. This variability is creating a 'washing out' of impact such that only clusters of impact can be ascertained.

The logic model presented in Figure 62 demonstrates the current positioning of the program.
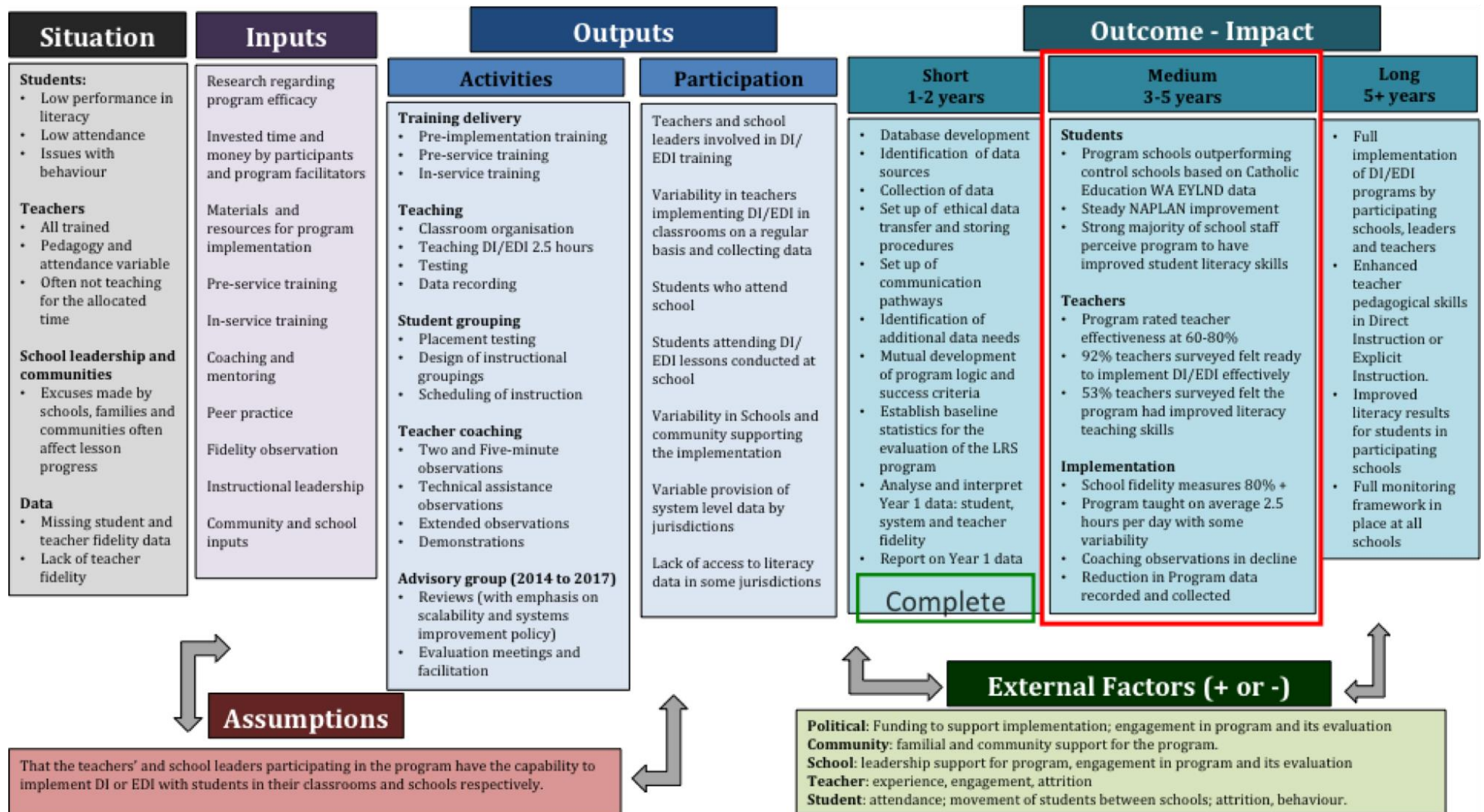
Situation

Students:
- Low performance in literacy
- Low attendance
- Issues with behaviour

Teachers
- All trained
- Pedagogy and attendance variable
- Often not teaching for the allocated time

School leadership and communities
- Excuses made by schools, families and communities often affect lesson progress

Data
- Missing student and teacher fidelity data
- Lack of teacher fidelity

Inputs

Research regarding program efficacy

Invested time and money by participants and program facilitators

Materials and resources for program implementation

Pre-service training

In-service training

Coaching and mentoring

Peer practice

Fidelity observation

Instructional leadership

Community and school inputs

Outputs

Activities

Training delivery
- Pre-implementation training
- Pre-service training
- In-service training

Teaching
- Classroom organisation
- Teaching DI/EDI 2.5 hours
- Testing
- Data recording

Student grouping
- Placement testing
- Design of instructional groupings
- Scheduling of instruction

Teacher coaching
- Two and Five-minute observations
- Technical assistance observations
- Extended observations
- Demonstrations

Advisory group (2014 to 2017)
- Reviews (with emphasis on scalability and systems improvement policy)
- Evaluation meetings and facilitation

Participation

Teachers and school leaders involved in DI/EDI training

Variability in teachers implementing DI/EDI in classrooms on a regular basis and collecting data

Students who attend school

Students attending DI/EDI lessons conducted at school

Variability in Schools and community supporting the implementation

Variable provision of system level data by jurisdictions

Lack of access to literacy data in some jurisdictions

Outcome - Impact

Short 1-2 years

- Database development
- Identification of data sources
- Collection of data
- Set up of ethical data transfer and storing procedures
- Set up of communication pathways
- Identification of additional data needs
- Mutual development of program logic and success criteria
- Establish baseline statistics for the evaluation of the LRS program
- Analyse and interpret Year 1 data: student, system and teacher fidelity
- Report on Year 1 data

Complete

Medium 3-5 years

Students
- Program schools outperforming control schools based on Catholic Education WA EYLND data
- Steady NAPLAN improvement
- Strong majority of school staff perceive program to have improved student literacy skills

Teachers
- Program rated teacher effectiveness at 60-80%
- 92% teachers surveyed felt ready to implement DI/EDI effectively
- 53% teachers surveyed felt the program had improved literacy teaching skills

Implementation
- School fidelity measures 80% +
- Program taught on average 2.5 hours per day with some variability
- Coaching observations in decline
- Reduction in Program data recorded and collected

Long 5+ years

- Full implementation of DI/EDI programs by participating schools, leaders and teachers
- Enhanced teacher pedagogical skills in Direct Instruction or Explicit Instruction.
- Improved literacy results for students in participating schools
- Full monitoring framework in place at all schools

Assumptions

That the teachers' and school leaders participating in the program have the capability to implement DI or EDI with students in their classrooms and schools respectively.

External Factors (+ or -)

Political: Funding to support implementation; engagement in program and its evaluation
Community: familial and community support for the program.
School: leadership support for program, engagement in program and its evaluation
Teacher: experience, engagement, attrition
Student: attendance; movement of students between schools; attrition, behaviour.

Figure 65: Revised logic model

Recommendations

In moving forward with a smaller subset of schools running the program in 2018 and given these schools have indicated a pedagogical preference for direct instruction approaches by choosing to continue with the program the following recommendations are proposed.

Overall

- Work with jurisdictions and government groups to develop stronger alignment between program and system policies and objectives
- Review resource needs in remaining schools and direct funding to areas of need to increase implementation
- Establish evaluation readiness and engagement with schools and jurisdictions
- Develop networks of schools for support, communities of program practice
- Move towards establishing the precursors for scale

Program Characteristics

- Increase flexibility of training and coaching model so that staff can be trained as they begin teaching DI or EDI
- Allow a more collaborative approach to program adaptation in schools that maintains fidelity but fits the program to context more appropriately

Implementation

- Explore building capacity of organisation structures within GGSA
- Work with GGSA to sustain rather than lose institutional knowledge due to staff turnover
- Actively support database use
- Consider working with the issue of attendance in schools rather than against it. How can the program fit around average attendance rates within schools?
- Consider record sharing across schools for itinerant students
- Consider ways of working with a fluctuating and very mobile workforce

A final word and the next phase of the evaluation

As previous evidence indicates, direct instruction can work and for the FLFRPSP it has worked in certain groups of students and shown that it has worked (Catholic Education WA). In other groups, we are unable to say if it has worked due to lack of supporting data. As with the previous report, the importance of context and implementation has arisen again. Context is crucial. As Harris and Jones (2018) suggest 'the process of policy implementation, in context, requires far more attention if the intended outcomes are to be achieved.' The FLFRPSP operates in multiple diverse contexts right down to the school level. This means that performance will be differentiated and influenced by many factors. Harris and Jones (2018) also propose that "explanations of differential performance need much more forensic attention and acknowledgement of cultural and contextual influences." It cannot be said more clearly than that. The evaluation has surfaced, defined and measured contextual factors. These factors mandate that the program must adapt. However, without additional substantive and proximal performance data, the evaluation cannot complete the final task of understanding has program has worked overall. To that end, we turn our attention to the final phase of the evaluation, working with a smaller group of schools and delving even further into both performance and context to understand the impact of this policy implementation on the literacy skills of students and the pedagogical skills of teachers.

5. References

Armbruster, B. B., Lehr, F., & Osborn, J. (2001). Put Reading First: The Research Building Blocks for Teaching Children To Read. Kindergarten through Grade 3.

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

Barth, A., & Elleman, A. (2017). Evaluating the Impact of a Multistrategy Inference Intervention for Middle-Grade Struggling Readers. Language, Speech, and Hearlng Services in Schools, 48, 31-41. doi:10.1044/2016

Batdi, V. (2016). A Meta-Analytic Comparison of Mastery Model and Traditional Methods? Effects on Academic Success, Retention, Achievement and Attitude. Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 1(40), 125-135. doi:10.21764/efd.18899

Beck, I. L., & McKeown, M. G. (1985). Teaching Vocabulary: Making the Instruction Fit the Goal. Educational Perspectives, 23(1), 11-15.

Bedard, C., Bremer, E., Campbell, W., & Cairney, J. (2018). Evaluation of a Direct-Instruction Intervention to Improve Movement and Preliteracy Skills among Young Children: A Within-Subject Repeated-Measures Design. Frontiers in Pediatrics, 5(January), 1-9. doi:10.3389/fped.2017.00298

Begeny, J. C., Levy, R. A., & Field, S. A. (2017). Using Small-Group Instruction to Improve Students' Reading Fluency: An Evaluation of the Existing Research. Journal of Applied School Psychology, 34(1), 36-64. doi:10.1080/15377903.2017.1328628

Benner, G., Nelson, R., Ralston, N., & Mooney, P. (2010). A Meta-Analysis of the Effects of Reading Instruction on the Reading Skills of Students With or at Risk of Behavioral Disorders. Behavioral Disorders, 35(2), 86-102.

Bereiter, C., & Engelmann, S. (1966). Observations on the use of direct instruction with young disadvantaged children. Journal of School Psychology, 4(3), 55-62.

Biemiller, A. (2001). Teaching Vocabulary: Early, Direct, and Sequential. American Educator, Spring.

Bloom, B. S. (1968). Learning for Mastery. Retrieved from

Brownson, R. C., Colditz, G. A., & Proctor, E. K. (2012). Dissemination and Implementation Research in Health. Ney York, NY: Oxford University Press.

Buckingham, J., Wheldall, K., & Beaman, R. (2012). A randomised control trial of a Tier-2 small-group intervention ('MiniLit') for young struggling readers. Australian Journal of Learning Difficulties, 17(2), 79-99. doi:10.1080/19404158.2012.717537

Buckingham, J., Beaman-Wheldall, R., & Wheldall, K. (2014). Evaluation of a two-phase experimental study of a small group ("MultiLit") reading intervention for older low-progress readers. Cogent Education, 1(1), 962786.

Bus, A., & van Ijzendoorn, M. (1999). Phonological Awareness and Early Reading: A Meta-Analysis of Experimental Training Studies. Journal of Educational Psychology, 91(3), 403-414.

