Evaluation of the Flexible Literacy for Remote Primary Schools Program

2015, 2016, 2017 and 2018 School Years

28 June 2019

Prepared for:
The Commonwealth Department of Education and Training

Investigators:
Professor Janet Clinton

Georgia Dawson

Project Team:
Meg Stonnill

Hanchao Hou

Ana Garcia-Melgar

Timoci O'Connor

Esther Kee

Benjamin Howes

Katla Hower

Megan Dennis

Chabel Khan

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

Citation: Clinton, J., Stonnill, M., O'Connor, T., Garcia-Melgar, A., & Dawson, G. (2019). Evaluation of the Flexible Literacy for Remote Schools Program: Summative Report. June 2019. Centre for Program Evaluation, the University of Melbourne.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full text
FLFRPSP	Flexible Literacy for Remote Primary Schools Program
GGSA	Good to Great Schools
DI	Direct Instruction
EDI	Explicit Direct Instruction
NAPLAN	National Assessment Program - Literacy and Numeracy
EYLND	Early Years Literacy and Numeracy Data
LBOTE	Language background other than English
NIFDI	National Institute for Direct Instruction

1. Introduction

This section of the report outlines the Flexible Literacy for Remote Primary Schools Program (FLFRPSP) and the evaluation of the program by the Centre for Program Evaluation (CPE) at the University of Melbourne.

1.1. Flexible Literacy for Remote Primary Schools Program

As part of the Quality Schools package, the Australian Government established the Reform Support Fund, which will provide $186.4 million for school reform across five years, from January 2018 (Australian Government Department of Education and Training [DET], 2017). This fund replaces the Students First Fund, which provided funding to non-government representative bodies from 2014-2017 (DET, 2017). As a key initiative of the Students First Fund, the Flexible Literacy for Remote Primary Schools Program (FLFRPSP) was developed in an effort to address significant and diverse learning barriers faced by students attending remote and very remote schools. These barriers are often distinctly different from those faced by students attending metropolitan or regional schools.

In 2014, Good to Great Schools Australia (GGSA) was contracted by the Commonwealth Government to implement the FLFRPSP using the key principles of the 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, specifically: 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 Carl Bereiter and Siegfried Engelmann (Bereiter & Engelmann, 1966). Teachers trained in DI implement a lesson-by-lesson approach to help students acquire a specific skill. Lessons are designed and delivered in sequence, whilst pacing is carefully managed to support incremental student progress. DI teachers seek to maximise on-task time, providing students with positive reinforcement to support success at each level of attainment. The DI literacy curriculum is based on five key components of reading: phonics, phonemic awareness, vocabulary, fluency and comprehension (NIFDI, 2011).

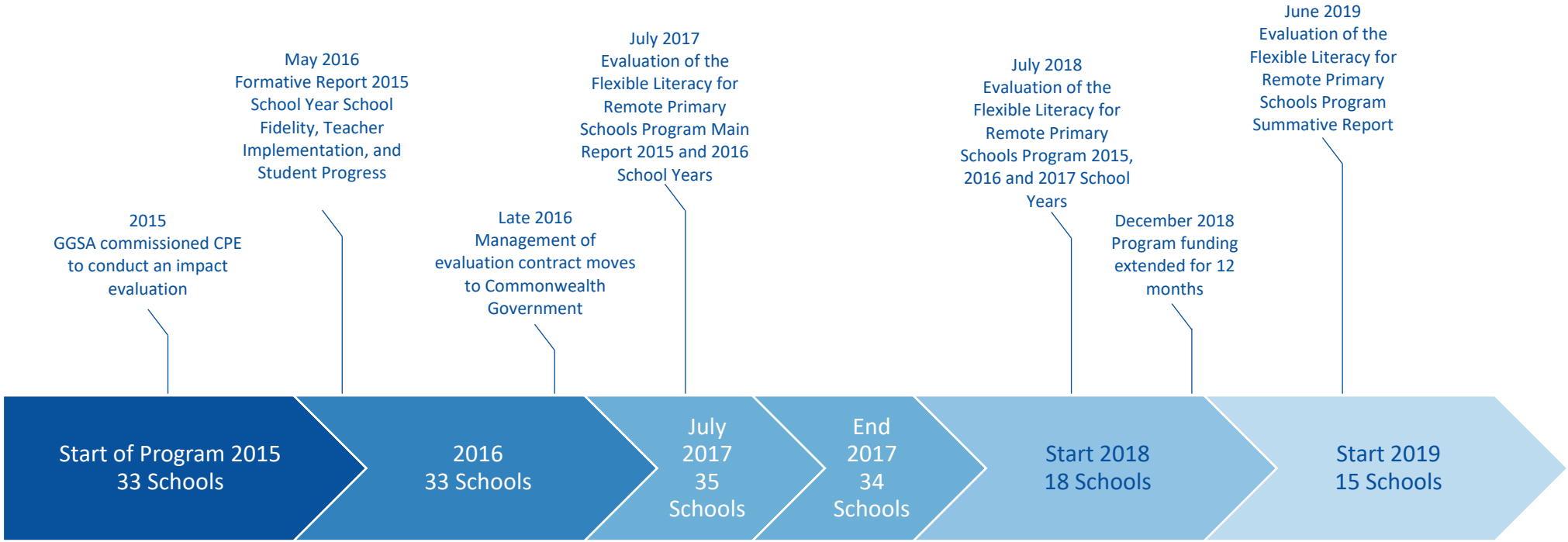
Explicit Direct Instruction (EDI) is a step-by-step guide to creating and delivering lessons across the Australian Curriculum. The model, based on DI theory in addition to contemporary educational theory and neuroscience, was developed by John Hollingsworth and Dr Silvia Ybarra from DataWORKS (Hollingsworth & Ybarra, 2017).

1.2. Good to Great Schools Australia

GGSA is a not-for-profit organisation that supports schools' transition through the following categories: 'Poor to Fair', 'Fair to Good', 'Good to Great' and onwards 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 participants. Currently, GGSA is seeking to address long-term trends of underachievement in Australia by supporting schools in the 'Poor to Fair' category in transitioning toward 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 address low literacy performance: one of several programs undertaken by GGSA to target and support school improvement within the 'Poor to Fair' category.

GGSA is responsible for introducing and implementing FLFRPSP, using either EDI or DI, in remote primary schools in the Northern Territory (NT), Queensland (QLD) and Western Australia (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, following the submission of the Main Report in 2017 and additional submissions by relevant stakeholders, a review of program funding determined that 18 schools would continue with the program in 2018 (see Figure 1). The Commonwealth Government has provided a further $2.8 million to extend the program for another 12 months within 15 schools (Ministers for the Department of Education and Training Media Centre, 2018). While the program has been extended, there is no additional funding for an ongoing evaluation. This report is the final evaluation report for the FLFRPSP and will provide summative information regarding the impact of the program up until the end of 2018.

Figure 1. Timeline of FLFRPSP evaluation 2015-2019.



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 FLFRPSP evaluation, including the development of relationships with the project team, database construction, the attainment of ethics approvals from relevant jurisdictions and the collection of 2015 program and performance data. This stage supported the appraisal of data quality and access, identifying data coverage to highlight where gaps were evident. Additionally, it helped to identify priorities for new data collection, highlighting the need to further develop evaluation capacity within GGSA and participating schools. The technical report produced in 2016 was limited by available data and reported on key implementation factors, ie school fidelity, teaching fidelity and student progress. To this end, the Centre for Program Evaluation created a customised Microsoft Access® database containing internal program data to be provided to GGSA at the end of the evaluation, allowing GGSA to continue monitoring and evaluating the program.

Overall, these efforts ensured that GGSA had the capacity for ongoing evaluation, incorporating a rigorous analysis of available data on the early impact of FLFRPSP against several outcomes, ie teacher practice and student literacy.

In 2016, contract management transferred from GGSA to the Commonwealth Department of Education and Training (DET) and the scope of the contract expanded to a full evaluation design, facilitating greater data collection. The new design enabled a more robust and nuanced analysis of FLFRPSP impact, guiding policy development and providing further credible evaluative information to GGSA and participating schools. This design described comprehensively in the 2017 report (Clinton, Dawson, McLaren, & Koelle, 2017), sought to investigate program effectiveness in improving the literacy results of students at participating schools and improving teachers' pedagogical skills in teaching literacy using the DI and EDI models.

The report submitted to the Commonwealth Government in 2018 (Dawson, Clinton, Koelle, & McLaren, 2018) presented the findings of the program evaluation for the 2015, 2016 and 2017 school years. The evaluation findings provided support that the FLFRPSP has enacted to improve student literacy outcomes. Large effect sizes were observed in NAPLAN Spelling across most schools, as well as small to large effect sizes in NAPLAN Reading and Writing. Notable gains were demonstrated in Catholic Education WA schools based on more specific literacy data. Overall, observed gains demonstrated similar effect levels as seen in previous compatible interventions with similar student populations (Dawson, Clinton, Koelle, & McLaren, 2018). 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 student wellbeing and attendance, suggesting these factors remain complex issues in the context of remote and rural primary education.

The evaluation also sought to understand teacher perceptions of how the FLRPSP has impacted their practice. This was captured through interviews and survey questions, which demonstrated largely positive responses. Teachers in WA Catholic Education schools were generally positive about the program's impact on their teaching practice, indicating that FLFRPSP participation had enhanced their ability to teach literacy as well as their general teaching skills. Similar responses were received from teachers in WA Government schools. Teachers in the NT were slightly less positive regarding whether the program improved their literacy and general teaching skills, yet more than 50 per cent 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.

This current report provides summative information about the impact of the FLFRPSP up to the end of 2018.

1.3.2. Evaluation Framework and Approach

While the fundamental principle behind evaluation is to ascertain 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 utilises an adapted form of the Centers for Disease Control and Prevention for Public Health Evaluation Framework (CDC, 1999), which aims to embed key evaluative principles within the entirety of the evaluation process (see Figure 2). The framework emphasises a continuous cycle of consultation and feedback between all stakeholders throughout the evaluation as measured against four key standards, ie utility, feasibility, propriety and accuracy (Yarbrough, Shulha, Hopson, & Caruthers, 2011). Given the nature of this evaluation and sensitivities that have become apparent in CPE's initial work with relevant jurisdictions, these standards are critically important.

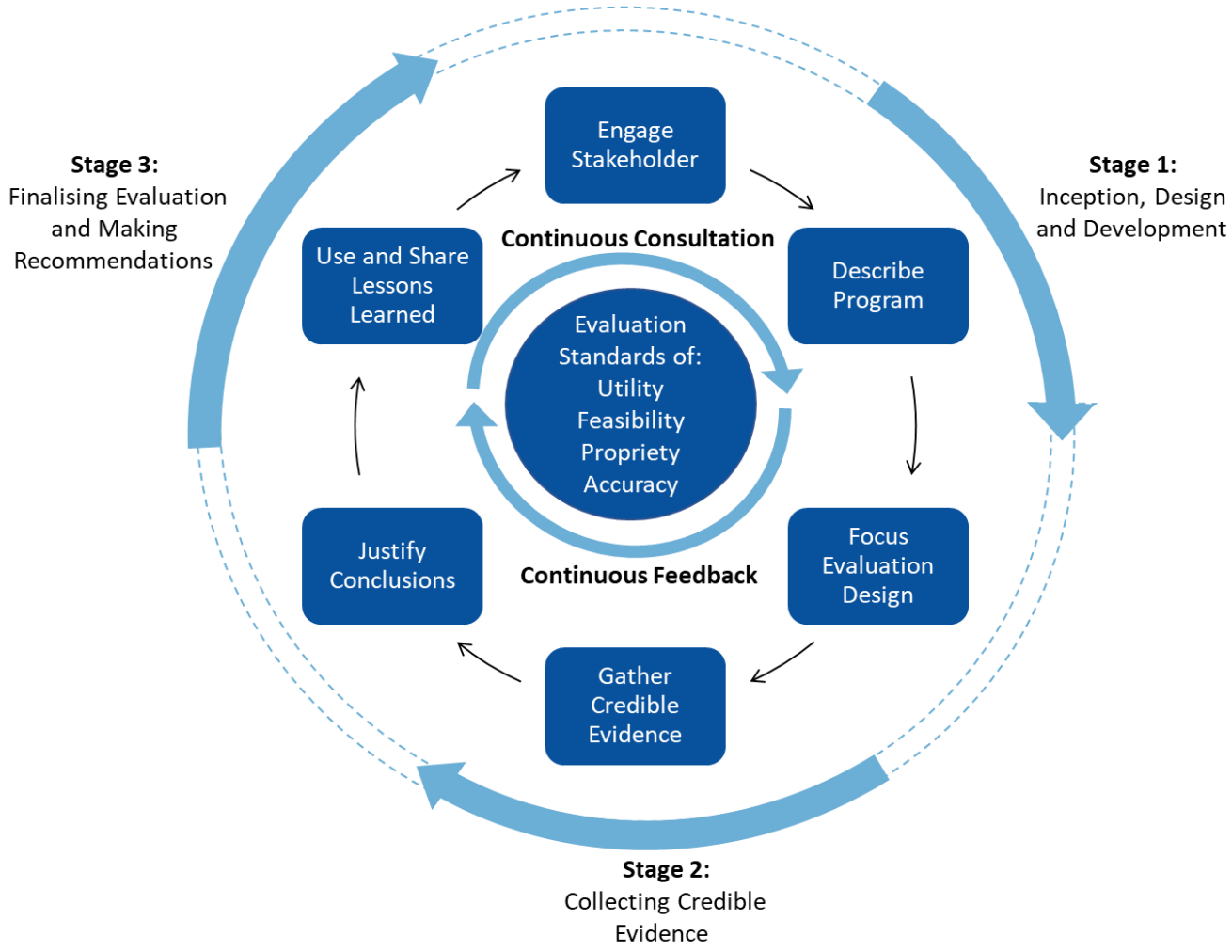


Figure 2. Evaluation framework.

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 FLFRPSP evaluation efforts. The CIPP model has been specifically developed in an educational context and provides a structure that allows for the formative and summative assessment of programs by considering four key components, ie Context, Input, Process, and Product (see Figure 3).

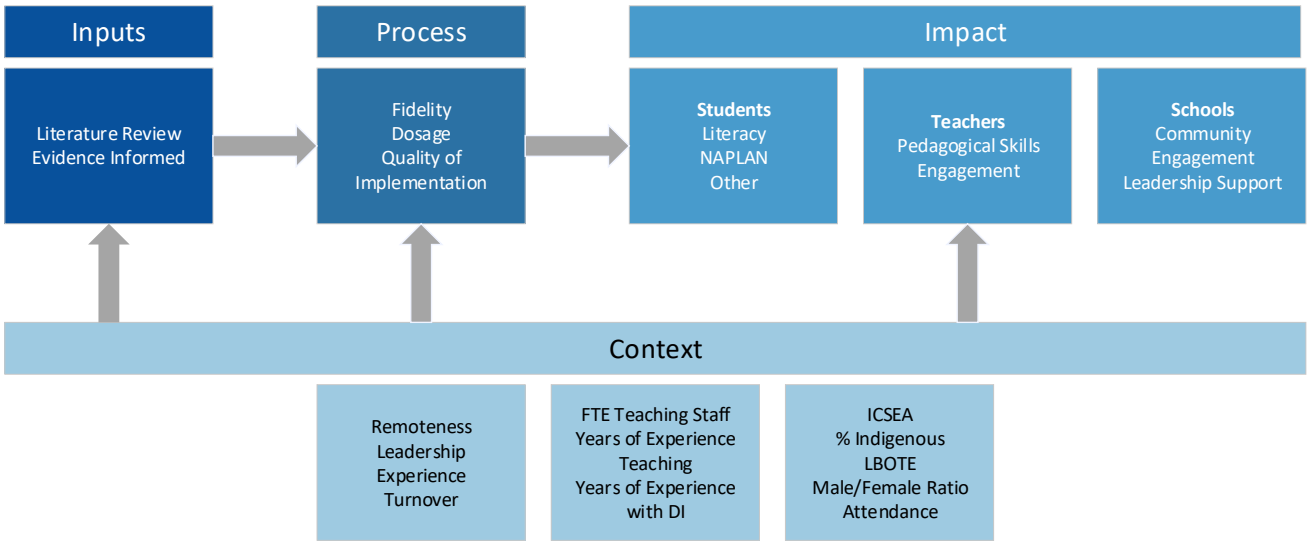


Figure 3. The CIPP model.