Cape York Aboriginal Australian Academy. (2016). Annual Report 2016. Retrieved from

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The Foundations of Spelling Ability: Evidence from a 3-Year Longitudinal Study. Journal of Memory and Language, 45(4), 751-774. doi:10.1006/jmla.2000.2785

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? Cognition, 91(1), 77-111. doi:10.1016/s0010-0277(03)00164-1

Castles, A., Coltheart, M., Wilson, K., Valpied, J., & Wedgwood, J. (2009). The genesis of reading ability: What helps children learn letter-sound correspondences? Journal of Experimental Child Psychology, 104(1), 68-88. doi:10.1016/j.jecp.2008.12.003

Centers for Disease Control and Prevention. (1999). Framework for Program Evaluation in Public.

Centre for Education Statistics & Evaluation. (2017). Effective Reading Instruction in the Early Years of School. Retrieved from Sydney:

Chard, D., Vaughn, S., & Tyler, B.-J. (2002). A Synthesis of Research on Effective Interventions for Building Reading Fluency. Journal of Learning Disabilities, 35(5), 386-406.

Chase, C. C., & Klahr, D. (2017). Invention Versus Direct Instruction: For Some Content, It's a Tie. Journal of Science Education and Technology, 26(6), 582-596. doi:10.1007/s10956-017-9700-6

Chitiyo, M., Makweche-Chitiyo, P., Park, M., Ametepee, L. K., & Chitiyo, J. (2011). Examining the effect of positive behaviour support on academic achievement of students with disabilities. Journal of Research in Special Educational Needs, 11(3), 171-177. doi:10.1111/j.1471-3802.2010.01156.x

Clinton, J. (2014). The True Impact of Evaluation: Motivation for ECB. American Journal of Evaluation, 35(1), 120-127.

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.

Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. ISBN 1-134-74270-3.

Collins, D., & Collins, A. (2004). A closer look at the five essential components of effective reading instruction: A review of scientifically based reading research for teachers. Learning Point Associates, 630-649.

Cunningham, J. W. (2001). The National Reading Panel Report. Reading Research Quarterly, 36(3), 326-335. doi:10.1086/509524

Darch, C., Eaves, R., Crowe, D. A., Simmons, K., & Conniff, A. (2006). Teaching Spelling to Students With Learning Disabilities A Comparison of Rule-Based Strategies Versus Traditional Instruction. Journal of Direct Instruction, 6(1), 1-16.

Datchuk, S. M. (2017). A Direct Instruction and Precision Teaching Intervention to Improve the Sentence Construction of Middle School Students With Writing Difficulties. Journal of Special Education, 51(2), 62-71. doi:10.1177/0022466916665588

De Nigris, R. P. (2017). The Effect of Explicit Instruction for Story Grammar Code Strategy on Third Graders' Reading Comprehension.

De Nobile, J., London, T., & El Baba, M. (2015). Whole school behaviour management and perceptions of behaviour problems in Australian primary schools. Management in Education, 29(4), 164-171. doi:10.1177/0892020615589135

Department of Education, S., & Training. (2005). Teaching Reading: Report and Recommendations (9780642775771 064277577X 9780642775757 0642775753). Retrieved from http://www.dest.gov.au/nitl/report.htm

Dunlap, G., Carr, E. C., Homer, R. H., Koegel, R. L., Sailor, W., Clarke, S., . . . Fox, L. (2010). A Descriptive , Multiyear Examination of Positive Behavior Support University of Kansas University of Kansas University of California at Davis University of South Florida. 35(August), 259-279.

Ehri, L., & Nunes, S. (2002). Systematic Phonics Instruction Helps Students Learn to Read: Evidence From the National Reading Panel's Meta-Analysis. Journal of Direct Instruction, 2(2), 121-166.

Ehri, L., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic Phonis Instruction Helps Students Learn to Read: Evidence from the National Reading Panel's Meta-Analysis. Review of Educational Research, 71(3), 393-447.

Elleman, A. M., & Compton, D. L. (2017). Beyond Comprehension Strategy Instruction: What's Next? Language Speech and Hearing Services in Schools, 48(2), 84-84. doi:10.1044/2017_LSHSS-16-0036

Elmore, R. F. (1996). Getting to scale with good educational practice. Harvard Educational Review, 66(1), 1-26.

Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2013). Statewide Implementation of Evidence-Based Programs. Exceptional Children, 79(3), 213-230.

Fulwiler, G., & Groff, P. (1980). The Effectiveness of Intensive Phonics. Reading Horizons, 21(1), 50-54.

Garan, E. (2001). Beyond the Smoke and Mirrors: A Critique of the National Reading Panel Report on Phonics. Phi Delta Kappa, 500-506. doi:10.1177/003172170108200705

Gillan, K. P., Mellor, S., & Krakouer, J. (2017). The Case for Urgency: Advocating for Indigenous voice in education.

Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? Learning and Instruction, 51, 26-35. doi:10.1016/j.learninstruc.2016.11.002

Grey, I., Lydon, H., & Healy, O. (2016). Journal of Intellectual & Developmental Disability Positive behaviour support: What model of disability does it represent? Positive behaviour support: What model of disability does it represent? , 41(3), 255-266. doi:10.3109/13668250.2016.1164304

Guskey, T. (2007). Closing Achievement Gaps. Journal of Advanced Academics, 19(1), 8-31.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., . . . Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. Journal of Educational Psychology, 96(3), 403-423. doi:10.1037/0022-0663.96.3.403

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. routledge.

Hattie, J. (2016). *Jack Keating Memorial Lecture*. University of Melbourne, Melbourne Australia.

Hart, R. (2010). Classroom behaviour management: Educational psychologists' views on effective practice. Emotional and Behavioural Difficulties, 15(4), 353-371. doi:10.1080/13632752.2010.523257

Hayes, B. (2008). Introductory Phonology. USA: Wiley-Blackwell.

Halsey, J. (2018). Independent Review into Regional, Rural and Remote Education–Final Report.

Hernandez, H. P. (2017). Direct Instruction in Second Language Acquisition: A critical review of related literature. Journal on English Language Teaching, 7(1), 43-57.

Hill, D., & Brown, D. (2013). Supporting inclusion of at risk students in secondary school through positive behaviour support. International Journal of Inclusive Education, 17(8), 868-881. doi:10.1080/13603116.2011.602525

Hughes, C. A., Morris, J. R., Therrien, W. J., & Benson, S. K. (2017). Explicit Instruction: Historical and Contemporary Contexts. Learning Disabilities Research and Practice, 32(3), 140-148. doi:10.1111/ldrp.12142

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A Meta-Analytic Review of Achievement Goal Measures: Different Labels for the Same Constructs or

Different Constructs With Similar Labels? Psychological Bulletin, 136(3), 422-449. doi:10.1037/a0018947

Hulme, C., Snowling, M. J., Caravolas, M., & Carroll, J. (2005). Phonological Skills Are (Probably) One Cause of Success in Leraning to Read: A Comment on Castle and Coltheart. Scientific Studies of Reading, 9(4), 351-365.

Instruction, N. I. f. D. (2018). Basic Philosophy of Direct Instruction (DI). Retrieved from https://www.nifdi.org/what-is-di/basic-philosophy

Johnston, R., & Watson, J. (2005). The effects of synthetic phonics teaching on reading and spelling attainment: a seven year longitudinal study.

Joshi, R. M., Binks, E., Graham, L., Ocker-Dean, E., Smith, D. L., & Boulware-Gooden, R. (2009). Do textbooks used in university reading education courses conform to the instructional recommendations of the national reading panel? Journal of Learning Disabilities, 42(5), 458-463. doi:10.1177/0022219409338739

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of Three Components of Reading Fluency to Reading Comprehension. Journal of Educational Psychology, 100(2), 310-321. doi:10.1037/0022-0663.100.2.310

Konza, D. (2014). Teaching Reading: Why the "Fab Five" should be the "Big Six". Australian Journal of Teacher Education, 39(39). doi:10.14221/ajte.2014v39n12.10

Krashen, S. (2001). More Smoke and Mirrors: A Critique of the National Reading Panel (NRP) Report on "Fluency". In.

Kuder, S. J. (2017). Vocabulary Instruction for Secondary Students with Reading Disabilities: An Updated Research Review. Learning Disability Quarterly, 40(3), 155-164. doi:10.1177/0731948717690113

Kulik, C.-L., Kulik, J., & Bangert-Drowns, R. (1990). Effectiveness of Mastery Learning Programs : A Meta-Analysis. Review of Educational Research, 60(2), 265-299.

Langenberg, D. (2018). Findings of the National Reading Panel. Retrieved from http://www.readingrockets.org/article/findings-national-reading-panel

Learning, E. f. (2018). MiniLit - active project. Retrieved from http://evidenceforlearning.org.au/lif/our-projects/minilit/

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association, 83(404), 1198-1202.

Llewellyn, L., Boon, H., & Lewthwaite, B. (2018). Effective behaviour management strategies for Australian Aboriginal and Torres Strait Islander students: A Literature review. Australian Journal of Teacher Education, 43(1). doi:10.14221/ajte.2018v43n1.1

López, P., Torrance, M., Rijlaarsdam, G., & Fidalgo, R. (2017). Effects of direct instruction and strategy modeling on upper-primary students' writing development. Frontiers in Psychology, 8(JUN), 1-10. doi:10.3389/fpsyg.2017.01054

Maddox, K., & Feng, J. (2013). Whole Language Instruction vs. Phonics Instruction: Effect on Reading Fluency and Spelling Accuracy of First Grade Students.

Mastropieri, M. A., Scruggs, T. E., & Graetz, J. E. (2003). Reading Comprehension Instruction for Secondary Students : Challenges for Struggling Students and Teachers. 26(2), 103-116.

McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking Reading Comprehension. Reading Research Quaterly, 44(3), 218-253. doi:dx.doi.org/10.1598/RRQ.44.3.1

Miles, K. (2010). Mastery Learning and Academic Achievement.

MultiLit. (2018). MiniLit Program. Retrieved from http://www.multilit.com/programs/minilit-program/

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. NIH Publication No. 00-4769, 7, 35-35. doi:10.1002/ppul.1950070418

O'Connor, R. E. (2018). Reading Fluency and Students With Reading Disabilities: How Fast Is Fast Enough to Promote Reading Comprehension? J Learn Disabil, 51(2), 124-136. doi:10.1177/0022219417691835

Özokcu, O., Akçamete, G., & Özyürek, M. (2017). Examining the Effectiveness of Direct Instruction on the Acquisition of Social Skills of Mentally Retarded Students in Regular Classroom Settings. Journal of Education and Training Studies, 5(4), 214-214. doi:10.11114/jets.v5i4.2294

Pearson, P. D. (2009). The Roots of Reading Comprehension Instruction. Handbook of research on reading comprehension(June), 347-372.

Piasta, S. B., & Wagner, R. K. (2010). Developing Early Literacy Skills: A Meta-Analysis of Alphabet Learning and Instruction. Read Res Q, 45(1), 8-38. doi:10.1598/RRQ.45.1.2

Richmond, A. S., Bacca, A. M., Becknell, J. S., & Coyle, R. P. (2017). Teaching Metacognition Experientially. Teaching of Psychology, 44(4), 298-305. doi:10.1177/0098628317727633

Rose, J. (2006). Independent Review of the Teaching of Early Reading. London: DfES.

Rowe, K. (2005). Teaching reading: Report of the National Inquiry into the Teaching of Literacy. *Canberra: Department of Education, Science, and Training*.

Scott, B. (2005). Getting to the Heart of the Matter: Examining the Efficacy of a Whole-School Approach to Behaviour Management. Kairaranga, 6(1), 29-34.

Sedita, J. (2005). Effective Vocabulary Instruction. Insights on Learning Disabilities, 2(1), 33-45.

Senko, C., Hulleman, C. S., & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. Educational Psychologist, 46(1), 26-47. doi:10.1080/00461520.2011.538646

Sermier Dessemontet, R., & de Chambrier, A. F. (2015). The role of phonological awareness and letter-sound knowledge in the reading development of children with intellectual disabilities. Research in Developmental Disabilities, 41-42, 1-12. doi:10.1016/j.ridd.2015.04.001

Shanahan, T. (2006). The National Reading Panel Report. Practical Advice for Teachers. Learning Point Associates/North Central Regional …(1), 41-47.