Context

In undertaking this evaluation, it is critical that the diverse contexts in which FLFRPSP is implemented are considered and understood. Challenges arise when contexts are heterogenous across student, school and jurisdictional levels. Hence, this evaluation is not only concerned with where and how the program is working, but what factors may help to explain why the program achieves comparatively better outcomes in specific contexts. Given that context encompasses the environment in which FLFRPSP is implemented, a consideration of contextual factors must inform the evaluation of impact, particularly where impact differs between schools or jurisdictions. These contextual factors are captured at the student, teacher and school levels respectively, providing an increasingly nuanced understanding of what variables are associated with differences in program outcomes across different contexts (see Table 1).

Table 1. Contextual Variables at Student, Teacher and School Level

Domain Level	Contextual Variable
Student	ICSEA; SEA; % Indigenous; Male-Female Ratio; Attendance
Teacher	Years' Experience Teaching; Years' Experience with DI; Teacher Turnover
School	School Remoteness; Leadership Turnover; Number Full Time Equivalent Staff

Input

The input component of this evaluation focuses on understanding FLFRPSP in the context of the problems it seeks to address, ie improving student literacy outcomes in the primary years. Accordingly, a synthesis of the literature on similar programs implemented within similar contexts has been provided as an input (see Section 1.4). This will aid in better understanding the impact that can be expected for these types of interventions, by establishing a relative benchmark and comparative degree of progress. The evaluation questions for the input component include:

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

Process

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

Product or Impact

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

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

1.4. Literature Review

The evaluation focuses on understanding the FLFRPSP in the context of the intended goals and objectives, ie improving student literacy outcomes in the primary years of schooling. Accordingly, a rapid synthesis of the literature on the DI teaching approach, and interventions to improve literacy using DI has been provided as an input.

The previous 2017 and 2018 reports presented results of a rapid literature review on the DI pedagogies and explored factors identified as important to the program's underlying foundation. A summary of previous findings is presented below and is expanded upon in this section with more recently published literature and additional understandings about the contextual factors that affect the implementation of DI and school-based interventions in rural and remote contexts and in Aboriginal and Torres Strait Islander communities.

1.4.1. Summary of Previous Literature Review Findings

Direct instruction is a general term for a set of teaching practices that include the explicit explanation of materials and explicit direction of students, grounded in a behaviourist approach to learning that emphasises the student mastery of carefully scripted and highly structured skill acquisition sequences (Engelmann & Carmine, 1982).

Teachers are guided to make behavioural and cognitive goals explicit to students (Luke, 2014), while content is further made clear through demonstrations, participation, guided practice, and discussion (Hattie, 2008).

Explicit Direct Instruction (EDI) is a specific application of direct instruction, developed through the work of Hollingsworth and Ybarra (2009). It caters for a more flexible approach to teaching, allowing teachers to modify their instruction in response to individual student needs.

The FLFRPSP incorporates the implementation of both DI and EDI instructional teaching approaches in an effort to best support the improvement of literacy results for children in remote primary schools. Accordingly, this literature review seeks to examine seminal existing reviews (ie narrative, systematic and meta-analytic reviews) and primary empirical studies on the implementation of DI and EDI across a range of educational contexts since 2000

Direct instruction as a pedagogical approach has been widely examined and reviewed in recent years (Edwards, Weinstein, Goetz & Alexander, 2014; Stockard, Wood, Coughlin & Rasplica Khoury, 2018).

The FLFRPSP's pedagogical approach

The FLFRPSP is based on the following theory of change statement:

> FLFRPSP is intended to support the improvement of literacy results for children in remote primary schools by implementing two instructional teaching approaches, Direct Instruction and Explicit Direct Instruction (Australian Government Department of Education, 2018).

Direct Instruction

Direct instruction (DI) is perhaps most often used to teach literacy skills, where teachers are given explicit lessons to teach, and studies investigating the effectiveness of DI have found it to be an effective method for improving literacy, although DI is sometimes used for other subject areas, such as mathematics. The FLSRPSP uses the National Institute for Direct Instruction (NIFDI) approach to DI. Created by Zig Wnglemann in the 1960s, the NIFDI approach to DI is a sequenced and scripted model of instruction, using small learning increments and clearly defined teaching tasks (NIFDI, 2015).

Explicit Direct Instruction

Explicit Direct Instruction (EDI) was developed from other instructional practices, such as DI (Dataworks, 2019). Like DI, EDI lessons are provided for the teacher; however, unlike DI, teachers have more control of the pace of the lesson by modifying it in real-time in response to students' responsiveness and learning progress (Dataworks, 2019).

EDI was developed by Hughes, Morris, Therrien, & Benson (2017), who reviewed 68 studies which examined explicit instructional teaching practices as a primary focus. Through this review, Hughes et al. (2017) identified the most

commonly used components of explicit instructional practices. Five main components were identified in at least 75 per cent of the publications. These comprise the essential components of explicit instruction and include:

1. segmenting complex skills, or 'chunking', in which complex tasks are broken down into more manageable units of instruction
2. drawing student attention to important features of lesson content through modelling or think-alouds, in which students are provided with clear and consistent descriptions and demonstrations of a skill
3. promotion of successful student engagement by using systematically faded supports/prompts, in which a strategy or skill performed by a student is promoted through gradually decreasing levels of guidance and prompts, with an aim for student independence
4. 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
5. creating purposeful practice opportunities, in which students can practice and therefore retain and generalise new skills and knowledge.

Evidence of the effectiveness of both DI and EDI has formed a key component of the literature review, and these results are discussed below.

Effectiveness of DI on literacy

Several studies have shown that DI can significantly improve literacy outcomes for 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 and were conducted predominantly outside of Australia.

Overall, studies measuring the effect of DI on literacy outcomes show varied effects among the specific student groups, with some studies indicating that more capable students have greater potential to benefit and others indicating that DI can support all students regardless of ability.

Individual studies

- Shippen, Houchins, Steventon, and Sartor (2005) found that 55 students enrolled in two different DI programs all made significant learning gains, but more capable students made greater gains than lower-achieving students.
- In a study of 41,233 students taught DI implemented by NIFDI, Rebar (2007) found that first-graders taught using the National Institute for Direct Instruction (NIFDI) model experienced increased literacy outcomes significantly faster than comparison schools; by the end of the project they reported results above the national mean and means of comparison schools, based on the Comprehensive Tests of Basic Skills (CTBS).
- From a sample size of 104 Grade 4 students, Lencioni (2013) found that lower-performing readers benefitted from DI with independent practice of comprehension strategies, whereas higher-performing readers benefitted from DI with cooperative learning groups to practice comprehension strategies.
- However, evidence is not exclusively positive; Gaston, Martinez, and Martin (2016) found DI to be less effective than other teaching methods aimed at literacy.

Meta-analyses

Meta-analyses examining the efficacy of different teaching pedagogies have found DI to be one of the most effective teaching methods. In a mega-analysis combining four meta-analyses consisting of a total of 304 primary studies, Hattie (2008) determined that DI produced an overall moderate effect size of d = .59. Overall effect sizes varied depending on the subject being taught. For example, the effects of DI were larger when implementing to teach reading (d = .89) compared with mathematics (d = .50). Other meta-analyses of DI have demonstrated effect sizes of varying magnitudes, ranging from small (d = .21; Borman, Hewes, Overman, & Brown, 2003) to moderate (d = .55; Haas, 2005).

Haas (2005) analysed teaching methods for mathematics and found that DI had the largest effect for both low-ability and high-ability students, while Borman et al. (2003) found that direct instruction models had an overall effect size of d = .21. A 2018 meta-analysis of DI effectiveness indicated positive results on all academic subjects, including reading,

language, spelling, and maths (Stockard, Wood, Coughlin, & Khoury, 2018). This meta-analysis synthesised 328 primary studies from 1966 to 2016 and found statistically significant positive effects of DI in all study areas, the largest weighted average effect size were for single-subject designs (.83) and spelling (.66).

Effectiveness of DI on learning difficulties

Many students in remote primary schools are identified as low progress readers due to several contextual factors and educational vulnerabilities. Therefore, it is also important to consider the effectiveness of DI for students with learning concerns. Direct Instruction can be an effective teaching pedagogy for students with learning difficulties. Hattie (2009) demonstrates that the impact of DI is similar for mainstream students compared with those with learning difficulties.

Other studies have demonstrated significant positive effects of DI in populations with learning difficulties, such as students with literacy difficulties (Barth & Elleman, 2017, n=66), geometry difficulties (Zhang, 2017, n=4), and global delay (Özokcu, Akçamete, & Özyürek, 2017, n=3). DI has been shown to be more effective than other teaching methods for increasing literacy in students with learning difficulties, compared with Simultaneous Prompting Procedure (Celik & Vuran, 2014), Activity-Based Intervention (Botts, Losardo, Tillery, & Werts, 2014), and follow-in labelling and incidental exposure to vocabulary (Lund & Douglas, 2016). However, DI was found to be less effective among students with more pronounced learning problems such as those diagnosed with Intellectual Disability (Blik, Harskamp, & Naayer, 2016).

Based on the evidence from these meta-analytic studies, it appears that DI has a positive effective on learning outcomes, with few negative effects found in the synthesised literature. Employing the standard set by Hattie that effect sizes above 0.3 are meaningful positive effects on student learning, all reviewed meta-analyses, except for Borman et al. (2003), demonstrate moderate, and in some cases, large effects of DI on student learning outcomes, particularly in literacy outcomes. These synthesised results provide support for DI as a literacy teaching strategy for primary school students with and without learning problems.

Considerations of the 'side effects' of teaching approaches in effectiveness research

Zhao (2017) conducted a review of the historical and ongoing debates within DI studies and education, specifically identifying the potentially negative 'side effects' of direct instruction. Side effects identified by Zhao (2017) include the rigidity of the approach, inconsistency with developmental theories, inappropriateness for some children and contexts, sustainability of effects, and suppression of student autonomy and creativity. 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). This claim extends beyond DI and further into the paradigms that underpin effectiveness research designs in education.

Zhao asserts that the dichotomy in the research is not from questioning the evidence-supported effectiveness of DI, but instead has arisen from a set of concerns regarding the prescriptive nature of the approach, inappropriateness of DI for certain groups of children and contexts, sustainability of observed effects, and suppression of learner autonomy and development of creativity as an artefact of the pedagogy (Zhao, 2017). Zhao suggests that the negative effects of DI be acknowledged along with the positive effects, noting that, for example, DI can improve student test scores while also suppressing creativity and problem solving.

Zhao's argument implicitly raises the need to consider *multiple* and more *holistic* learner outcomes in effectiveness research, if all intended (and unintended) effects of a classroom intervention are to be investigated.

Influence of fidelity and adherence to DI effectiveness

Davis (2018) raises difficulties associated with measuring the effectiveness of DI, due to challenges capturing and monitoring fidelity and adherence to the approach in a classroom environment. In other words, "unless direct instruction can be clearly and unambiguously specified, robust empirical research into its 'effectiveness' is simply not feasible" (Davis, 2018, p. 135). Specificity in the reporting of research on most DI approaches falls down when it comes to details, such as seating arrangements, exact language used, and the type of lesson taught. As Davis points out, certain kinds of lessons are impossible to teach using a strict DI approach, such as drama and music; however, on this basis it would be inaccurate to conclude that DI is ineffective for teaching such subjects. Effectiveness research on DI falls prey to the challenges of examining effectiveness of an 'in-situ', complex, social intervention which is being delivered by a skilled individual, who has their own individual expertise and in many cases their own style of teaching.

Given this, literature on complex intervention implementation was briefly examined, and it should be noted that this literature was not restricted to the education sector.

1.4.2. Implementation of complex interventions

There is considerable evidence indicating that DI is effective on literacy outcomes for mainstream student populations. However, this evidence in isolation is insufficient to understand whether a DI program like FLFRPSP would be successful in remote Australia. This is partly due to a lack of investigation into the influence of contextual factors on DI effectiveness, and most importantly a comparatively small number of studies conducted on the effectiveness of DI in Australian educational contexts (Lyon, 2018; Moir, 2018). Consequently, literature in the area of implementation science was examined to understand in more detail, the program mechanisms within DI that may be involved in yielding an effect, and the contextual factors that influence these program mechanisms (Moir, 2018).

There is considerable agreement across the literature concerning the key drivers of the implementation of education programs. In most cases, the drivers are common irrespective of the nature of the program, policy or intervention, and include staff characteristics (eg years of experience), staff professional practice development, structured design and planning for implementation and leadership support for the program, policy or intervention being implemented within the school setting (Moir, 2018; Fullan, 2012; Clinton et al., 2015; Honig, 2000; Spillane, 2002). These drivers are described in more detail below. It should be noted that where possible, these drivers have been related specifically to DI, but perhaps more accurately reflect drivers common to all classroom interventions, particularly those that are implemented within contexts with a great degree of complexity. The complexity might be observed in the characteristics of the student population (a greater proportion of students with known vulnerabilities that will influence their learning); the school (leadership style, climate, geographic location and institutional history); and the broader socio-political environment within which the intervention is based. Therefore, models from implementation science, such as the Consolidated Framework for Implementation Research, or the Theoretical Domains Framework developed from considerably strong evidence bases, are useful to articulate *how* the drivers interact with program mechanisms, and specifically highlight the *direct relationship* between implementation (both practices and contexts) of complex interventions and its effects (Cane, O'Connor & Michie, 2012; French et al. 2012). The evidence from implementation science, and more broadly complex classroom interventions, can be applied in a life course model of progression in the context of understanding the effectiveness of DI.

Life course progression and program success

In complex social interventions there is a direct relationship between implementation drivers, implementation itself and, consequently, intervention effectiveness. Applying this to a DI scenario means that effectiveness (or ineffectiveness) is closely related to the specified drivers. Synthesising results in meta-analaytic studies to understand effectiveness of DI only provides a view of one piece of the puzzle. Therefore, evaluations of effectiveness should account for or at least monitor these drivers and *understand implementation and effectiveness* in relation these drivers.

Putting this into practice requires a conceptual model that reflects these considerations and aligns with approaches to measuring implementation and effects. The life course model of progression has been applied as a conceptual model which has been developed and tested in several Australian education evaluations including, but not limited to, the evaluation of the implementation of the Australian Professional Standards for Teachers; the evaluation of the Western Australian Independent Public Schools Initiative; and the evaluation of the Notschool Australia Pilot Program (Clinton et al., 2015; Clinton et al. 2013; Clinton et al. 2014).