Slavin, R. E. (1987). Mastery Learning Reconsidered. Review of Educational Research, 57(2), 175-213. doi:10.3102/00346543057002175

Slavin, R. E. (1990). Mastery Learning Re-Reconsidered. Review of Educational Research, 60(2), 300-302. doi:10.3102/00346543060002300

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

Soiferman, L. K. (2017). Adapting Effective Writing Instruction 1. (2007), 1-37.

Spörer, N., Brunstein, J. C., & Kieschke, U. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. Learning and Instruction, 19(3), 272-286. doi:10.1016/j.learninstruc.2008.05.003

Stahl, S., & Fairbanks, M. (1986). The Effects of Vocabulary Instruction: A Model-Based Meta-Analysis. Review of Educational Research, 56(1), 72-110.

Stockard, J., & Wood, T. W. (2017). The Threshold and Inclusive Approaches to Determining "Best Available Evidence": An Empirical Analysis. American Journal of Evaluation, 38(4), 471-492. doi:10.1177/1098214016662338

Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research. Review of Educational Research. doi:10.3102/0034654317751919

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

Madaus, G F, & T. Kellaghan (Eds.), (2nd ed.). Boston: Kluwer Academic Publishers.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., . . . Ruef, M. B. (2000). Applying Positive Behavior Support and Functional Behavior Assessment in Schools. Journal of Positive Behavior Interventions, 2, 131-143. doi:10.1177/109830070000200302

Swain, K., Elizabeth, L.-J., & Conley, P. (2013). Effects of Repeated Reading and Listening Passage Preview on Oral Reading Fluency. Reading Improvement, 5(1), 12-19.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta analysis. Remedial and Special Education, 25, 252-261.

U.S. Department of Education. (2014). Repeated Reading. Retrieved from

U.S. Department of Education. (2015). Spelling Mastery. Retrieved from

Vadasy, P., Sanders, E., & Nelson, R. (2015). Effectiveness of Supplemental Kindergarten Vocabulary Instruction for English Learners. Journal of Research on Educational Effectiveness, 8(4), 490-529. doi:10.1080/19345747.2015.1033120

Wagner, R. K., Kim, Y. S., & Foster, E. (2011). Relations Among Oral Reading Fluency, Silent Reading Fluency, and Reading Comprehension: A Latent Variable Study of First- Grade Readers. Scientific Studies of Reading, 15(4), 338-362. doi:10.1080/10888438.2010.493964.Relations

Wheldall, K. (2009). Effective instruction for socially disadvantaged low-progress readers: The Schoolwise Program. *Australian Journal of Learning Difficulties, 14*(2), 151-170.

White, W. (1988). A Meta-Anaylsis of the Effects of Direct Instruction in Special Education. Education and Treatment of Children, 11(4), 364-374.

Wimmer, H., Landerl, K., Linortner, R., & Hummer, P. (1991). The relationship of phonemic awareness to reading acquisition: More consequence than precondition but still important. Cognition, 40(3), 219-249. doi:10.1016/0010-0277(91)90026-Z

Wolf, M., & Katzir-Cohen, T. (2001). Reading Fluency and Its Intervention. Scientific Studies of Reading, 5(3), 211-239. doi:10.1207/S1532799XSSR0503

Wright, T. S., & Cervetti, G. N. (2017). A Systematic Review of the Research on Vocabulary Instruction That Impacts Text Comprehension. Reading Research Quarterly, 52(2), 203-226. doi:10.1002/rrq.163

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

Zhang, D. (2017). Effects of visual working memory training and direct instruction on geometry problem solving in students with geometry difficulties. 15(1 OP - Learning Disabilities: A Contemporary Journal. Spring, 2017, Vol. 15 Issue 1, p117, 22 p.), 117-117.

Zhao, Y. (2017). What works may hurt: Side effects in education. Journal of Educational Change, 18(1), 1-19. doi:10.1007/s10833-016-9294-4

6. Appendices

Appendix 1: Interview Protocol

Non-teaching stakeholders (GGSA, Government, NIFDI)
Introduction
1. Tell me about your role within the organisation?
2. How long have you been with this organisation?
3. Tell me about your involvement with the FLRPS?

Implementation
1. I'm interested in your feedback on the FLRPS
    a. If you think about the implementation of the program in schools, what do you think has worked well?
    b. What do you think has not worked so well?
2. From your perspective, what are the most critical aspects of the program that facilitate implementation?
3. Which aspects, if any, could be removed?
4. From your perspective, do you think that some schools have implemented the program more successfully than others? If yes, without naming the schools, can you identify the factors within those schools that have made the program implementation more successful?
5. Conversely, without naming the schools, can you identify the barriers to implementation in less successful schools?

Outcomes/Impact
1. From your perspective what do you consider the key indicators of success of the FLRPS program? What should we be seeing in schools implementing the program?
    a. Teaching practices
    b. Student behaviour
    c. Student learning
2. What changes have you observed in schools since the program was implemented?

Teachers/Principals

Opening question:

1. How long have you been using the DI/EDI approach to teaching literacy?

Fidelity

2. Take us through a typical literacy session
   a. What happens at the beginning, middle and end?
   b. Do sessions run at the same time each day?
   c. How long do you generally spend teaching literacy each day?
      i. Is this all DI/EDI or do you integrate other components?
   d. If there are teachers' aides- how many? Are they in everyday? What is there role in teaching literacy
3. Do you follow the script exactly, or have you changed it to better suit your students?
   a. If yes, how has that worked out?
4. Do you follow the order of lessons, or have you adapted them?
5. How do you deal with students missing classes?
6. How do you deal with students progressing at a different rate?

Perspective of the program

7. From your perspective, what factors are critical to the program's success?
8. Has the program benefited your school community, in what ways?

Perspective of student change

9. In your opinion, how do students feel about the way literacy is taught in your school?
10. Have student's literacy levels changed since you began using the program?
    a. For all students or just for some?
       i. Why do you that students differ?
    b. What aspects of the program are most critical for student literacy improvement?

Perspective of teacher change

11. How has the way you teach literacy changed since you started using the EDI/DI approach?
    a. What were you doing previously?
    b. Have you integrated these teaching strategies into other areas?

Implementation:

12. What was the process of implementation like?
    a. What was the training like?
    b. How long until it became embedded in the school?
    c. How long until you felt proficient and confident?
13. Has there been anything that has helped you and your school with program implementation/maintenance?

Perspective of the program

14. Are the program outcomes what you were expecting?
15. From your perspective, what factors are critical to the program's success?
16. Has the program benefited your school community, in what ways?
17. Have there been any negative impacts of the program?
18. Is there anything else you would like to add about the FLRSP that has not been covered?

Appendix 2: NVivo Nodal Structure

Summary of final NVivo nodal structure used by coders in the overall thematic reporting.

Parent Node	First Level Child Node	Second Level Child Node
Perceptions of program impact on students	Students literacy improving because of the program	Oral language
		Phonics
		Vocabulary
		Reading
		Writing
	Students literacy not improving because of the program	Oral language
		Phonics
		Vocabulary
		Reading
		Writing
	Student literacy negatively impacted because of the program	Oral language
		Phonics
		Vocabulary
		Reading
		Writing
	Students not engaged with the program	
	Students engaged with the program	
	Student engagement unchanged by program	
	Student behaviour positively impacted by program engagement	
	Student behaviour negatively impacted by program engagement	
	Student perception of program positive	
	Student perception of program negative	
	Student self-efficacy increase	

	Student self-efficacy decrease	
Perceptions of program impact on practitioners	Teachers find program structure positive	Time allocation to literacy
		Prescribed practice (i.e. cues, rotation of groups etc)
	Teachers find program structure negative	Time allocation to literacy
		Prescribed practice (i.e. cues, rotation of groups etc)
	Teachers find program materials useful/practical	Matching cohort needs (pre-existent literacy levels and sociocultural needs)
		Accessibility of resources (online or hardcopy)
	Teachers find program materials impractical	Matching cohort needs (pre-existent literacy levels and sociocultural needs)
		Accessibility of resources (online or hardcopy)
	Teacher improvement in pedagogical practice	
	Teacher decline in pedagogical practice	
	Teacher self-efficacy increase	
	Teacher self-efficacy decrease	
	Principal leadership improved	
	Principal leadership weakened	
Perceptions of implementation	Fidelity	*Change* – Community- Adapting program to fit local context
		Change - Student- Adapting program structure to suit student needs
		Change -Teacher- Adapting program structure to suit teacher needs
		Withdrawal from program – Community
		Withdrawal from program – Student
		Withdrawal from program—Teacher

		No change
		Teacher attitude/commitment
		Teacher turnover
		Teacher quality
		Teacher knowledge and training
		Teacher support (administration and classroom)
	Barriers	Student attendance
		Student behavioural management
		Student pre-existent literacy levels
		Student sociocultural needs
		School leadership engagement
		School leadership turnover
		School financial resourcing
		School location
		Community program commitment ('buy in')
		Community context (unrest/volatility)
		Program timelines (for implementation)
		Program supports (GGSA, government, NIFDI)
		Program administration and materials
		Program prescribed classroom structure
		Program leadership
		Teacher attitude/commitment
		Teacher turnover
		Teacher quality
		Teacher knowledge and training

		Teacher support (administration and classroom)
		Student attendance
		Student behavioural management
		Student pre-existent literacy levels
		Student sociocultural needs
		School leadership engagement
		School leadership turnover
		School financial resourcing
		School location
		Community program commitment ('buy in')
	Enablers	Program timelines (for implementation)
		Program supports (GGSA, government, NIFDI)
		Program administration and materials
		Program prescribed classroom structure
		Program leadership

Appendix 3: NAPLAN Average Scores over Time, NT Government Program Schools

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, 2016 and 2017. A series of one-way ANOVAs were 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 such, sensitivity testing was also conducted using Kruskal-Wallis H tests.

The results indicate a statistically significant difference over 2014 to 2017 in the Spelling scores of Year 5 students (*Figure 67*), $F(3, 52) = 5.65$, $p = .002$, with students performing better on spelling in 2017 on average. However, there were no significant difference over 2014 to 2017 for the other NAPLAN domains in the NT government cohort for Year 3 (Figure 66) or Year 5 (*Figure 67*)[18].
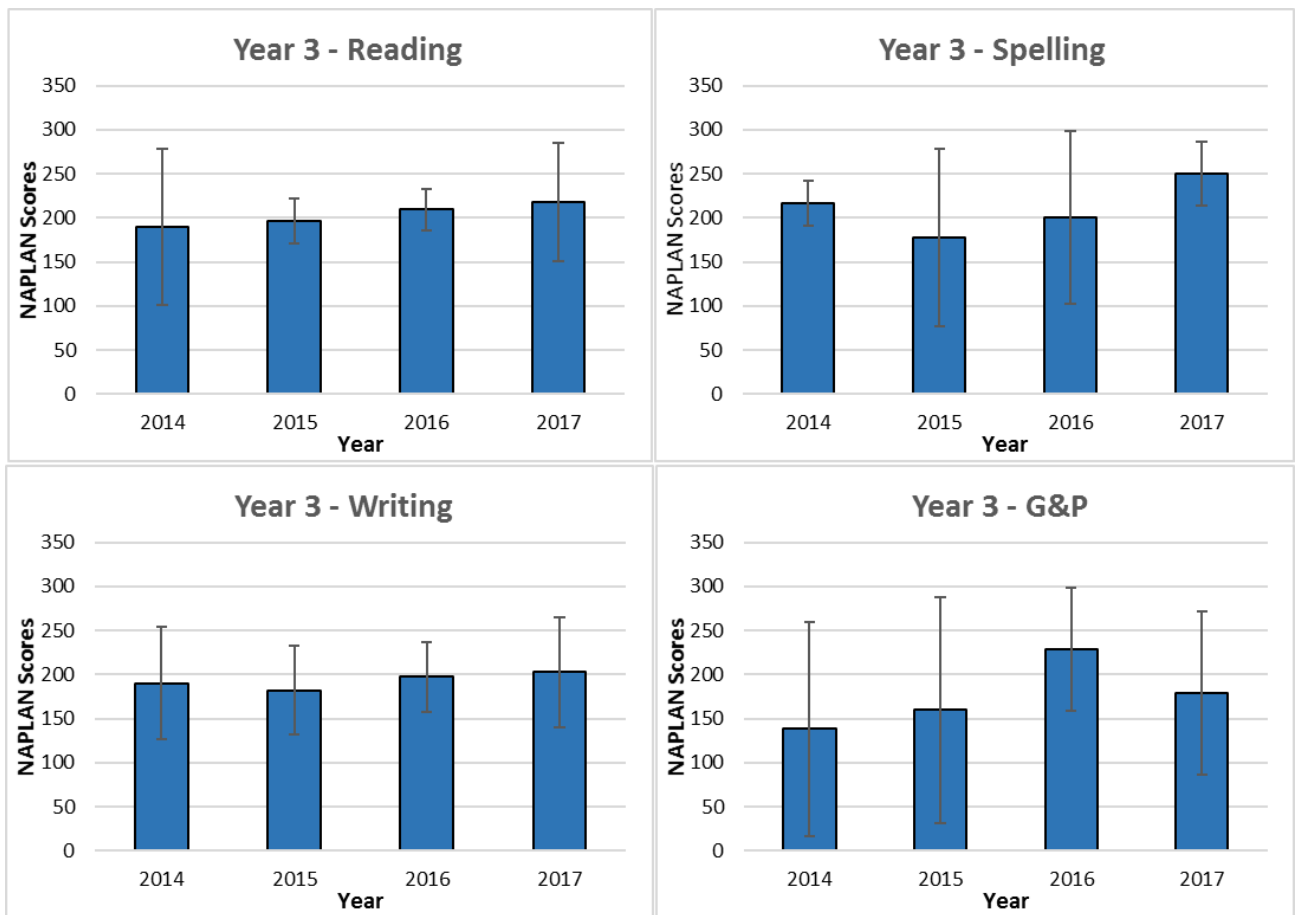

Figure 66: NT government mean NAPLAN scores for Year 3, 2014 -2016.