The life course model below outlines the behaviour change process involved in implementing a classroom intervention, as well as highlighting the drivers of this process. Impact is modelled as a progressive curve, where impact may be observed only when implementation activity and behaviour change has occurred. The figure also indicates that a negative experience in implementation activity or behaviour change can impact the progressive trajectory, as it can alter the attitude of implementers both towards the program and the implementation activity occurring.

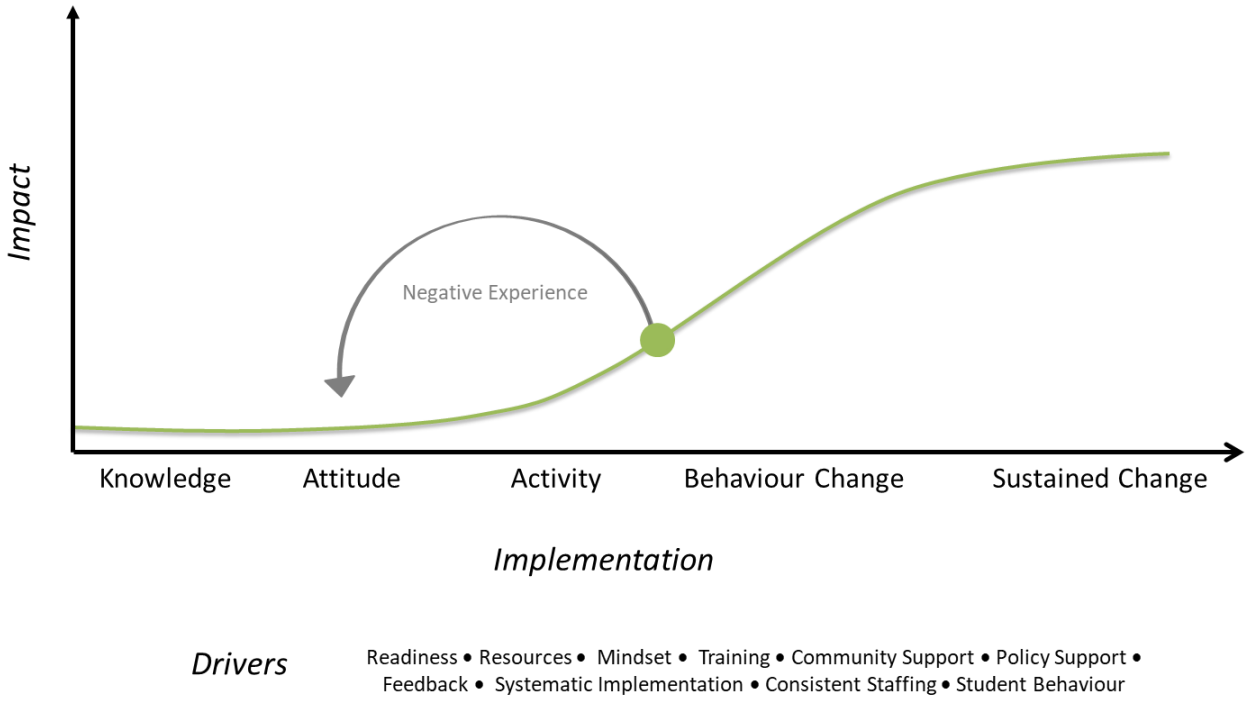


Figure 3. Life course model (adapted from Clinton et al., 2015).

Well-prepared, committed staff is a key factor in high quality, sustainable interventions (Maher, et al., 2008; Denton, et al., 2003; Moir, 2018). A study on the impact of practitioner and site characteristics on fidelity found that teacher belief of success, perceived support from external organisations (ie the program organisation) and personality characteristics such as openness were positively correlated with fidelity (Klimes-Dougan, et al., 2009).

Relatedly, another key factor in successful education programs is staff development (Joyce & Showers, 2002). Staff development in successful programs is characterised by programs which create ongoing systems in which goals for student learning are set, training is provided, staff collaboration is embedded in the program, and student data is collected and analysed (Joyce & Showers, 2002).

The continual monitoring of implementation and student outcomes can also be considered a driver of implementation separate to staff development (Moir, 2018). Programs that embed data systems are more likely to facilitate monitoring of implementation and outcomes.

Finally, leadership is considered important for facilitating implementation by providing a supportive work environment, supporting positive staff attitudes towards the program and providing a commitment to change (Aarons, et al., 2014).

Each of the identified drivers of implementation influence the implementation process; that is, the shift from knowledge to attitude, fidelity to the intended design of the intervention, and the degree and quality of adaptations to this implementation process in response to the contextual environment within which the implementation is occurring. Fidelity and adaptation are not mutually exclusive; both can and should occur in complex interventions. Complete fidelity is in fact considered suboptimal, as it has been shown in complex interventions to compromise effectiveness (Moir, 2018).

1.4.3. Rural and Remote School Considerations

Schools in remote and very remote areas face their own contextual considerations not encountered in more metropolitan schools. In addition to the practical challenges of being more rurally located, schools in these areas also tend to be based in communities and student populations with relatively lower socio-economic advantage, which contributes to a larger number of students who may come to school with less opportunity to learn. For instance, students may not eat breakfast at home, and they are more likely to have learning difficulties and other difficulties associated with exposure to teratogens during the prenatal period and early childhood (such as foetal alcohol syndrome) (Lock, Budgen, Lunay & Oakley, 2012; Clarke & Wildy, 2011; Godinho, Woolley, Scholes & Sutton, 2017).

There are also high rates of absenteeism in remote and very remote schools (some students missed more than 80 per cent of the school year in one WA district; Clarke & Wildy, 2011); low academic achievement, especially in literacy (Clarke & Wildy, 2011; Godinho, Woolley, Scholes, & Sutton, 2017); and in some cases "little parental engagement with the school because of a profound mistrust of the education being offered" (Clarke & Wildy, 2011, p. 28).

In the Independent Review into Regional, Rural and Remote Education, Halsey (2018) identified nine factors that have a major impact on student achievement in rural and remote schools. These nine factors include: curriculum and assessment, teachers and teaching, leaders and leadership, school and community, information and communication technology, entrepreneurship and schools, improving access, diversity, and transitioning beyond school.

The Independent Review into Regional, Rural and Remote Education was informed by 340 submissions lodged by the public, government, and education authorities, among others, as well as a literature review and consultations (Halsey, 2018). During the review, participants identified teachers as one of the most important factors in student learning (Halsey, 2018).

High teacher turnover

The Independent Review into Regional, Rural and Remote Education also found that there is a high level of teacher turnover in remote schools. Beginning teachers have traditionally been a major staffing source for remote schools. While some beginning teachers are retained in these educational contexts, there is a high rate of teacher turnover (Halsey, 2018). This high teacher turnover makes it more difficult to establish sustained school improvement programs, as well as maintain a high level of quality teaching (Godinho, Woolley, Scholes, & Sutton, 2017).

Teachers in remote and very remote areas tend to experience burnout and report specific challenges working in these areas including: geographical, social, and cultural isolation; inadequate accommodation; student management; salary; or a lack of opportunity for professional development (Lock, Budgen, Lunay, & Oakley, 2012). Sharplin (2002) found that many pre-service teachers had stereotypical and underinformed assumptions about living and teaching in remote areas, which made their transition into those areas much harder and made them more likely to leave.

One of the recommendations from the Independent Review into Regional, Rural and Remote Education was to ensure challenges, and opportunities, of working in rural and remote schools and contexts are explicitly included in the selection processes for teacher education and initial appointment processes (Halsey, 2018). Attracting and retaining teachers in rural and remote schools remains a persistent challenge in education (Halsey, 2018). This was evident in the implementation of the FLFRPSP.

Education and assessment among Aboriginal and Torres Strait Islander student populations

In addition to the above considerations, there are other challenges specific to Aboriginal and Torres Strait Islander students who attend remote and very remote schools. Given that Aboriginal and Torres Strait Islander students form a large component of the target population for the FLFRPSP, considerations specific to this population have been identified.

There is a persistent inequality in achievement between Aboriginal and Torres Strait Islander and non-Aboriginal and Torres Strait Islander students in Australia (De Bortoli & Thomson, 2010; Ewing, 2011; Luke et al., 2011). The determinants of this inequality are multifactorial and have not been examined here in their entirety for the purposes of brevity, but it is important to note that access to quality education differs considerably among the two student populations (Bodkin-Andrews, 2014). Bodkin-Andrews (2014) argues that the "insidious effects of epistemological racism still plagues the Indigenous Australian educational research agenda" (p. 786). McRae et al. (2000) argues that a shift must occur in which Indigenous students are respected as individuals, their cultures are "respected and included in the education system", they have "the benefit of quality teaching practices", and they are able to attend school on a regular basis (p. 180).

While several education reforms have been implemented to support addressing the inequality, much of the reforms have focussed on improving student literacy outcomes, given this has the greatest potential for improving long-term educational outcomes. Consequently, DI-oriented programs have emerged and been trialled because of the evidence support their effects on literacy outcomes. Some examples include the Schoolwise Program in NSW (Wheldall, Beaman, & Langstaff, 2002), which compared academic improvement between Indigenous and non-Indigenous students, and the Cape York Aboriginal Australian Academy (CYAAA) project, which has demonstrated mixed results overall (Australian Council for Educational Research, 2013).

However, there is important debate regarding the efficacy and implementation of literacy interventions with Aboriginal and Torres Strait Islander students. Several researchers have advocated that Aboriginal and Torres Strait Islander students require different education strategies on the basis that existing models of education are designed, tested and developed for a non-Indigenous cultural context (Dow, 2011; Luke et al., 2011; Stewart, 2002). Going further, Luke et al. (2011) argue that instead of focusing on comprehension and academic performance, teaching models for Indigenous students should focus on "bridging community cultural practices and epistemologies with systematic introductions to the specialised genres and registers of school and institutional texts" (p. 154). This aligns with findings by De Bortoli and Thomson (2010) that certain cognition, beliefs, and actions, for example a student's interest in learning and belief in their own ability, are associated with learning outcomes. In consultations with community and education leaders, teachers, parents and students, the review of Regional, Rural and Remote education highlighted the disconnect between the Australian curriculum and the needs of Aboriginal and Torres Strait Islander students, with Elders seeking greater opportunity to contribute to the curriculum to support students learning about Homelands knowledge and life skills (Halsey, 2018).

In addition to the need for a relevant curriculum that engages students' interest in learning and a belief in their own ability, language concerns are also made apparent; students whose first language is often not Australian English must in many literacy interventions engage with materials that are not designed in their language, adding another barrier to their learning and to implementation of any programs (Dow, 2011; Goldfeld, Quach, O'Connor & Aston, 2015). As presented in section 4.2.2 of this report, in the context of the FLFRPSP on average 85 per cent of students in program schools have a language background other than English (LBOTE), and in some school this is 100 percent. In addition to the intervention content, DI lessons are highly scripted and fast-paced, which does not allow for specific corrections or wider explanations when misunderstandings (often culturally-based) occur (Dow, 2011).

Assessment of achievement in rural and remote schools

Finally, it would be remiss not to highlight considerations associated with assessing achievement, specifically literacy outcomes for students in rural and remote schools (including Aboriginal and Torres Strait Islander students).

The National Assessment Program - Literacy and Numeracy (NAPLAN) provides national literacy and numeracy data for Year 3, 5, 7 and 9 students (ACARA, 2016). The achievements of regional, rural and remote students have been consistently below that of urban students for decades; further to this, the 2017 trends showed that the gap increases with increasing remoteness (Halsey, 2018).

One of the nine factors which were identified as having a major impact on student achievement in the Independent Review into Regional, Rural and Remote Education was curriculum and assessment. An associated recommendation that emerged from the review in regard to this was a need to create more flexibility in the application of the curriculum to ensure relevance for regional, rural and remote students and increase their engagement with it.

Given that assessment and curriculum are closely related, as they should be, submissions to the review highlighted concerns with the relevance of NAPLAN, specifically that it could be considered 'metro-centric' and does not provide enough connection to the lives of regional, rural and remote students for them to adequately respond to assessment questions (Halsey, 2018).

The limitations of using NAPLAN data in the evaluation of the FLFRPSP to measure the literacy of rural and remote students are acknowledged. The purpose of the analysis of data was to examine growth patterns in outcomes rather than compare and make absolute judgements about academic performance of regional, rural and remote students.

Limitations to the evidence base on DI

Drawing this rapid and focussed review together, several limitations of the evidence base on DI should be acknowledged. While there is distinct evidence of the effectiveness of DI, study designs illustrating effects tend to have small sample sizes and are weighted towards mainstream students located in metropolitan areas. Further, the rigour in these study designs is largely questionable, with limited high-quality randomised controlled trials.

As highlighted earlier, there is also little evidence about the factors that influence the effectiveness of DI; for instance, the evidence highlights the importance of teachers, school leadership support, and a supportive socio-political environment, but it does not articulate in detail the extent to which these drivers contribute to implementation fidelity, adaptation, degree, and over-time intervention effectiveness.

Finally, there is almost no research on DI in regional, rural and remote Australian educational contexts, and where international evidence exists for DI in regional or remote contexts, the generalisability of these findings is questionable.

Summary

Therefore, drawing on these findings, it can be concluded that DI should improve literacy, and potentially numeracy, outcomes for a mainstream student population in metropolitan areas. Although there is strong evidence underpinning DI, it is not possible to conclude based on the evidence alone whether DI would be effective in regional, rural and remote Australian contexts with both Aboriginal and Torres Strait Islander students and non-Aboriginal and Torres Strait Islander students.

2. This Report

2.1. Purpose of this Report

The purpose of this report is to present the summative findings of the FLFRPSP evaluation for the 2015, 2016, 2017, and 2018 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. It will provide a culminated view of the program from the beginning of its implementation through to the end of 2018.

2.2. Structure of this Report

This report began with an executive summary, introduction and literature review 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 presents aggregated results from each state or jurisdiction under the components of the CIPP model outlined earlier:

1. Context (see Section 4.2)
2. Outcomes (see Section 4.3 and Section 4.4)
3. Process (see Section 4.5)

In the subsequent discussion, all findings are critically examined and contextualised within relevant research and theory. The next sections describe data sources and analytical approaches underpinning quantitative and qualitative data analyses.

3. Methodology

This section outlines the methodological approach underpinning data collection and analysis. Although methodological considerations described in this section specifically refer to data collection and analyses procedures undertaken in 2019, this approach leverages and builds on evaluation findings to date. The complexity of the educational reform under evaluation calls for methodological approaches that collect, analyse and combine multiple data sources. Consequently, this evaluation employs a mixed methods research design to address the following 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?

In utilising a mixed methods approach, this evaluation seeks to capture data from a range of sources; these are described in the next section.

3.1. Description of the Data Sources

The table below provides an overview of data sources in the evaluation, participants and how the data was used.

Table 2. 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 and student outcomes.	Students, teachers, school leadership.

Data source	Description	Type	Use	Participants
NAPLAN	Standardised annual assessment for students in Years 3 and 5. Tests utilised in this evaluation were Reading, Writing, Spelling, and Grammar and Punctuation domains for program and control schools for the assessment years 2015, 2016, 2017 and 2018.	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 a paper-based survey relating to their perspectives and experiences with the program. 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 program's 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 implementation and its impact on student, teachers, schools, and the community more broadly.	Qualitative	Stakeholder perceptions of the program's 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.	Principals

3.2. Quantitative Methods

This section describes data collection and analysis procedures of FLFRPSP program data, teacher survey data, and NAPLAN data.

To analyse gains in literacy, NAPLAN scores of program and control (ie schools with similar socio-demographic background not participating in the program) were compared. The initial NAPLAN sample consisted of 487 students from 15 schools participating in the program in 2018 and their respective control schools. NAPLAN data was obtained through a formal application and approval process with the Department of Education of each state, in alignment with University of Melbourne Human Research Ethics Committee and Department of Education ethics guidelines and code of conduct. The analysis of NAPLAN data included 2015 and 2016 Year 3 NAPLAN scores, and 2017 and 2018 scores. Student individual scores were used to calculate mean scores by NAPLAN domain for each participating and control school by year. Ten students were excluded from the analysis as they had transferred between program and control schools. The analysis of missing data (individual student NAPLAN scores) indicated no anomalies in the data set and thus, methods of multiple imputation to account for missing data was not conducted. Mean NAPLAN scores were then calculated for program and control schools by year, alongside state average, national remote, and national very remote NAPLAN scores. Mean scores and standard deviations were weighted by the number of school participants.

To calculate differences in NAPLAN mean scores and compare literacy gains across schools and years, a growth variable (2018-2016 and 2017-2015) was calculated. Pooled mean differences were calculated for program and control schools to account for the variability in sample size of program and control schools, and then compared to national and very remote mean difference scores. Pooled effect sizes of the differences in mean scores in the 2016-2018 and 2015-2017 period were calculated using Hedge's g coefficient, then averaged by type of school: program and control. Finally, a series of independent t-tests was conducted to determine statistically significant differences in effect sizes between program and control schools across the two periods. Upon inspection of test assumptions, and due to unequal variances between groups, Welch's correction was applied to independent t-tests to account for unequal variances.

Descriptive statistics, including measures of central tendency and frequency counts, were calculated for program and teacher survey data. Mean scores and percentage of responses by level of agreement/disagreement were calculated for teacher survey data.

3.3. Qualitative Methods

Qualitative data obtained via surveys and individual interviews was 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 findings of the analysis of individual interviews were collated and synthesised into a cross-case analysis matrix. This synthesis involved identifying major themes across interview participants in relation to their perceptions about the program, potential limitations and strengths, and impact on students, teachers and the community.

3.3.1. 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, 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. There were ten survey respondents from five different schools (Schools 1, 3, 9, 18 and 20). Seven respondents indicated they were classroom teachers, one respondent indicated they were a teacher aide, and one respondent indicated they were a school principal. Responses to open-ended survey questions were thematically coded using open coding techniques to ascertain response categories

In addition to the surveys, semi-structured interviews (n=9) were conducted with principals according to a co-designed interview protocol by senior evaluation team members. The interview protocol can be found in Appendix 1. Participants were from across two jurisdictions: WA and NT. Seven of the principals were from current program schools, and two principals were from recently withdrawn schools, one in 2018 and not included in any other analysis, and one in 2019 and included in the quantitative data sources.

3.3.2. Descriptive Analysis

Qualitative descriptive analysis occurred at the parent nodal level in NVivo. This descriptive analysis, or a 'broad brush' descriptive coding, derived directly from the interview protocol reflecting the topic structures contained within the protocol. 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.

3.3.3. 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. During this analysis stage, first-level nodes obtained through descriptive analysis (phase 1) were expanded into second level or child notes that represent specific and emerging sub-themes. For instance, interview data on program impact on students was further broken down into specific outcomes described by interview participants. Whilst this stage of the analysis was primarily conducted by one coder for consistency purposes, a second coder double-coded 30 per cent of interview data. Discrepancies between coders were analysed and resolved by a senior evaluation team member.

3.3.4. Componential 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 the figure below.

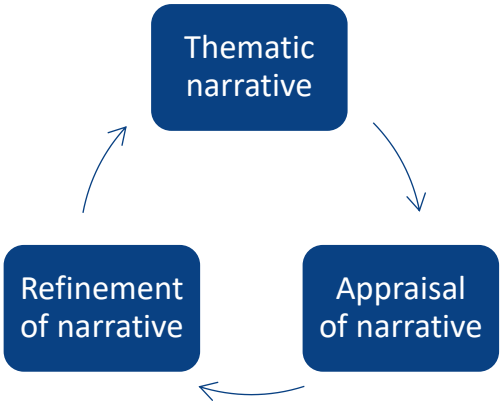


Figure 4: Iterative process of componential or comparative analysis.

Finally, divergences and convergences across interview participants were analysed using a cross-case analysis matrix. Within the matrix, coders noted thematic analysis findings - including first and second level nodes - alongside strength and salience of each theme (ie how frequently they were mentioned by participants, and emphasis ascribed to each theme). In addition to identifying common themes across participants, the cross-case analysis was employed to explore relationships between school contextual factors, program implementation practices, and outcomes for students, teachers and the community.

4. Results

This section of the report presents the aggregated results from each jurisdiction under the components of the CIPP model outlined earlier:

1. Context
2. Outcomes
3. Process

4.2. FLFRPSP Context

4.2.1. State and Territory Demographics

The FLFRPSP is implemented across four jurisdictions: Western Australian Government Schools, Western Australian Catholic Schools, Northern Territory Government Schools and Queensland Government Schools. In addition to these jurisdictions there are also a number of independent schools. The spread of program schools across these four jurisdictions and the program they are implementing (DI/EDI) can be found in the table below.

Table 3. Spread of Program Schools Across Jurisdictions

Jurisdiction	Number of DI Program Schools	Number of EDI Program Schools
WA Government	1	1
WA Catholic	2	1
NT Government	7	
NT Independent	2	
QLD Government		1
Total	12	3

In previous years of the evaluation (Dawson, Clinton, Koelle, & McLaren, 2018), outcomes have been reported for each jurisdiction. However, given the reduced number of program schools included in this report, results have been aggregated across jurisdictions to protect the anonymity of program schools.

4.2.2. School Demographics

The table overleaf provides an overview of demographic characteristics of the 15 schools participating in the FLFRPSP. The table shows that there is a significant amount of variation in the demographics of the schools implementing the FLFRPSP program. Number of full-time equivalent staff ranged from 3.9 to 85.5, depending on the size of the school. The Index of Community Socio-Educational Advantage (ICSEA) is a scale which provides an indication of the socio-educational backgrounds of students (ACARA, 2015). ICSEA has a set average of 1000, all program schools fall under this average. Socio-Educational Advantage (SEA) Quarters represent a scale of relative disadvantage (bottom quarter) through to relative advantage (top quarter) (ACARA, 2017). On average 86 per cent of program school students were considered in the bottom SEA. On average 85 per cent of students in program schools have a language background other than English (LBOTE), in some cases this was 100 per cent. On average there was a high Aboriginal and Torres Strait Islander student population in program school.

Table 4. Demographic Information for the 15 Program Schools Participating in the FLFRPSP

	Teachers*	Staff*	ICSEA	Students in the bottom SEA quarter %	Total Students	Boys	Girls	Indigenous %	LBOTE %
Mean	13.53	10.64	637.93	86.00	149.73	74.93	74.80	87.53	85.67
Standard Deviation	20.38	15.42	119.13	20.23	229.98	115.27	115.02	23.70	21.53
Range	3.9-85.5	0-60.9	546-960	32-99	17-946	7-470	8-476	12-100	33-100

**full time equivalent*

4.3. FLFRPSP Effects on Outcomes

Within this FLFRPSP evaluation, the product component of the CIPP model is referred to as impact and includes measuring the outcomes of the program among each of the stakeholders. This section of the report presents the outcomes of the evaluation for teachers, students and schools.

4.3.1. Teacher Outcomes

This section presents the teacher level outcome data from the GGSA School Reporting Tools (SRT) for 2015, 2016, 2017 and 2018. Data for each of the 15 schools has been aggregated to provide an overall understanding of a range of teacher outcomes. Teacher survey and principal interview data is also used to explore teacher outcomes.

Teacher implementation effectiveness across classroom organisation

Program data provides a measure of teacher implementation effectiveness for three program domains that include classroom organisation, instructional delivery and behaviour management and engagement. This allows us to look at the effectiveness of teachers in implementing the FLFRPSP, rather than overall teacher effectiveness. Classroom organisation measures teachers' use of allocated time, organisation and accessibility of materials, and recording of data. Overall, teacher effectiveness scores were consistently highest in 2016–2017, during peak program implementation.

The figure below shows teacher implementation effectiveness trends across the program implementation period, from 2015-2018 as reported in the GGSA School Reporting Tools for classroom organisation. The classroom readiness domain measures seven factors including that all students have been placement tested, placed into instructional groups, and at least 2.5 hours of instructional delivery is scheduled. For each year, teachers' average scores for classroom organisation was the highest, and remained consistent across all years.

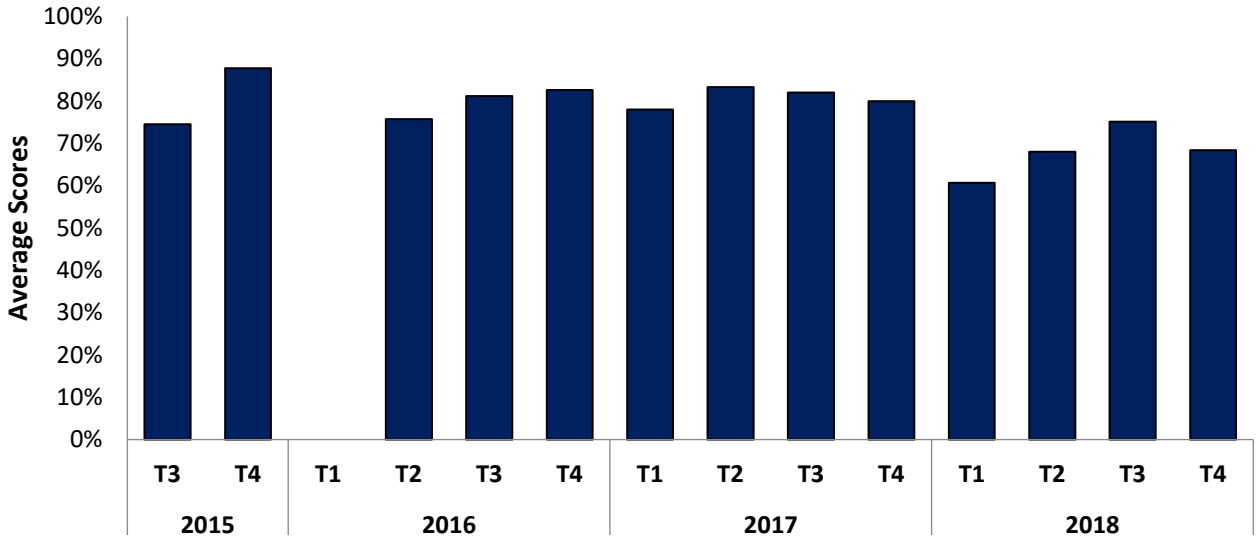


Figure 5. Teacher implementation effectiveness classroom organisation.

Teacher implementation effectiveness across instructional delivery

Instructional delivery entails the use of pedagogical activities, including assessments, corrections, use of repetition, pacing, and use of signals.

The figure below shows teacher implementation effectiveness trends across the program implementation period, from 2015-2018 as reported in the GGSA School Reporting Tools for instructional delivery. Instructional delivery in the FLFRPSP includes the use of signals and lesson pacing to allow for 'think time'. Teachers' instructional delivery scores show slight fluctuations across the years.

Figure 6. Teacher implementation effectiveness instructional delivery.

Principals who were interviewed had varied perceptions of the FLFRPSP impact on pedagogy. The interviews found that the structure of the program enhances consistency across teaching styles and supports a whole school approach to literacy that can be highly beneficial to new teachers. However, views on the benefits for teachers also varied, again in relation to the structured nature of the program: it provides consistency, but also limits teacher agency, for example in using a variety of methods and teaching styles.

Several principals mentioned positive outcomes for teachers in terms of improving their teaching practice and strategies, which one of the principals associated with the accountable nature of DI program:

> I definitely think it sort of, increases the concept of explicit instruction, data reliability, constantly tracking data and things like that, because we are required to be accountable. I think it does set up lots of good practices for teachers. (School 12, Principal)

In contrast, a few principals reported that opportunities for teachers to engage in critical reflection and review of their teaching practices were limited due to the rigid structure of the program:

> …there's a lack of critical literacy or critical… I mean we want people to be critical of their practice, and try and improve all the time. (School 39, Principal)

The staff survey also asked respondents if the FLFRPSP changed the way they teach. While half of survey participants responded that they did not believe the FLFRPSP had changed the way they taught, when asked to respond to seven statements about how the program had impacted upon a range of aspects including pedagogy, alignment of practice with school context, supports for teachers and individual benefit, there was a more positive response, as shown in the figure below. Fifty per cent of respondents agreed or strongly agreed that the program had improved their ability to teach literacy, 70 per cent somewhat agreed, agreed or strongly agreed the program benefits them as a teacher and 70 per cent somewhat agreed, agreed or strongly agreed that the program guides their teaching practice. Eighty per cent of respondents saw the program as a good match for their school.

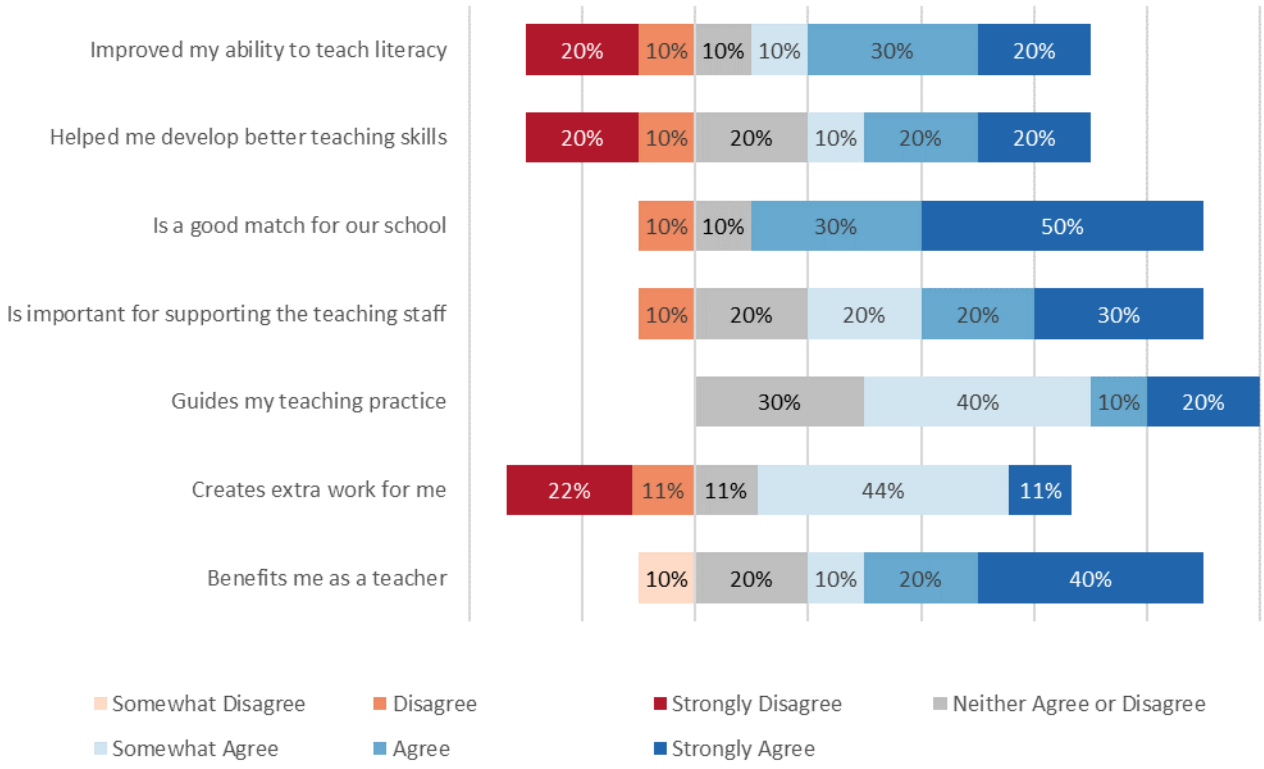


Figure 7. Teacher perception of the program's impact on their teaching practice (n=9-10).