---

[18]Year 3: Reading $F(3, 52) = 0.48$, $p = .60$; Writing $F(3, 52) = 0.23$, $p = .88$; Spelling $F(3, 52) = 2.12$, $p = .11$; Grammar and Punctuation $F(3, 52) = 1.88$, $p = .15$; Year 5: Reading $F(3, 51) = 2.11$, $p = .11$; Writing $F(3, 52) = 0.89$, $p = .45$; Grammar and Punctuation $F(3, 52) = 1.88$, $p = .15$.
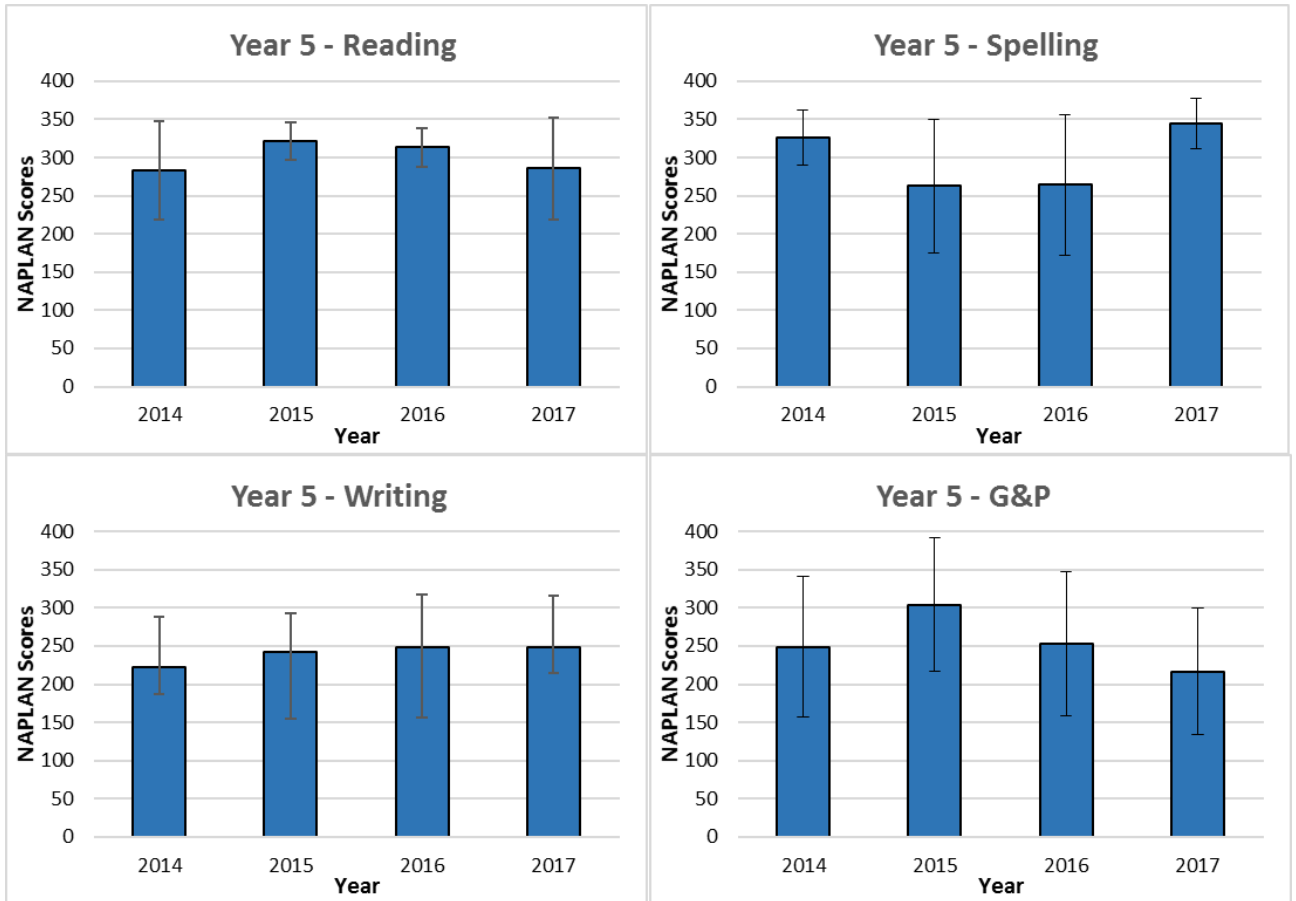
Figure 67: NT government mean NAPLAN scores for Year 5, 2014-2017

Appendix 4: Percentage of Students below NMS, NT Government Program Schools

Analysis of the change in the percentage of students below National Minimum Standard (NMS) from 2014 to 2017 for Years 3 and 5 are presented in *Figure 68* and *Figure 69* respectively.

For Grade 3, the results are mixed. Students below the NMS in Reading, Writing, and Grammar and Punctuation showed an initial decrease in 2015, then an increase in 2017. However, Spelling demonstrated an increase in percentage of students below NMS through 2014 to 2017.

In general, Grade 5 demonstrated quite high percentages of students below NMS. Students below the NMS in Reading and Spelling showed a decrease in 2017; however, percentages fluctuated over these four years. Overall, these results do not suggest that there have been any material improvements in the percentage of students below NMS according to NAPLAN testing. However, 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, these scores may not be expected to have changed appreciably.
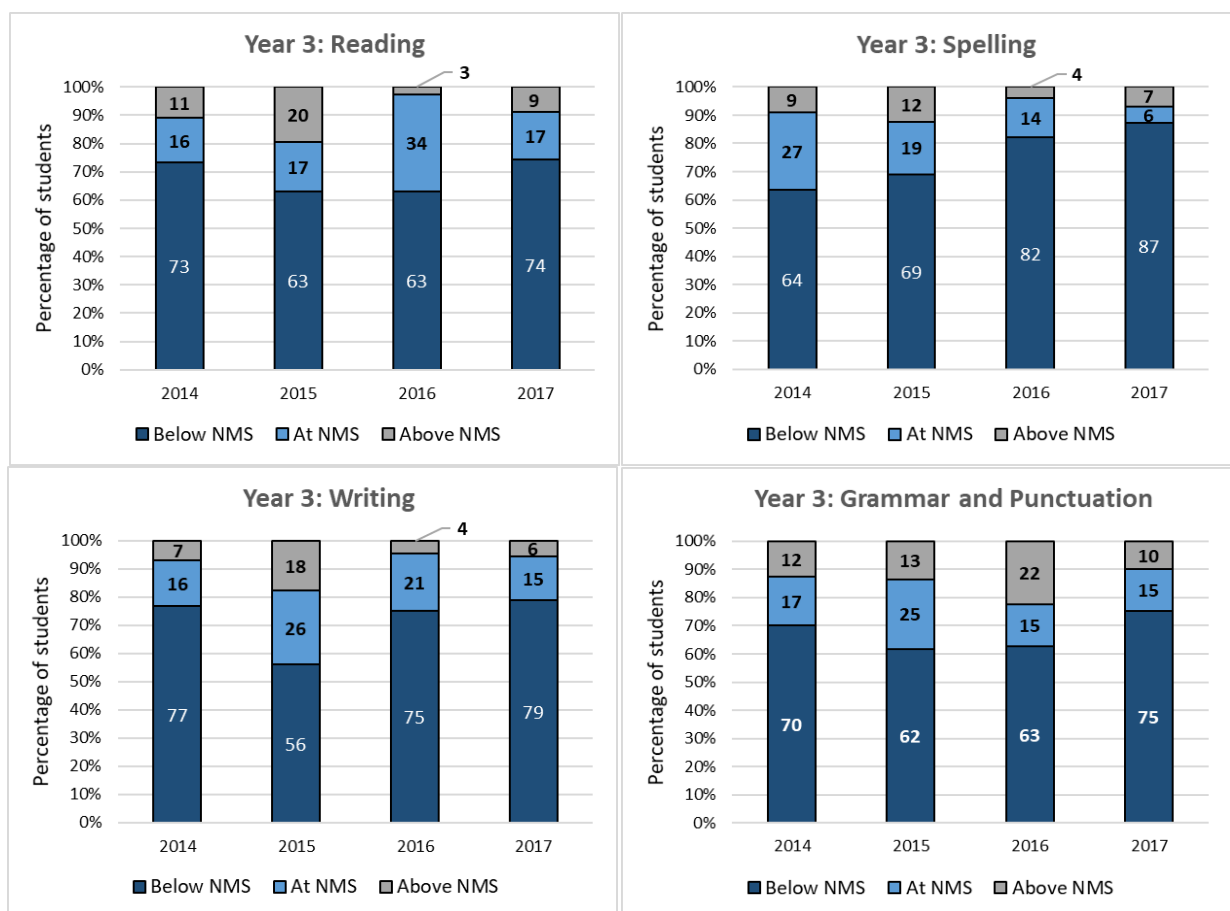
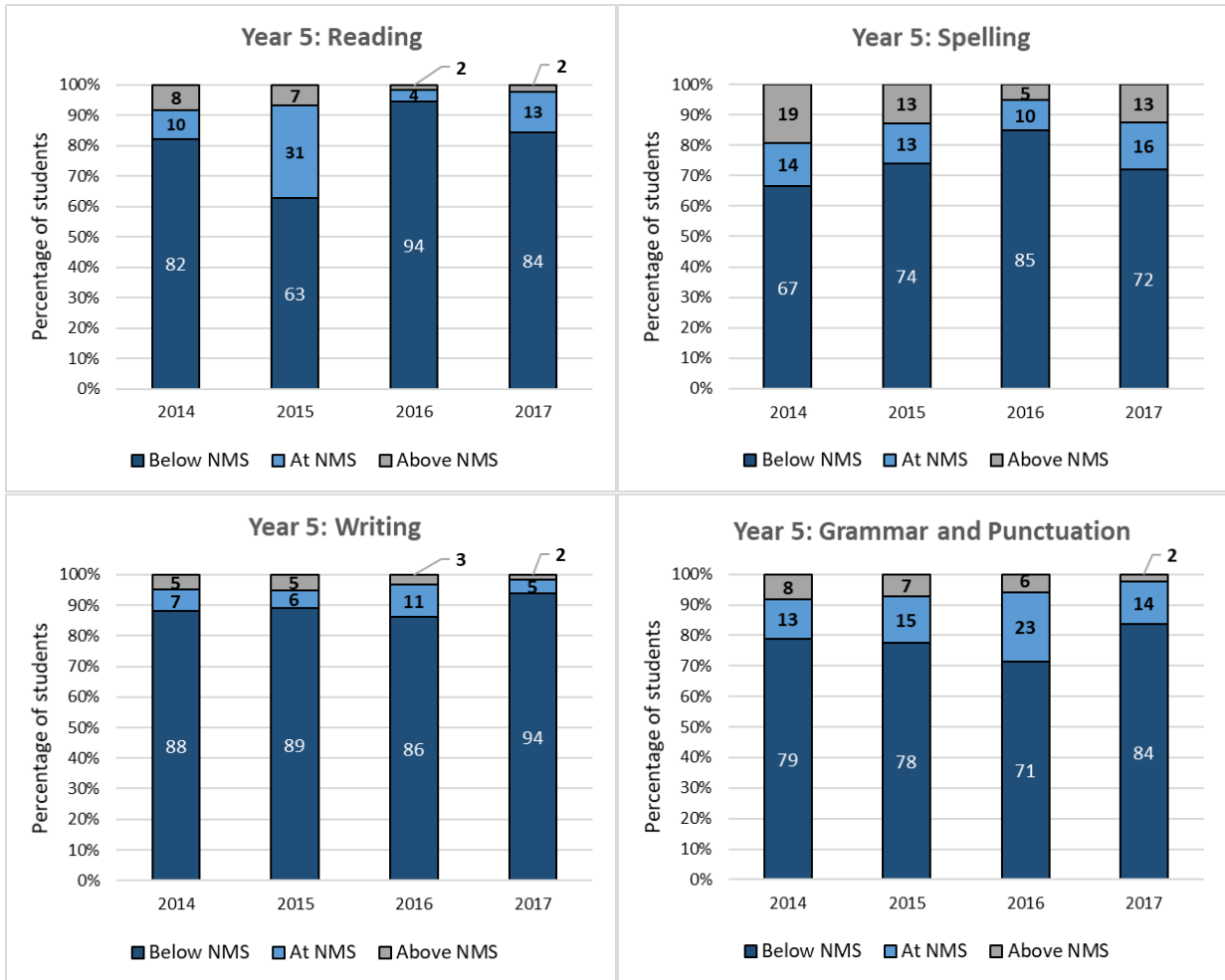*Figure 68: Percentage of Year 3 students in relation to NMS for NT government schools, 2014 to 2017.*