Teacher implementation effectiveness across behaviour management

Behaviour management entails task transitions, teacher engagement during independent and group work, maintaining student motivation, and providing positive behavioural support to students.

The figure below shows teacher implementation effectiveness trends across the program implementation period, from 2015-2018 as reported in the GGSA School Reporting Tools for behaviour management. Behaviour management in the FLFRPSP includes positive management techniques and the use of motivation systems. Teachers' behaviour management scores show slight fluctuations across the years.

Figure 8.Teacher implementation effectiveness behaviour management.

Consistent with principal interviews, survey respondents held varying views of the FLFRPSP impact on teaching. Half of survey respondents agreed that FLFRPSP had changed the way they taught and provided responses that described the effects of FLFRSP on teaching practice. This included positive effects on how they managed student behaviour, the way they now structure their class to deliver lessons, and positive effects on self-efficacy to improve student learning outcomes. One respondent indicated that the program had not changed the way they taught, but suggested that the program was restricting the teaching methods and styles that were used.

Behaviour management

When asked to expand on the ways their practice had changed, a few respondents believed it had improved the way they managed student behaviour by using positive reinforcement strategies that include positive interventions and rewarding behaviour immediately:

> Use of positive interventions eg rules such as sit tall - rewarding behaviours immediately. Telling children then getting children to give answers (eg more support)! Getting children to answer on signal. (School 18, Teacher 3)

Implementation structure

When asked how their practices had changed, one teacher indicated that their implementation structure had changed. Highlighting that implementation had moved from whole class delivery to group-based work:

> Not delivering whole class literacy. (School 1, Teacher)

Teacher self-efficacy

Finally, one respondent commented on their belief in their ability to increase student literacy skills by using the program, demonstrating an increase in self-efficacy:

> I feel I finally teach something that stays with the kids and actually progress their literacy skills quickly. (School 9, Principal)

Limited their Teaching Style

When asked if the program had changed their teaching practice, one teacher indicated that the program had not fundamentally changed the way they taught, but highlighted that they felt implementing the program restricted their ability to use their preferred teaching style:

> Yes and no. I am teaching the program but feel I cannot bring my own skills and strengths in literacy to my teaching. (School 3, Teacher 2)

Overall Summary of Teacher Outcomes

- Overall teacher effectiveness scores were at their highest in 2016- 2017, during the peak of program implementation, and have decreased in 2018
- The staff survey and principal interviews showed a varied perception of the impact of the FLFRPSP on teacher practice
- The majority of survey respondents believed implementing the FLFRPSP had positively impacted their teaching; one survey respondent believed it had limited their teaching practice
- Principal interviews were consistent with survey responses. Several principals mentioned positive outcomes for teachers in terms of improving their teaching practice; however, a few principals believed it limited teaching practice and opportunities for teachers to engage in critical reflection

4.3.2. Student Outcomes

The following section presents the results of standardised assessments that are considered indicators of program impact on student outcomes, as well as program data on student outcomes and school staff perception of student outcomes.

In addition to NAPLAN data, state/jurisdictional literacy data such as Early Years Literacy and Numeracy Data (EYLND) was sought. However, only one school gave consent to access their EYLND data, and as such it has not been included in this report.

Positive direction of growth

To illustrate the direction of growth, the average NAPLAN scores for Year 3 2015 and Year 5 2017, and Year 3 2016 and Year 5 2018 for program and control schools are presented against the national, remote, and very remote averages. There are limitations in this comparison, as the national, remote, and very remote average NAPLAN scores use matched data. However, due to the limited sample size and student turnover in the program and control schools, these scores are unmatched, as use of matched data for the program schools would have resulted in a significant reduction in available data for analysis.

When comparing growth of program and control groups across 2015-2017 and 2016-2018 time periods, both groups showed a positive direction of growth. The mean changes for these results are further discussed in the following section.

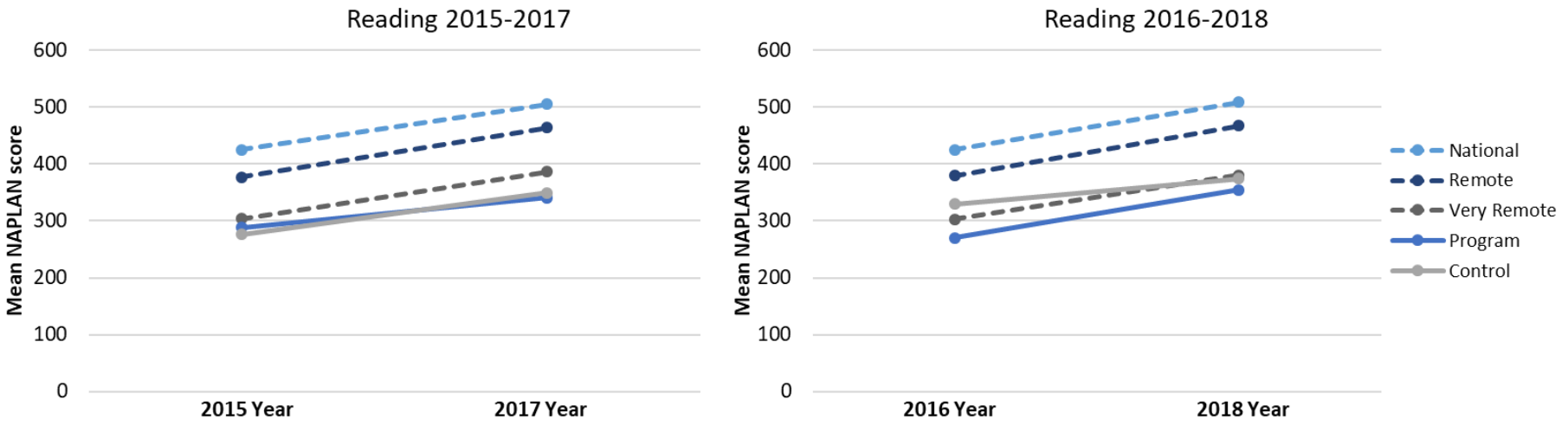


Figure 9. Mean NAPLAN Score Reading 2015-2017, 2016-2018.

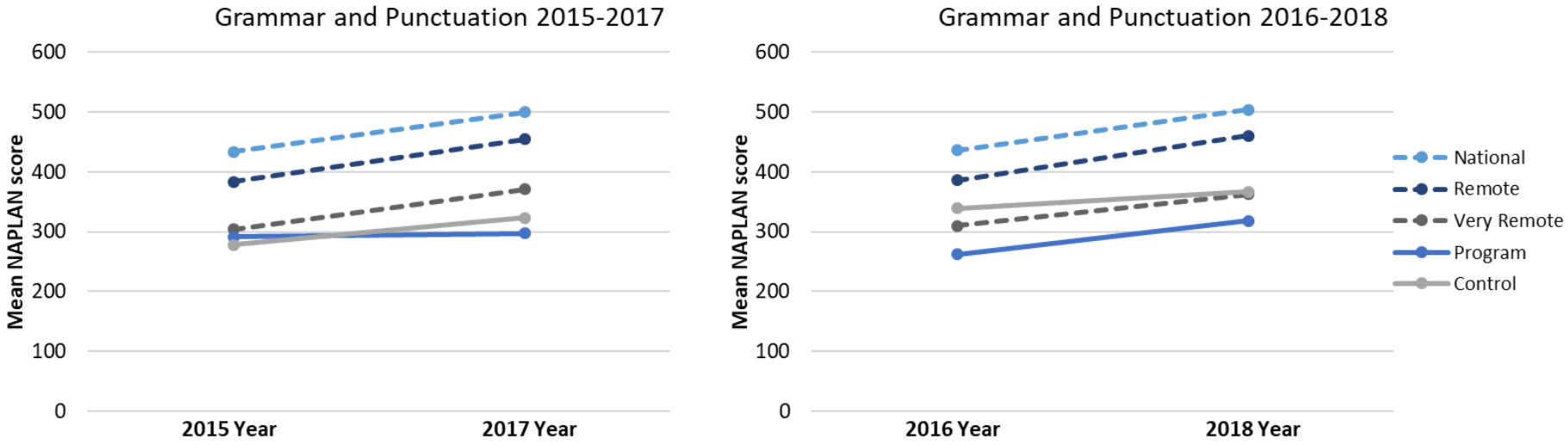


Figure 10. Mean NAPLAN Score Grammar and Punctuation 2015-2017, 2016-2018.

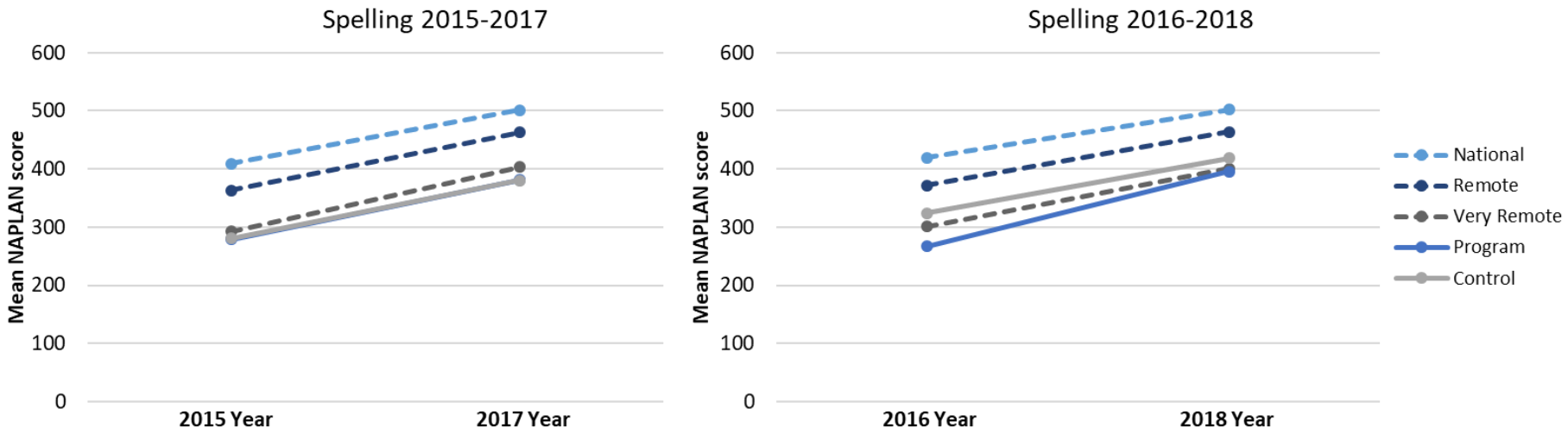


Figure 11. Mean NAPLAN Score Spelling 2015-2017, 2016-2018.

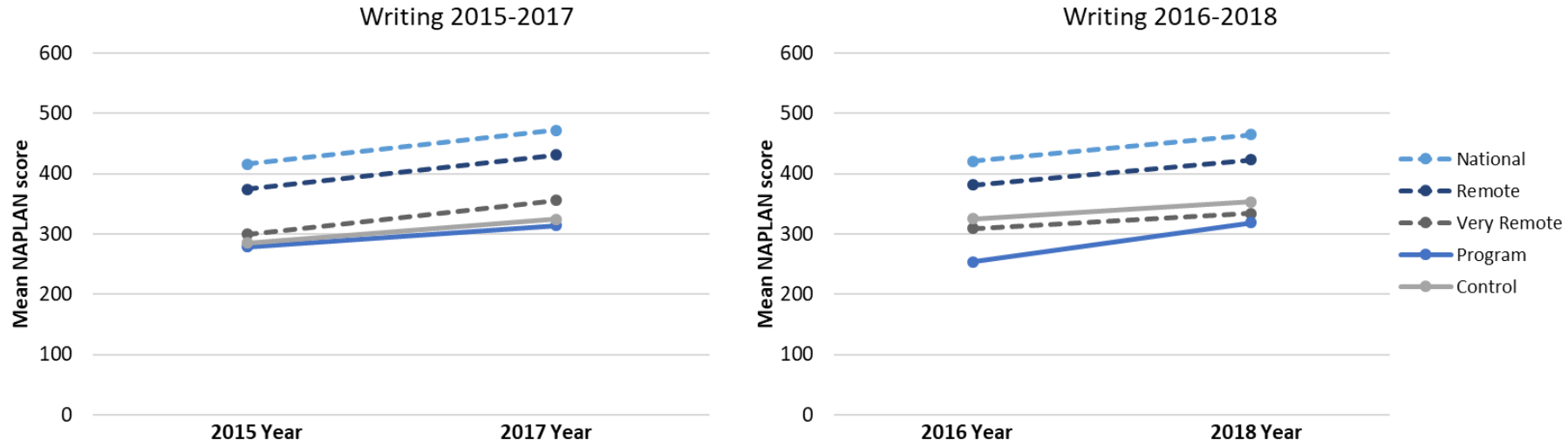


Figure 12. Mean NAPLAN Score Writing 2015-2017, 2016-2018.

Mean changes

An analysis of NAPLAN data was conducted to compare change in mean scores for Year 3 and Year 5 in 2015-2017, and Year 3 and Year 5 in 2016-2018 for all cohorts. This was undertaken for NAPLAN Reading, Writing, Spelling, and Grammar and Punctuation domains, and is presented in the figure overleaf. Due to limited availability of 2018 NAPLAN scores for some schools, change scores could not be calculated for all schools.

The majority of program schools demonstrated a positive mean change in all four NAPLAN literacy domains as consistent with previous years of the evaluation; the greatest effect was observed in Spelling. All schools recorded a positive mean change in Spelling. A positive mean change was observed for School 27 in this domain. However, as the confidence interval includes zero, it cannot be confidently determined if the mean change was either positive or negative. Schools showed the least improvement in Grammar and Punctuation (GP), with Schools 20, 11 and 17 recording negative mean changes. School 27 also recorded a negative mean change. However, as the confidence interval includes zero, it cannot be determined if the mean change was positive or negative. Negative mean changes were also recorded in Reading (Schools 7 and 3), and Writing (School 27). While there has been an overall increase in mean score, most schools are below the national average in Reading, Writing and Grammar and Punctuation. Eight schools were above the national average in spelling.

Several challenges were encountered with the analysis of 2016-2018 data compared to previous analyses of 2015-2017 data and, in particular, effect size calculations (see Section 4.2.3). The reduced number of program schools across all jurisdictions impacted overall effect sizes compared to previous years.

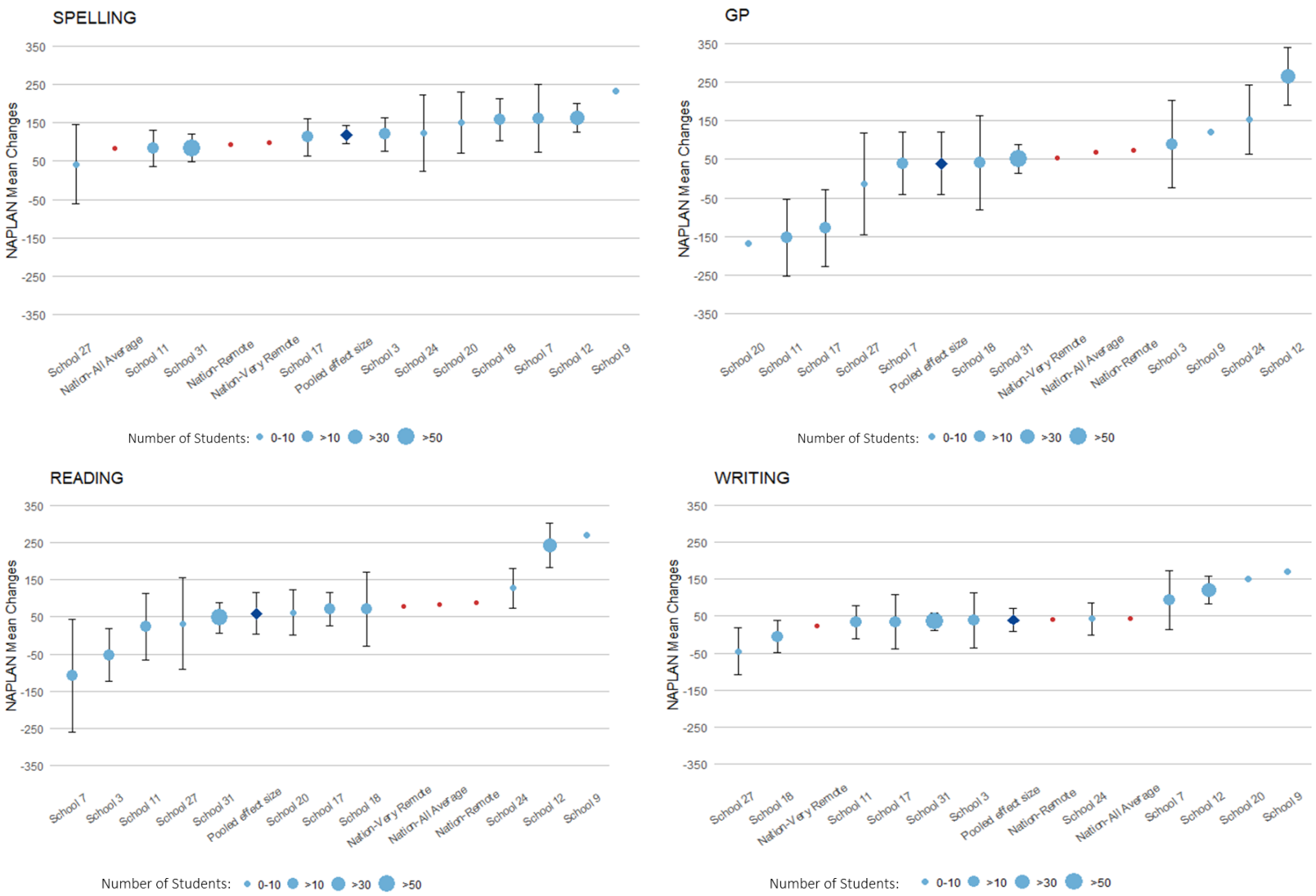


Figure 13. Changes in NAPLAN mean changes compared to national, remote and very remote averages.

The table below provides the pooled mean difference for 2015-2017 and 2016-2018 for each NAPLAN domain. The pooled mean takes into consideration the variation in number of students in each of the schools in the sample. For both 2015-2017 and 2016-2018, the largest pooled mean difference was observed in the domain of Spelling for both program and control schools, indicating the most growth.

Table 5. NAPLAN Pooled Mean Difference and Standard Error

NAPLAN Domain	Group	2015-2017 Pooled mean difference *(Standard Error)*	2016-2018 Pooled mean difference *(Standard Error)*
Writing	Program	66.1 *(13.0)*	40.0 *(15.8)*
	Control	40.2 *(11.3)*	38.2 *(11.7)*
Spelling	Program	130.3 *(9.0)*	119.5 *(12.1)*
	Control	101.8 *(9.1)*	118.8 *(15.3)*
Reading	Program	63.8 *(14.0)*	59.7 *(28.5)*
	Control	72.9 *(10.1)*	92.9 *(21.0)*
Grammar and Punctuation	Program	29.8 *(22.6)*	39.8 *(41.7)*
	Control	45.6 *(13.1)*	43.2 *(12.0)*

Differences in mean changes between program and control schools

The magnitude of the difference in mean NAPLAN scores from 2015 to 2017 and 2016 to 2018 is discussed in this section using pooled effect size coefficients. All pooled effect sizes were calculated using Hedges' g, as this coefficient is considered to be more suitable for small sample sizes (n < 20). Pooled effect sizes take into consideration the variation of number of students in the sampled schools. Pooled effect sizes are calculated across the two-year period for program and control schools 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) is used. As seen in the table below, a score of 0.80 or above is considered large.

Table 6. Hedges' g Effect Size Interpretation

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

Effect Size Explained

To assess the magnitude of the observed effect or difference in mean scores between groups, effect sizes for each result are calculated. There are different effect size statistics associated with specific tests; in this report, we have employed Hedge's g.

Effect sizes of the mean difference in scores between 2016 and 2018, by NAPLAN domain, were aggregated to calculate a pooled effect size for program and control schools. The figure overleaf shows that for program schools, a large effect size was observed for Spelling (Hedge's g= 1.95); medium effect sizes were observed for Reading (Hedge's g= 0.59) and Writing (Hedge's g= 0.71) and a small effect size was observed for Grammar and Punctuation (Hedge's g= 0.27). Overall, program schools showed a greater effect size than control schools in the Writing and Spelling domains, and a smaller effect size in Reading, and Grammar and Punctuation. Welch's t-test was used to determine if there was a significant difference in effect size coefficients between program and control schools for each NAPLAN domain. No statistically significant differences (p> 0.05) were detected, for Reading (t(20) = -0.7, p = 0.49), Spelling (t(19) = 0.1, p = 0.92), Writing (t(12.02) = -0.7, p = 0.5), or Grammar and Punctuation (t(11.61) =- 0.11, p= 0.92).

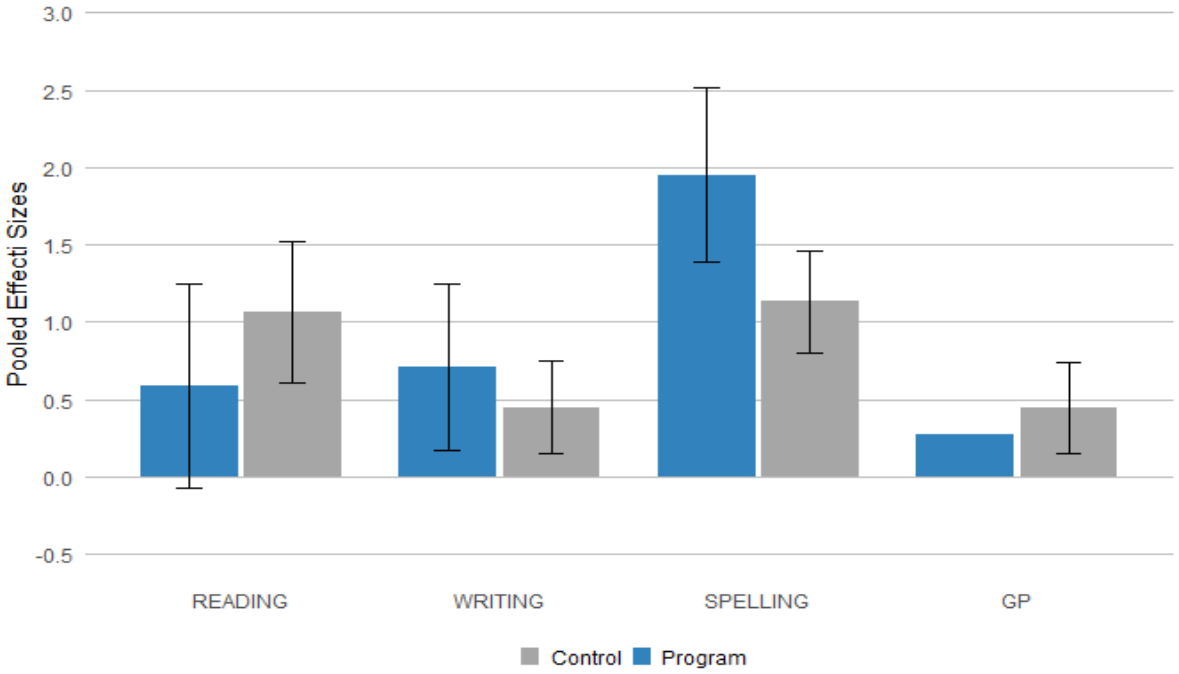


Figure 14. Program and control NAPLAN pooled effect sizes 2016-2018.

The figure below presents the effect sizes from 2015-2017, compared to the effect sizes from 2016-2018. The figure demonstrates that program schools have experienced slightly smaller effect sizes in 2016-2018 in Reading (Hedge's g= 0.59), Writing (Hedge's g= 0.71) and Spelling (Hedge's g= 1.95) compared to 2015-2017 (Hedge's g= 0.71, Hedge's g= 0.78 and, Hedge's g= 2.04). For program schools, there has been a slight increase in effect size in Grammar and Punctuation (GP) (Hedge's g 2015-2017= 0.22, Hedge's g 2016-2018= 0.27). The effect size for Grammar and Punctation has remained relatively consistent. While there is no significant difference in any of the NAPLAN domains, program schools have shown a greater effect size in Spelling and Writing compared to control schools across the duration of the program.

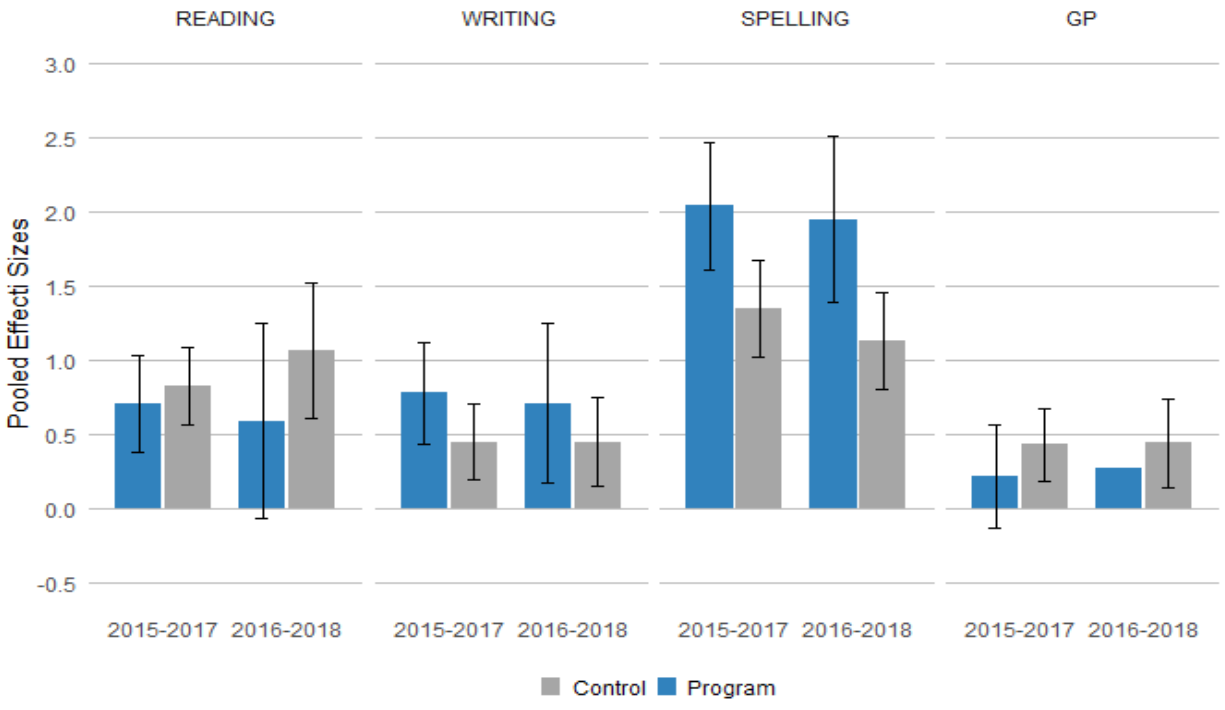


Figure 15. Program and control NAPLAN pooled effect sizes 2015-2017, and 2016-2018.

Program student mastery data

An analysis of student mastery data was conducted to compare student mastery for each year of implementation across three different FLSRPSP programs, Language, Reading and Spelling. Student mastery is collected by teachers implementing the program and is calculated by the total of students passing a program divided by the total number of opportunities to test. The average number of students reaching mastery has been variable over the life of the program. Average mastery in spelling has fluctuated between 2016 and 2018, from 81.82 per cent in 2016 to 90.77 percent in 2017, and then decreasing to 77.78 per cent in 2018. Average language master increased from the beginning of implementation in 2015 to 2018, with an average language mastery of 61.64 per cent in 2015 and an average language mastery of 77.25 per cent in 2018. However, similar to the trend in spelling, the average language mastery has fluctuated over the life of the program. Overall reading mastery has remained relatively stable, maintaining 74 per cent mastery between 2015 and 2016, before decreasing slightly to 72.33 per cent in 2017 and then increasing slightly to 76.25 percent in 2018.

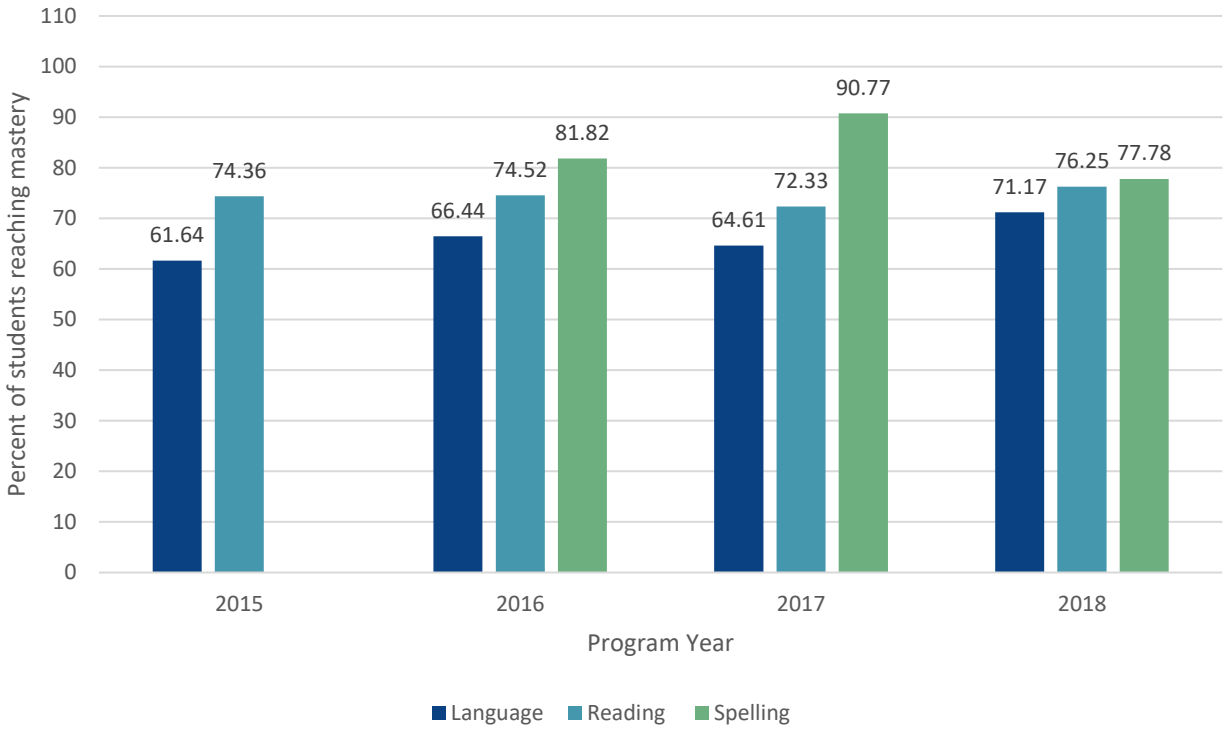


Figure 16. Average student mastery

Mastery data was further split by school. Twelve of the 15 schools recorded student mastery across 2015, 2016, 2017 and 2018 for language and reading. Nine of the 15 schools recorded student mastery across 2016, 2017 and 2018. The three figures overleaf show that student mastery was variable across schools and across years for each of the three programs. There was less mastery data recorded for the spelling program, compared to the language and reading programs. The table below shows that there were less opportunities to test for spelling than language and reading, indicating that overall there are fewer spelling program lessons taught compared to language and reading program lessons.