Figure 69: Percentage of Year 5 students in relation to NMS for NT government schools

ff

Appendix 5: NAPLAN Participation Rates, NT Government Program Schools

NAPLAN participation rates were analysed to determine if increases were evident since FLFRPSP commenced in 2015 (see Figure 70). Results indicate a relatviely consistant particapnt rate over the years 2014 to 2017 for NT program schools, with students participanting at the same rate as the state overall.

NT Government School Participation Rates

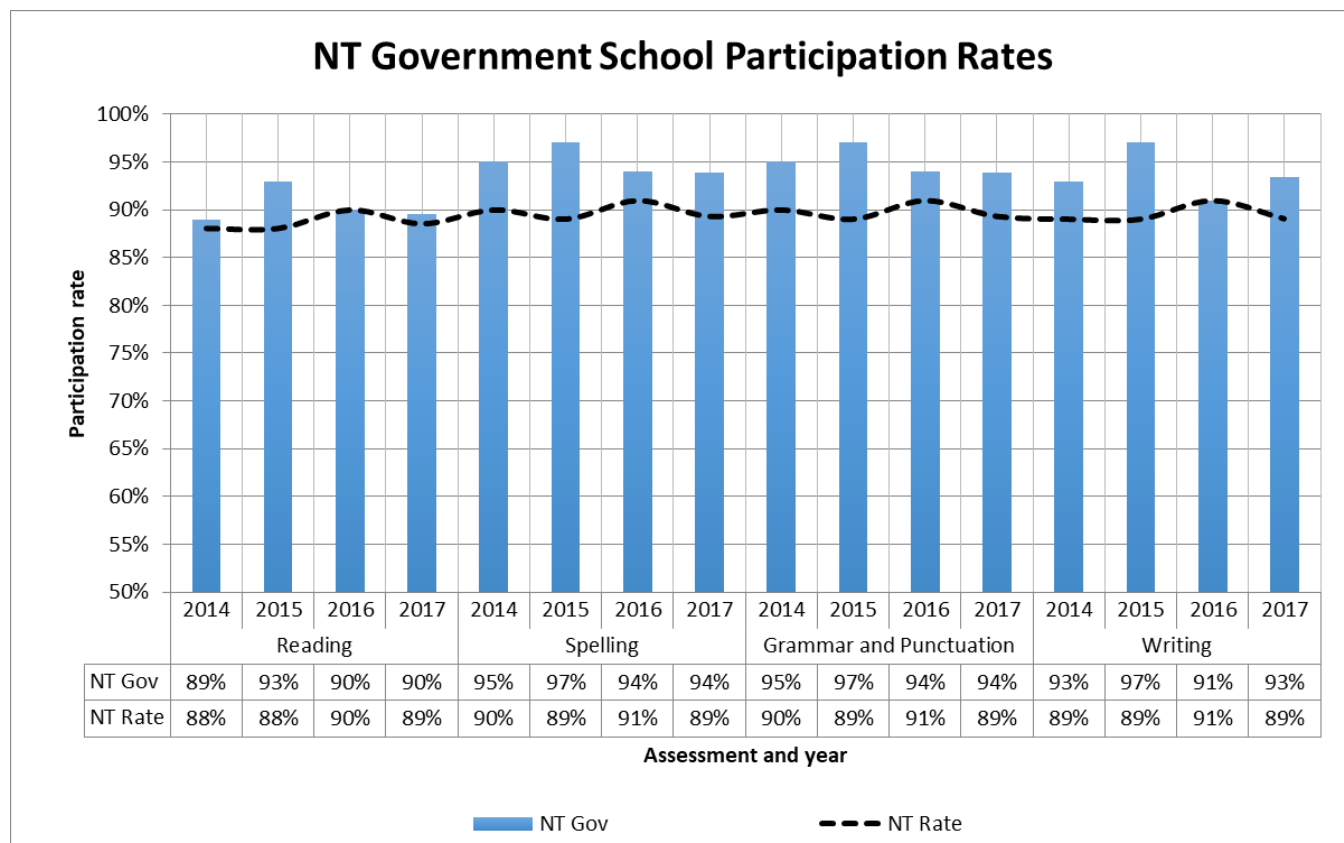| Assessment and year	Reading				Spelling				Grammar and Punctuation				Writing			
	2014	2015	2016	2017	2014	2015	2016	2017	2014	2015	2016	2017	2014	2015	2016	2017
NT Gov	89%	93%	90%	90%	95%	97%	94%	94%	95%	97%	94%	94%	93%	97%	91%	93%
NT Rate	88%	88%	90%	89%	90%	89%	91%	89%	90%	89%	91%	89%	89%	89%	91%	89%

Figure 70: NAPLAN participation rates by assessment type and year for NT government program schools.

Appendix 6: Summary of Fidelity Measures

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
Instructional Leadership	3	Principal engagement and participation with the program
Overall school fidelity	5 scales	Average of the above 4 School readiness measure scores

Appendix 7: NAPLAN Average Scores over Time, WA Government Program Schools

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 to 2017. 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 such, sensitivity testing was also conducted using Kruskal-Wallis H tests.

There was a statistically significant difference over 2014 to 2017 for Year 5 Spelling scores $F(3, 16) = 1.57$, $p = .24$. However, The results indicate no significant difference[19] for the other WA government NAPLAN domains in Year 3 or Year 5 from 2015 to 2017[20].
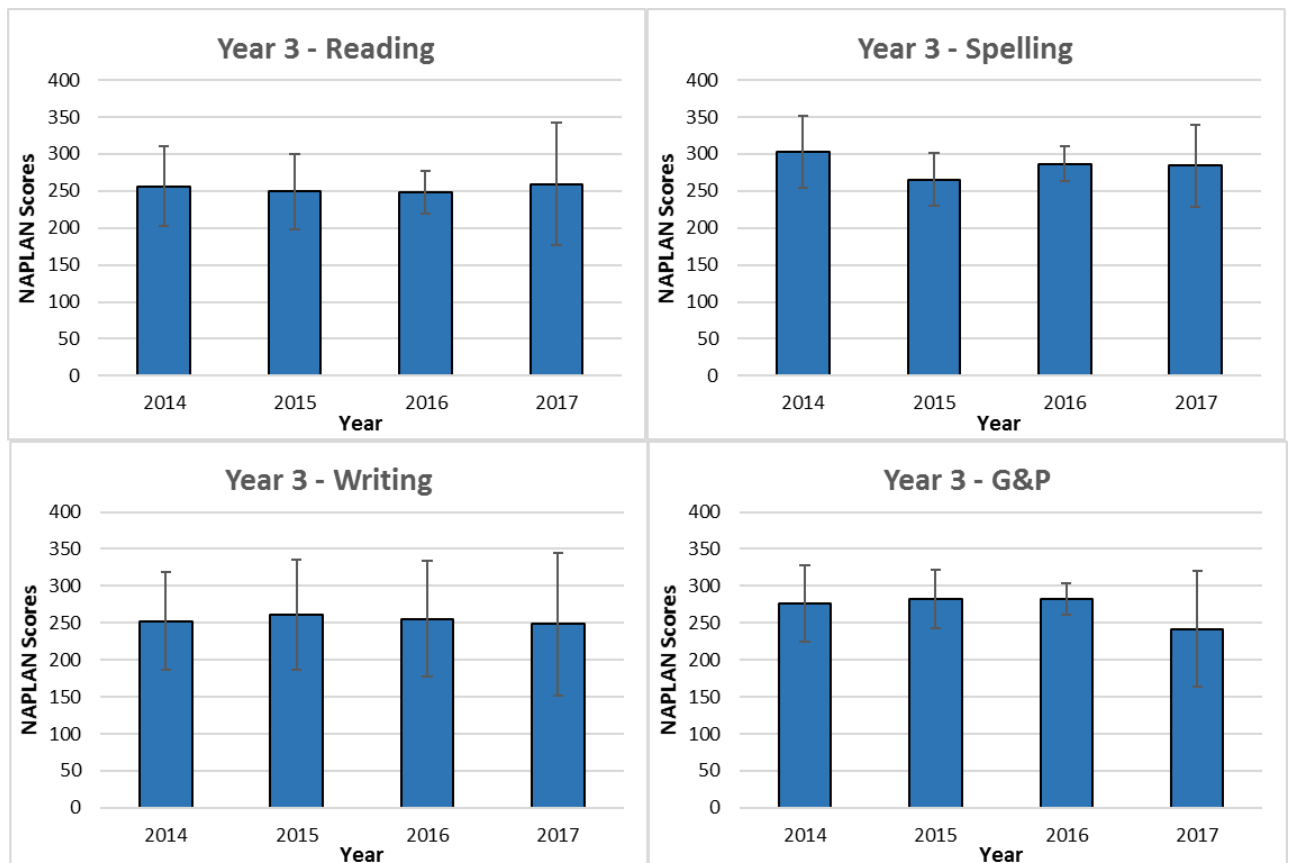

Figure 71: WA government mean NAPLAN scores 2014-2017, Year 3.

---

[19] Confirmed by Kruskal-Wallis H tests.