Table 7. Opportunity to test and number of students passed

	2015		2016		2017		2018	
Subject	Students passed	Opportunities to test	Students passed	Opportunities to test	Students passed	Opportunities to test	Students passed	Opportunities to test
Language	421	683	495	745	1331	2060	2967	4169
Reading	1894	2547	1392	1868	2198	3039	7243	9499
Spelling			72	88	118	130	623	801

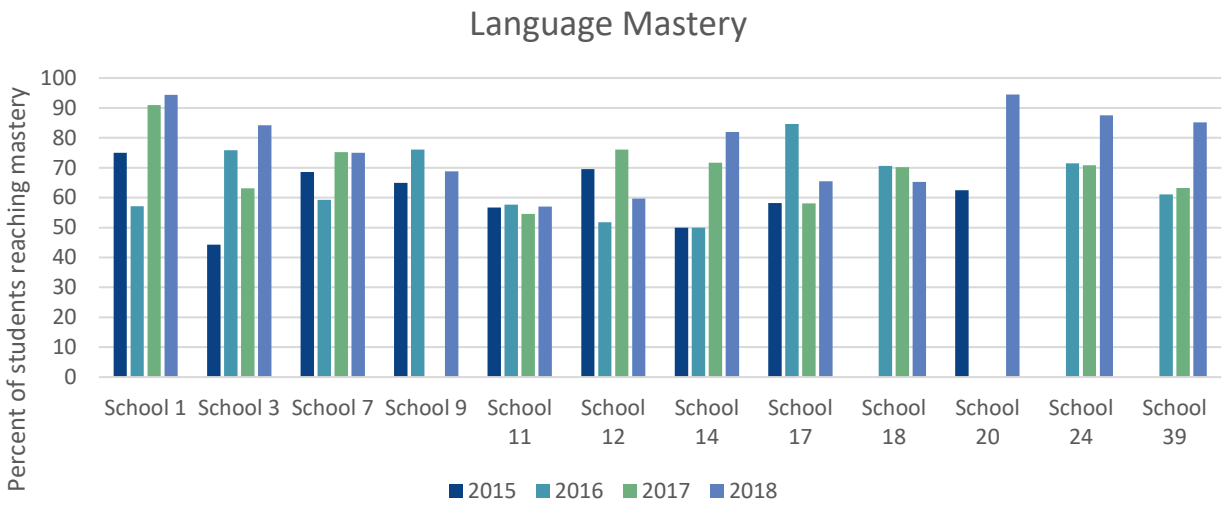


Figure 17. Average student language mastery by school

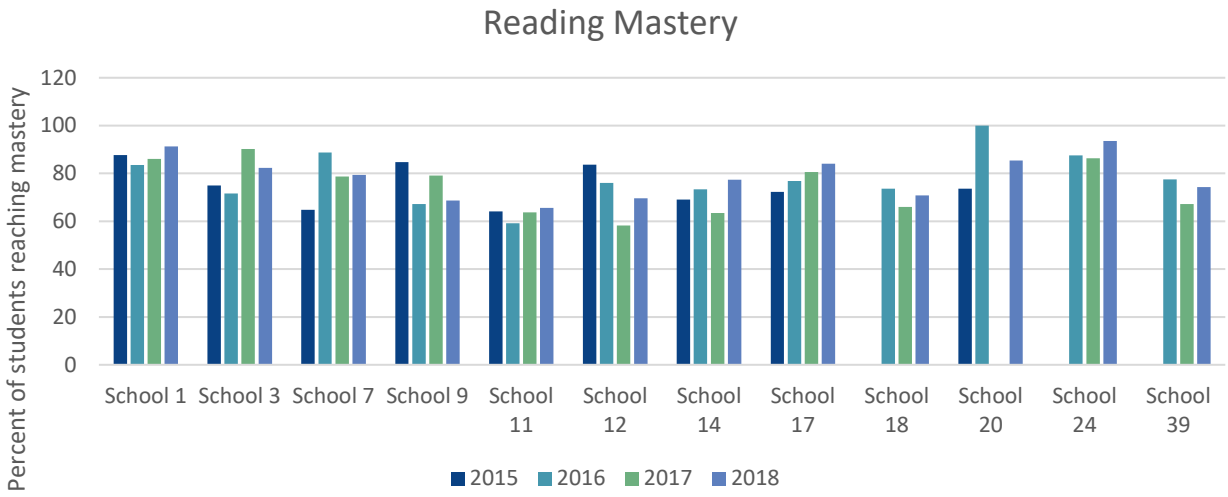


Figure 18. Average student reading mastery by school

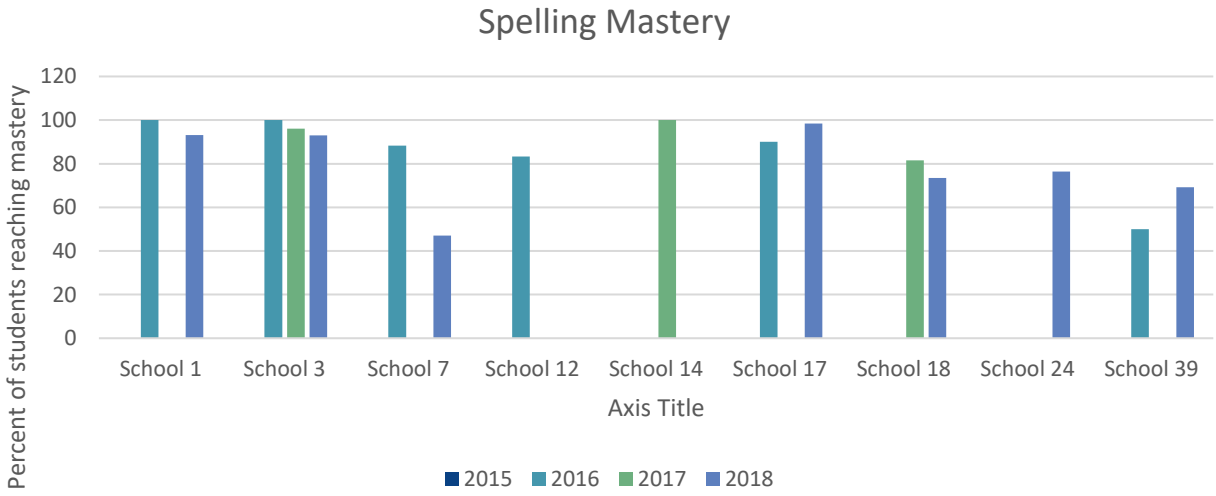


Figure 19. Average student spelling mastery by school

Program year level equivalent change

Each FLFRPSP language, reading and spelling program has a designated year level equivalent. An analysis of year level equivalent change was conducted as an indication of student progress. The DISE and LANG programs did not have a year level equivalent and were excluded from analysis. Using the starting program and the ending program of each term, year level equivalent change was calculated for 2017 and 2018 language, reading and spelling programs. Results were then group by year by summing the grade change for each term in a single year, while this means the data is not matched across the year, it gives an indication of change across one year. The end program was not recorded for 2015 and 2016 and therefore year level equivalent change was unable to be calculated. The four pie charts below and two overleaf demonstrate that the majority of students maintained their year level for language, reading and spelling over both 2017 and 2018. For each program, in each year a small number of students increased their year level equivalent. As is consistent with NAPLAN data, the most growth was observed in spelling, with seven per cent of students increasing their year level equivalent in 2017 and 11 per cent increasing their year level in 2018. Only one student decreased their year level, in the 2018 reading program.

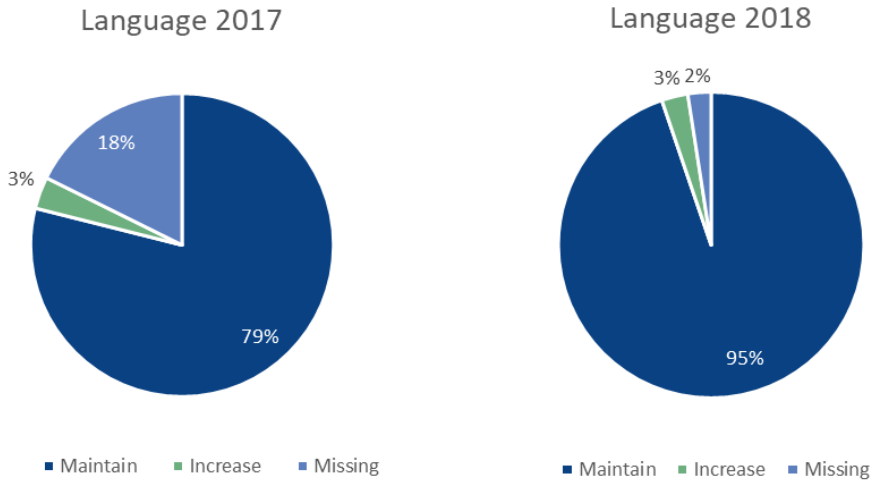


Figure 20. Language program year level equivalent change 2017 (n=147) and 2018 (n=325)

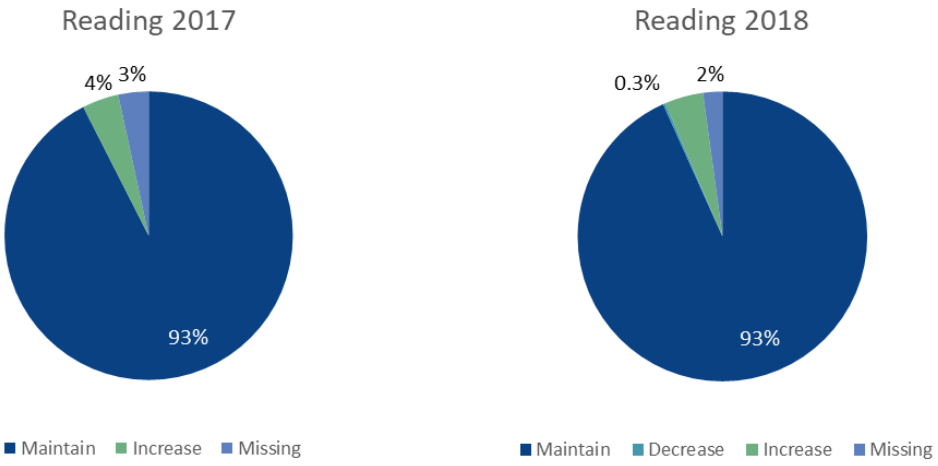


Figure 21. Reading program year level equivalent change 2017 (n=147) and 2018 (n=372)

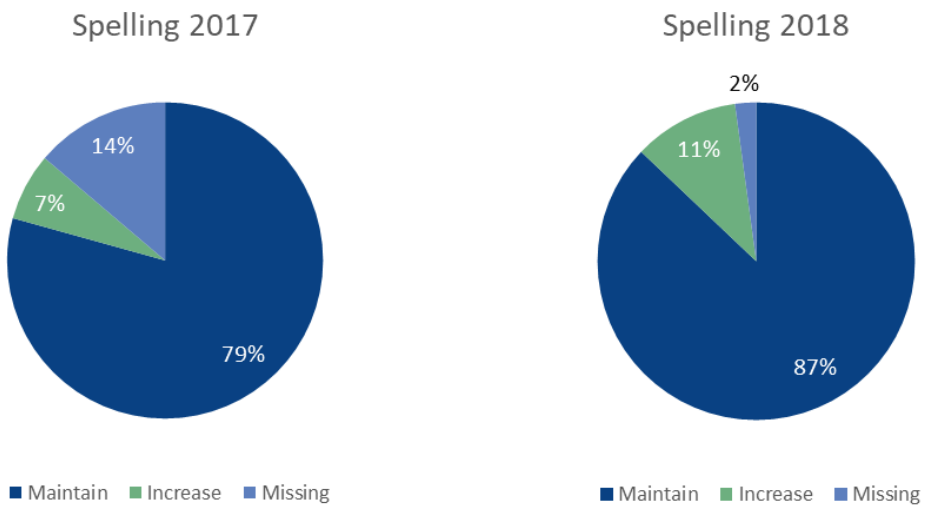


Figure 22. Spelling program year level equivalent change 2017 (n=29) and 2018 (n=93)

Overall summary of student assessment

- Mean changes between 2016 and 2018 in all NAPLAN literacy domains were varied
- Most program schools demonstrated a positive mean change in all NAPLAN literacy domains. Four schools demonstrated a negative mean change across the different domains, and a few recorded confidence intervals included zero, so we are unable to determine the direction of the mean change
- The largest effect size was observed in Spelling in program and control schools across 2015-2017, 2016-2018
- Overall, there was no significant difference in effect sizes between program and control schools
- Consistent with NAPLAN data, there was variation in student mastery across all FLFRPSP programs for the duration of the program
- While overall spelling mastery decreased between 2017 and 2018, consistent with NAPLAN data, the most growth was observed in spelling in both 2017 and 2018
- The majority of students maintained their year level across 2017 and 2018, with a small portion of students increasing their year level.

Perception of the program in their school

All respondents answered why or why not they felt the program was positive for their school. Overall, the majority of survey respondents felt the program has been positive for their school and three key themes were emergent; improved student learning, the program mitigated teacher turnover, and teaching students in ability groups rather than at their age level as a factor that supported student learning at their own pace:

> Yes. It's awesome for us to have each student being on their own level and being able to progress at their own pace. (School 1, Teacher)

One survey respondent felt the program was not positive for their school as it did not provide enough flexibility for teachers to use different teaching methods and to cater to different needs of students, and one respondent had not yet decided whether they thought it was positive or not.