[20] Year 3: Reading F(3, 16) = 0.04, p = .99; Writing F(3, 16) = .02, p = 1.00; Spelling F(3, 16) = .65, p = .60; Grammar and Punctuation F(3, 16) = .70, p = .57; Year 5: Reading F(3, 16) = 1.25, p = .32; Writing F(3, 16) = .23, p = .87; Grammar and Punctuation F(3, 16) = .97, p = .43.
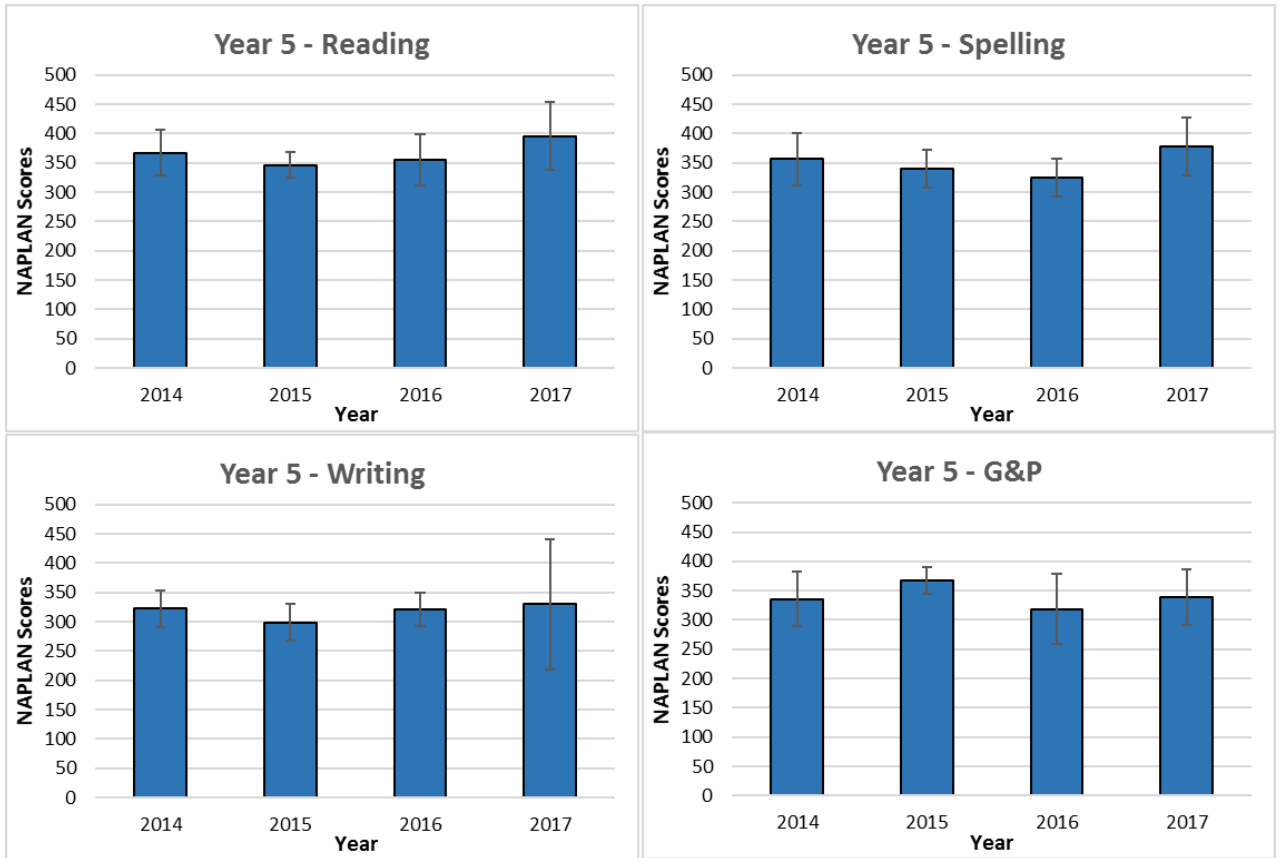
Figure 72: WA government mean NAPLAN scores 2014-2017, Year 5

Appendix 8: Percentage of Students below NMS, WA Government Program Schools

Analysis of the change in the percentage of students below National Minimum Standard (NMS) for years for 2014 to 2017 and for Grades 3 and 5 are presented in Figure 23 and Figure 24.

Overall, there was an increase in the percentage of students below the National Minimum Standard for students in both cohorts. In the Year 3 cohort, this was seen across all four subscales in 2017 compared the two previous years of the program (2015 and 2016), except for Grammar and Punctuation which also dropped considerably from the year before, from 50% to 33%. Although all students were all still all slightly higher than in 2014 (pre-program),

In Year 5 students, the percentage of students above the National Minimum Standard has increased in 2017 compared to results from 2014, 2015 and 2016 in all domains. A particularly dramatic change was seen in writing and reading with a twofold increase in the number of students above the NMS compared to previous years. There is also a higher percentage of students which are currently at the NMS compared to previous years across all four domains in Year 5 compared to 2014 (pre intervention), except for spelling which is roughly equal.
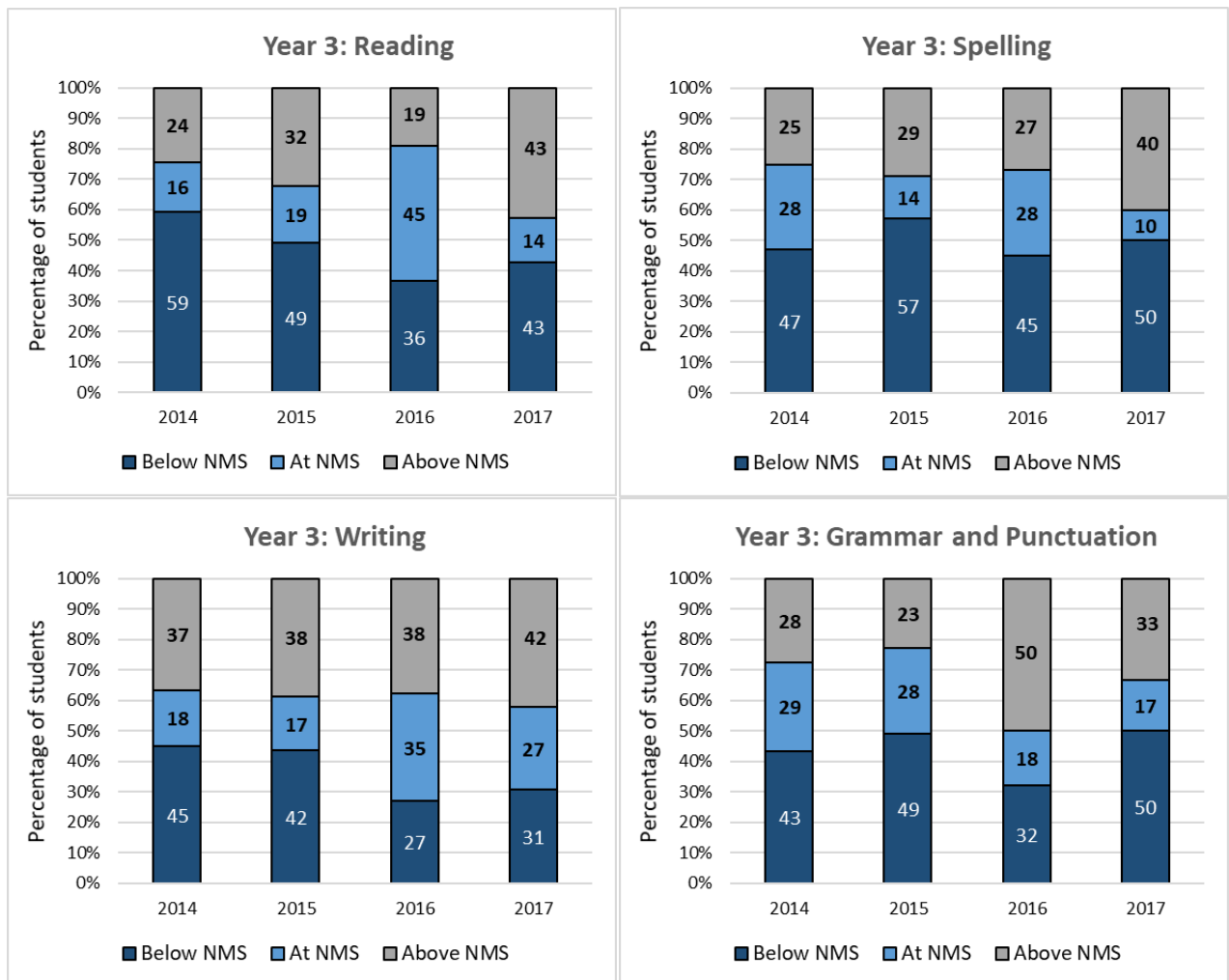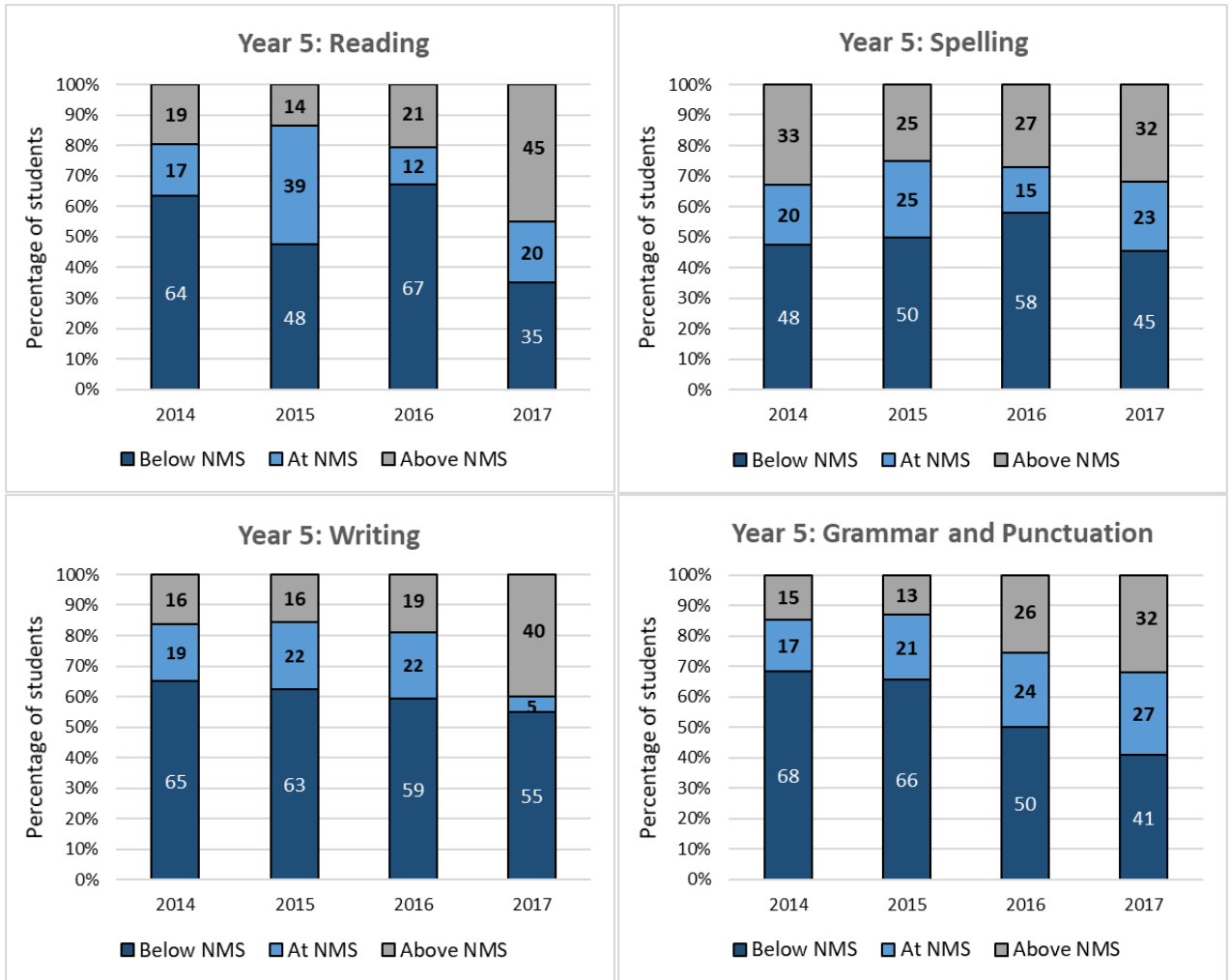

Figure 73: Percentage of Year 3 students in relation to NMS for years 2014-2017, WA government program schools.

Figure 74: Percentage of Year 5 students in relation to NMS for Years 2014-2017, WA government program school

Appendix 9: NAPLAN Participation Rates, WA Government Program Schools

NAPLAN participation rates were analysed to determine if increases were evident since FLFRPSP commenced in 2015 (see Figure 62). Results did not indicate a positive trend over the years 2014 to 2017 for WA government program schools. Participants rates are also below the average participant for the state. Nonetheless, it should be noted that for the here was an increase in 2017 compare to 2016.
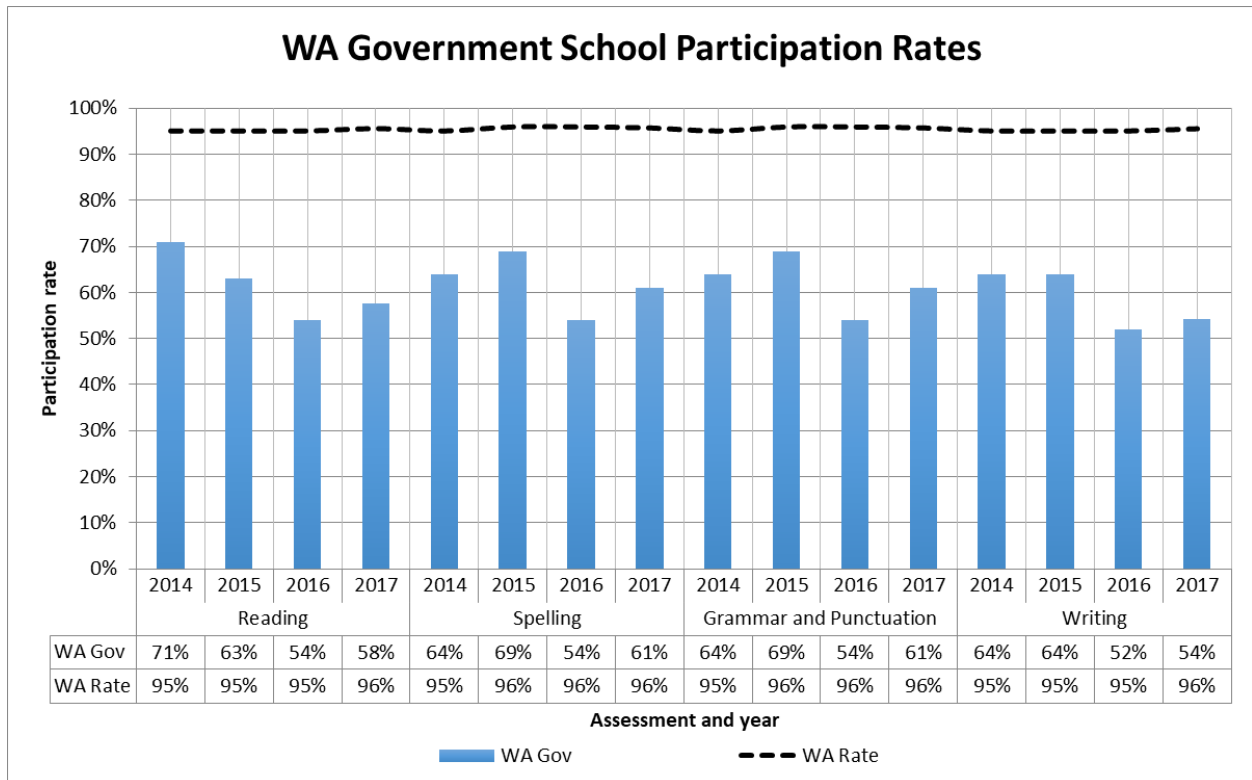

WA Government School Participation Rates

	Reading				Spelling				Grammar and Punctuation				Writing			
	2014	2015	2016	2017	2014	2015	2016	2017	2014	2015	2016	2017	2014	2015	2016	2017
WA Gov	71%	63%	54%	58%	64%	69%	54%	61%	64%	69%	54%	61%	64%	64%	52%	54%
WA Rate	95%	95%	95%	96%	95%	96%	96%	96%	95%	96%	96%	96%	95%	95%	95%	96%

Figure 75. NAPLAN participation rates by assessment type and year for WA government program schools.

Appendix 10: NAPLAN Average Scores over Time, WA Catholic Education Program Schools

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 to 2017. 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 such, sensitivity testing was also conducted using Kruskal-Wallis H tests.

The results indicate no significant difference[21] over 2014 to 2017, for NAPLAN domains in the WA Catholic Education Schools cohort, both years 3 and 5. The trend in the year 3 cohort showed a general trend for growth until 2017 where it seems there has been a drop for all domains. The year 5 cohort also generally decreased from the year before, including Grammar and Punctuation which seems to have a trend of deterioration.
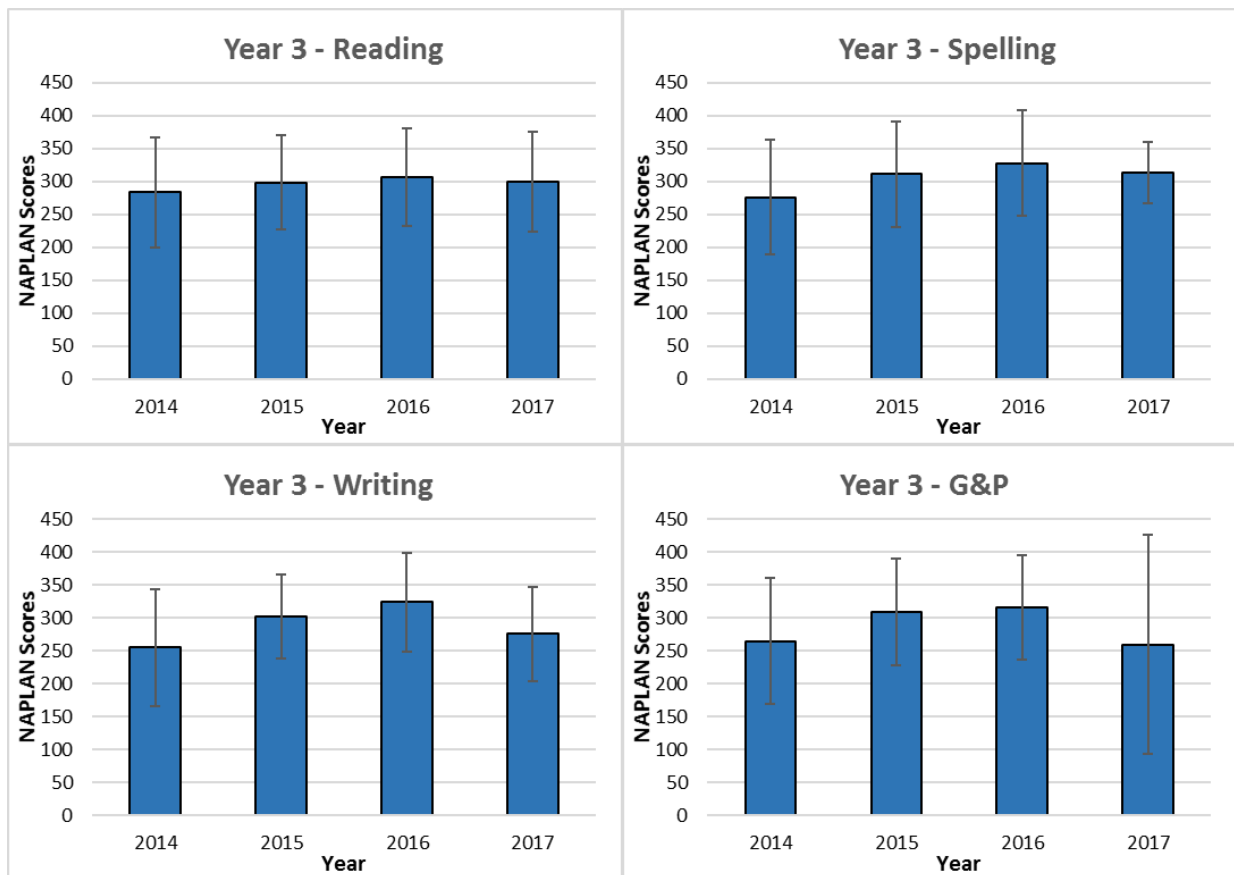

Figure 76: WA Catholic education program schools - mean NAPLAN scores 2014-2017, Year 3.

---

[21] (Year 3) *reading* $F(3, 23) = 0.11$, $p = .96$; writing $F(3, 23) = 1.11$, $p = .37$; spelling $F(3, 23) = .59$, $p = .63$; or grammar and punctuation $F(3, 23) = .50$, $p = .69$ (Year 5) reading $F(3, 24) = .11$, $p = .96$; writing $F(3, 24) = 0.68$, $p = .58$; spelling $F(3, 24) = .86$, $p = .48$; or grammar and punctuation $F(3, 24) = .48$, $p = .70$. Confirmed by Kruskal-Wallis H tests.
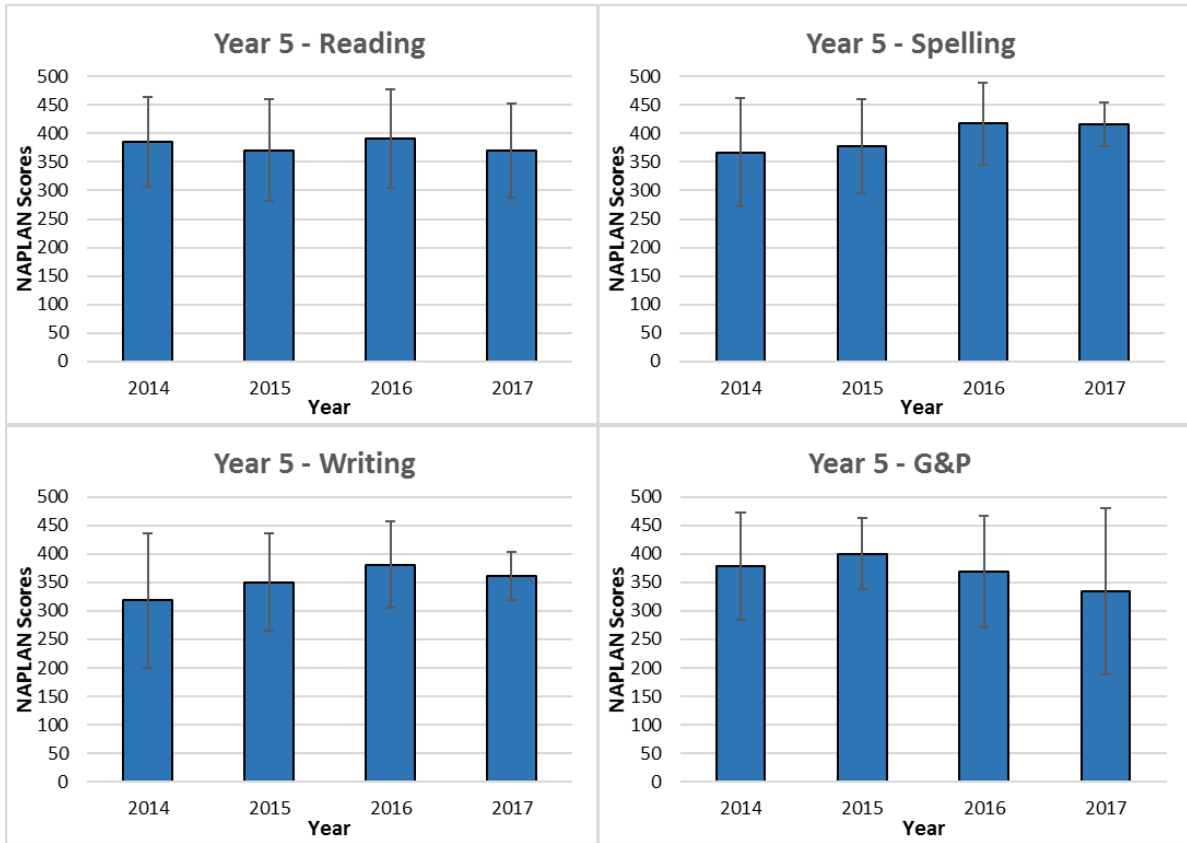
Figure 77: WA Catholic Education program schools - mean NAPLAN scores 2014-2017, Year 5

Appendix 11: Percentage of Students below NMS, WA Catholic Education Program Schools

Analysis of the change in the percentage of students below National Minimum Standard (NMS) for years for 2014 to 2017 and for Grades 3 and 5 are presented in and Figure 78 and Figure 79.

Overall, there was an increase in the percentage of students below the National Minimum Standard for students in both cohorts. In the Year 3 cohort, this was seen across all four subscales in 2017 compared the two previous years of the program (2015 and 2016), although Reading and Spelling is slightly higher than in 2014 (pre-program), except for Grammar and Punctuation which also dropped considerably from the year before, from 81% to 53%.

In Year 5 students, the percentage of students above the National Minimum Standard has decreased in 2017 compared to results from 2014, 2015 and 2016 in all domains. A particularly dramatic change was seen in writing which has dropped dramatically with 32% of students above the NMS compared to previous years of 62%, 54% and 62% respectively. There has, however, been a decrease in the percentage of students below the NMS in Spelling in 2017 compared to other years. There is also a higher percentage of students which are currently at the NMS compared to previous years across all four domains in Year 5.
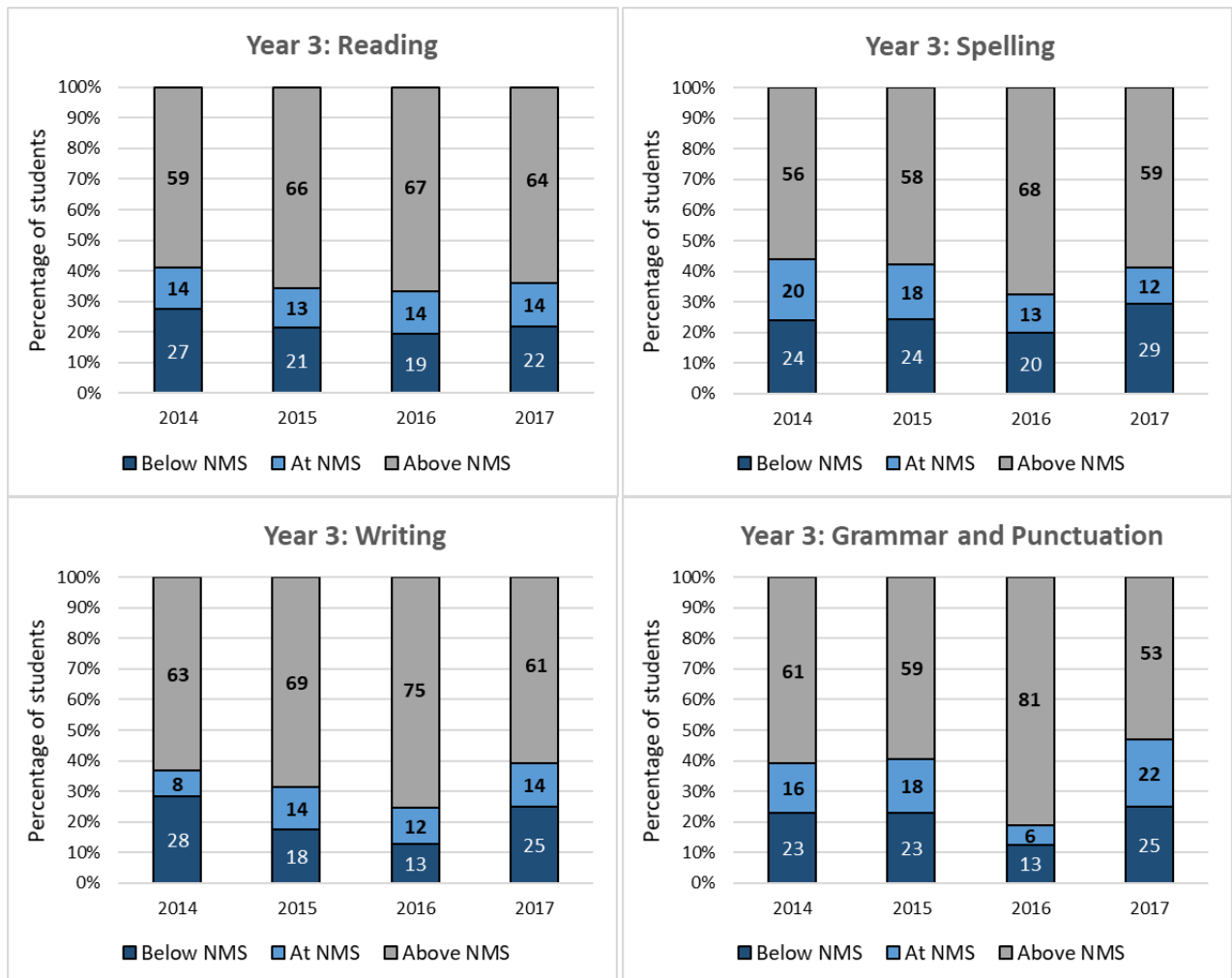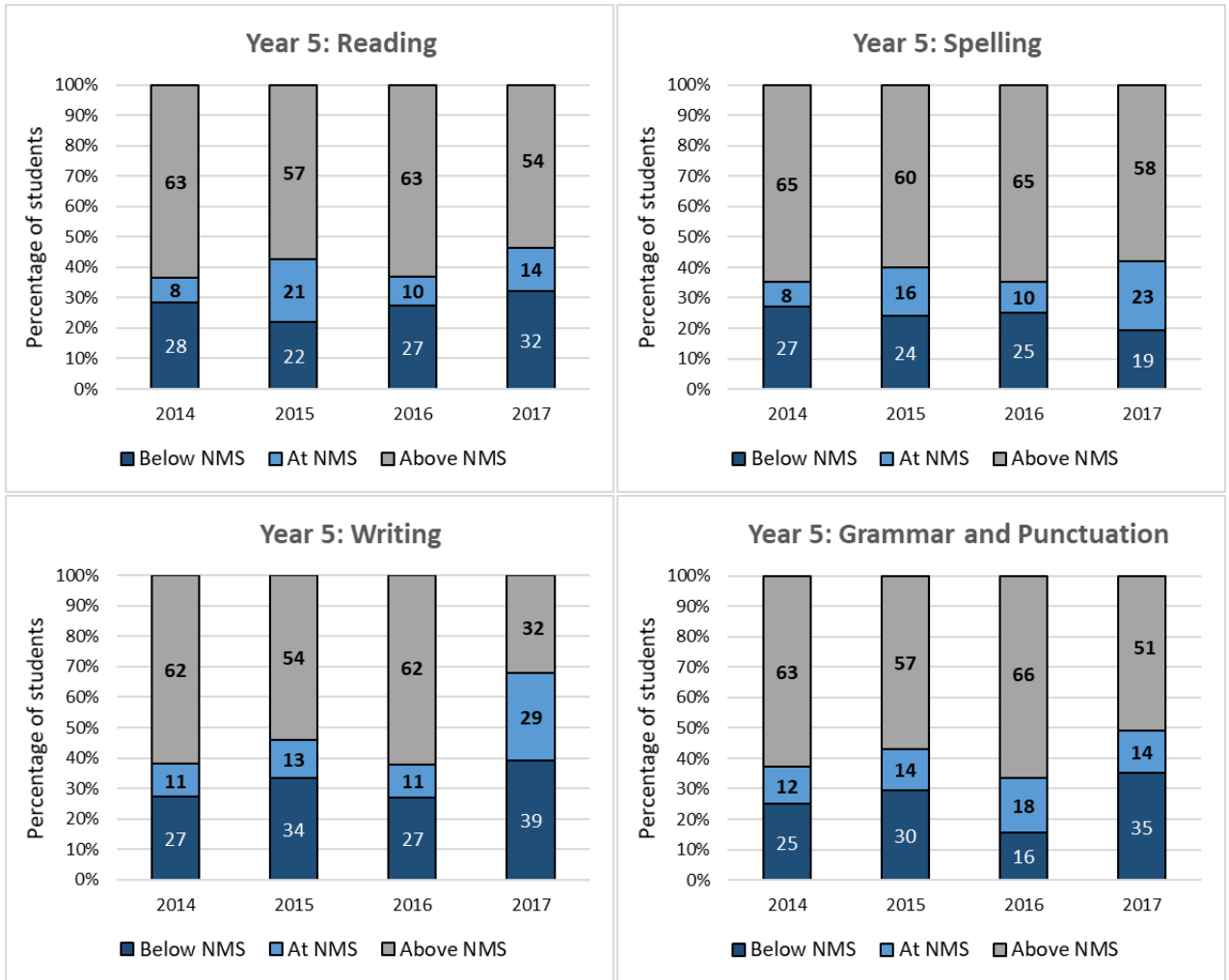

Figure 78: Percentage of students below NMS for years 2014-2017, Year 3, WA Catholic education program schools.

Figure 79: Percentage of Students Below NMS for years 2014-2017, Year 5, WA Catholic Education program Schools

Appendix 12: NAPLAN Participation Rates, WA Catholic Education Program Schools

NAPLAN participation rates were analysed to determine if increases were evident since FLFRPSP commenced in 2015 (see Figure 80). Results indicate a positive trend over the years 2014 to 2017 for WA Catholic Education program schools. While overall parcitpant rates are below the average participant for the state they have been seadily increase from 2014 to 2017.

WA Catholic School Participation Rates

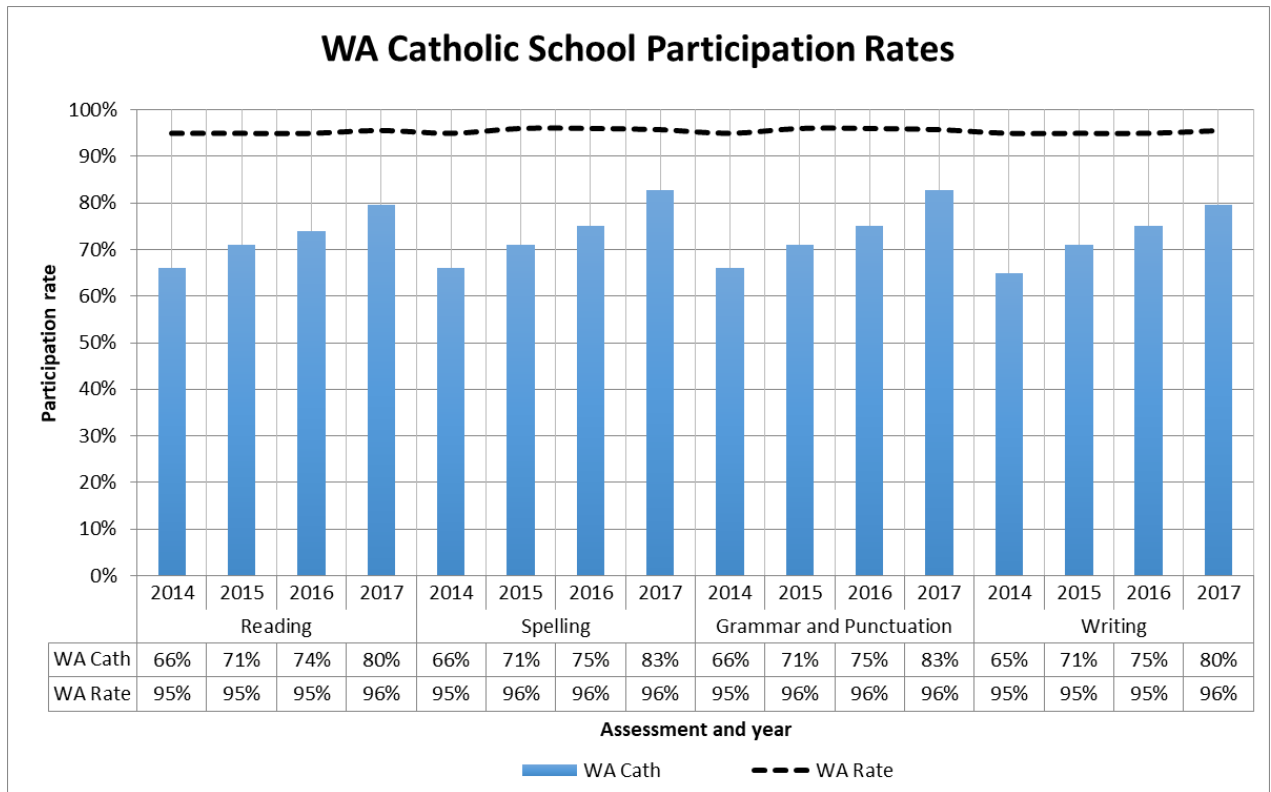|	2014	2015	2016	2017	2014	2015	2016	2017	2014	2015	2016	2017	2014	2015	2016	2017
	Reading				Spelling				Grammar and Punctuation				Writing			
WA Cath	66%	71%	74%	80%	66%	71%	75%	83%	66%	71%	75%	83%	65%	71%	75%	80%
WA Rate	95%	95%	95%	96%	95%	96%	96%	96%	95%	96%	96%	96%	95%	95%	95%	96%

Figure 80. NAPLAN participation rates by assessment type and year for WA Catholic education program schools.