Improved student learning

Of those respondents that felt the program was positive overall, several respondents identified improved student learning as the reason why. In contrast to the NAPLAN data which identified the greatest student growth in Spelling, and a slight decrease in reading growth, a few respondents specifically mentioned improved reading:

> The students themselves recognise that they are readers. Generally, students do not remember their reading growth, but these students do. They keep records of books that they can read. (School 20, Teacher 1)

The perception of improved student learning was supported by principal interviews. The majority of principals still engaged with the program reported positive student outcomes as attributed to the program. Most of the principals mentioned positive outcome associated with improvements in students' literacy level, alongside other outcome areas including behaviour and attendance. One principal associated the improvement in students' behaviours with the element of routine of the program:

> When we first brought the program in, as I said, the kids that were used to being able to sleep in the back of the room, we had 20, 30 kids sometimes running around the school refusing to go to class, so to start with, it made behaviour much, much worse. But as we got the kids in, as we got into the routines, it's helped a lot. (School 3, Principal)

Varied success

One survey respondent did not respond to whether they thought the program was positive overall, but provided a response in the open-ended section to expand on the fact they felt it did not allow for differentiation for varied student needs. This teacher was from the same school as a respondent who felt the program was positive overall for their school, highlighting the variation in perception even within schools:

> I am still on the fence. It is successful for a minority. It causes complete disengagement from learning for other cohort for students. (School 3, Teacher 2)

Consistent with survey respondents, interview participants' perceptions about the impact of the program on students were varied. Whilst some participants reported increased students' literacy gains, and perceptions of improved confidence and classroom behaviour, a few participants reported no change or improvement in standardised test scores. There was no clear relationship between program implementation/adaptation and outcomes for students. The schools either received low NAPLAN and Progressive Achievement Test (PAT) results or no obvious change in NAPLAN results:

> NAPLAN results are no better and no worse than what they were four years ago. (School 39, Principal)

Specifically, two principals mentioned that the program was not suitable and may result in detrimental outcomes for students with special needs, for example, hearing loss or learning difficulties.

Overall perception of program impact on student outcomes

Survey respondents were asked to provide their perspectives and experiences of the FLFRPSP in the context of their school and community, and the perception of student outcomes. All respondents agreed that the program had made a difference to students and that it had contributed to improved literacy, while most respondents agreed that students had engaged with the program. There was more variation in responses to other outcomes relating to wellbeing and attendance.

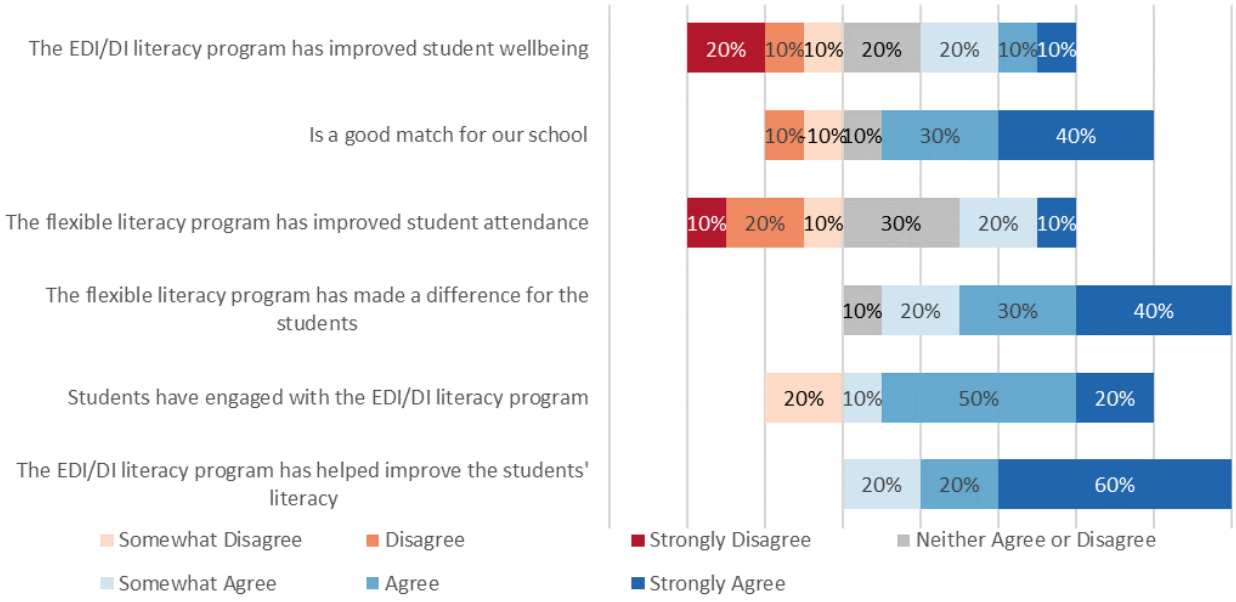


Figure 23. Teacher perception of the program's impact on students (n=10).

Outcomes other than literacy gain

Besides literacy, survey respondents were asked if there were any other outcomes for students. The majority of participants perceived that there were other benefits of the program besides literacy. Other benefits identified by respondents included improved student behaviour and/or attention and improved student motivation to learn and recognition of their own learning.

Improved student behaviour and/or attention

Several respondents observed improved student behaviour and/or attention as a result of the program and the increased use of positive reinforcement behaviour strategies (as mentioned in Section 4.3.1), allowing students to focus longer improving learning:

> The children are able to focus for longer periods. The positive behaviour strategies taught have improved their general behaviours in general. (School 18, Teacher 3)

Student motivation and awareness of learning

A few teachers highlighted improvements in student motivation to learn and an improvement in their ability to observe their own learning:

> Students have a sense of progression and learning. (School 3, Teacher 1)

Reasons for overall program success

The majority of survey respondents felt that the program had been successful overall. The reasons for success identified by respondents were varied and included an increase in student literacy, the structure of the program that requires mastery, and an improvement in other student outcome areas such as behaviour, attendance, awareness of learning and confidence.

Improved student literacy

A few survey respondents believed the FLFRPSP had been successful overall because of their perceived improvements in student literacy:

> My students are reading and starting to write. (School 3, Teacher 1)

Students work at their ability level

A few survey respondents believed it was the structure of the program that requires students to work at their ability level before they reach mastery and move onto the next level, that contributed to the overall success of the program:

> This program helped our kids learn in their own time. (School 3, Teacher Aide)

As well as the requirement of mastery, one respondent also highlighted the programs use of reinforcing previous learning as reason for the program's overall success:

> I believe the key to the success of the program is that it continually reinforced previous learning. Insistence on mastery. Students working at the level that they succeed at. No surprises for the students. Predictable. (School 18, Teacher 1)

Improved student behaviour, attendance and confidence

In addition to literacy outcomes, a few survey respondents also attributed other student factors to the overall success of the program, including improved behaviour and improved attendance:

> [School 20] students were unruly in the classroom. They struggled with reading and gave up, put no effort into reading. These days they are focused, self-motivated and they enjoy the reading experience. Generally, a positive experience. Sometimes we have the odd day of "this is boring". (School 20, Teacher 1)

Improved student confidence was another student factor identified by one participant as why they consider the program successful overall:

> I have seen the children grow in confidence in this program. They have made great progress with good attendance. It is a very supportive program with excellent resources. (School 18, Teacher 3)

Limited staff as a barrier to realising the full potential of the FLFRPSP

One survey respondent felt the program had been successful overall, but pointed to the challenges of implementing the program with limited staff as a factor that meant it was not as successful as it had the potential for:

> It would have been very successful if schools stuck with it more than three years and if staffing levels had been appropriate. DI came at a time of lowest staffing levels in years. (School 9, Principal)

Poor attendance and engagement as barriers to success

The two survey respondents who felt the program had not been successful overall highlighted poor attendance and poor engagement in some students as the reason for the perceived lack of success. One of the respondents who felt the program had not been successful, was from a school where two other teachers believed the program had been successful overall:

> There has been some success with our older students, but those children not engaged with the program are slipping through the net and falling behind. (School 18, Teacher 2)

Overall summary of perception of student outcomes

- There was a varied perception of the success of the FLSPRSP for student outcomes
- Several survey and interview participants identified improved student literacy as an outcome; however, a few survey and interview participants believed the FLFRPSP had limited or no impact on literacy
- The largest effect size was observed in Spelling; however, survey respondents observed change in Reading
- Survey and interview participants identified improved behaviour as an additional student outcome

4.4. School Profile Cross Case Analysis

The data from 15 school profiles was analysed to determine key themes relating to enablers and barriers. The results were triangulated across all profiles.

Once the themes were identified, a separate researcher read each of the profiles and triangulated results by recorded how each school met each of the key enablers and barriers. Eleven emerging themes were identified and are defined in the table below along with their source in the school profiles.

The following matrix presents the findings of the cross-case analysis. The data sources available for each school varied, resulting in limited information on some of the criteria, namely, systems support, infrastructure to support implementation, resources, flexibility for context, evaluation and assessment capability, data use, and continuity.

For ease of reading the cross-case analysis table overleaf has been colour coded to denote positive (green), negative (red), varied (orange) and not specified in interview or survey (blue) responses to enablers and barriers.

Table 8. Key Themes Relating to Enablers and Barriers

School	Knowledge, skill and training	Buy-in from teachers	Buy-in from leadership	Systems support	Infrastructure to support implementation	Resources	Consistency in delivery	Flexibility for context	Evaluation and assessment capability	Continuity
School 3	Declining	Varied	Varied	Not specified	Good	Good	Consistent	Variable	Good	High
School 31	Varied	High	High	Good	Good	Good	Varied	Good	Not specified	High
School 18	Varied	Varied	Low	Varied	Not specified	Good	Consistent	Variable	Not specified	Low
School 24	Declining	Varied	Varied	Varied	Varied	Not specified	Varied	Poor	Not specified	Not specified
School 7	Declining	Varied	Varied	Not specified	Not specified	Not specified	Increasing	Not specified	Not specified	Not specified
School 9	Declining	Varied	Varied	Not specified	Poor	Not specified	Varied	Good	Not specified	Not specified
School 29	Declining	High	High	Not specified	Not specified	Not specified	Declining	Not specified	Not specified	Not specified
School 11	Varied	Varied	Varied	Not specified	Good	Good	Declining	Not specified	Not specified	Low
School 12	Declining	Varied	Varied	Poor	Good	Good	Varied	Poor	Poor	High
School 20	Declining	Varied	Varied	Poor	Varied	Varied	Declining	Variable	Poor	Not specified
School 14	Declining	Low/Varied	Varied	Not specified	Not specified	Not specified	Varied	Not specified	Not specified	Not specified
School 1	Declining	Varied	Low	Not specified	Not specified	Not specified	Consistent	Variable	Not specified	Not specified
School 17	Varied	Varied	Varied	Poor	Varied	Varied	Consistent	Poor	Not specified	Not specified
School 27	Declining	Varied	Varied	Not specified	Not specified	Not specified	Varied	Not specified	Not specified	Not specified
School 39	Declining	Varied	High	Poor	Not specified	Varied	Consistent	Poor	Poor	Not specified

While the key enablers and barriers varied for all schools across the years, several elements are common across all the schools, in particular the lack of accessible data to explore these various themes. The following table defines the themes.

Table 8. School Profile Enablers and Barriers Definitions

Enablers and Barriers	Definition	Source
Knowledge, skill and training	The implementation of teacher training for knowledge and skill development	GGSA school reporting tools (training and observations)
Buy-in from teachers	Program engagement of teachers through the duration of the program	GGSA school reporting tools (staff turnover), principal interviews, staff survey
Buy-in from leadership	Program engagement of leadership through the duration of the program	GGSA school reporting tools (staff turnover), principal interviews, staff survey
Systems support	Quality of system support received	Principal interviews
Infrastructure to support implementation	Perception of school factors and infrastructure that support implementation (eg leadership support)	Principal interviews, staff survey
Resources	Quality of program resources	Principal interviews, staff survey
Consistency in delivery	Fidelity of program implementation across years	GGSA school reporting tools (school fidelity)
Flexibility for context	Perceived flexibility of the program to school's context	Principal interviews, staff survey
Evaluation and assessment capability	Capability of evaluation and assessments to determine student outcomes	Principal interviews, staff survey
Data use	Use of data by teachers to determine student outcomes	Principal interviews, staff survey
Continuity	Indication of staff as to whether their school will continue with the program	Principal interviews, staff survey

Overall summary of school profile cross-case analysis

- Thematic analysis of school profiles identified eleven key enablers and barriers
- Most schools showed varied engagement with each of the key enablers and barriers

4.5. Process Factors Mediating the Effects of the FLFRPSP

The process component of this evaluation focuses on several implementation factors, including fidelity, dosage, and barriers and enablers to implementation quality. These factors are considered at school, teacher and student levels and help explain the outcomes observed.

4.5.1. Engagement in Training and Observations

This section presents the teacher level data from the GGSA School Reporting Tools (SRT) for 2015, 2016, 2017 and 2018. Data for each of the 15 schools has been aggregated to provide an overall understanding of implementation and fidelity.

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.

The figure below shows the average number of teacher training and observation sessions for the 15 program schools running either the DI or EDI program.

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.

Based on aggregated school data, observation sessions are most frequently conducted at the start of the year (Term 1 and Term 2) and gradually decrease towards the end of each year. There is a more consistent trend observed across each year for the number of training sessions conducted.

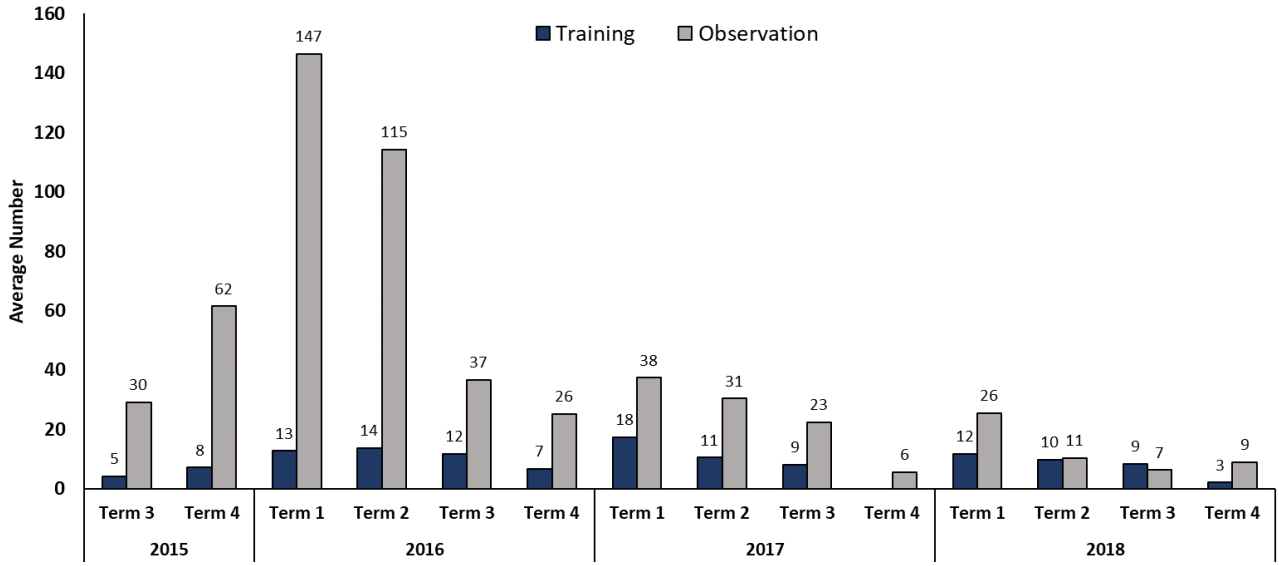


Figure 24. Average training and observations.

Perception of training

Nine survey respondents had not previously completed the survey and answered questions on implementation and training, while one respondent had previously completed the survey and continued on to questions regarding program impact. All continuing respondents somewhat agreed, agreed or strongly agreed that they had a good understanding of the pedagogical approach to EDI/DI, their role within the EDI/DI program, and the specific tasks and activities students are required to perform as a result of the training. While all respondents also agreed that they had a good understanding of how to teach literacy, 11 per cent of respondents only somewhat agreed, as shown in the figure overleaf.

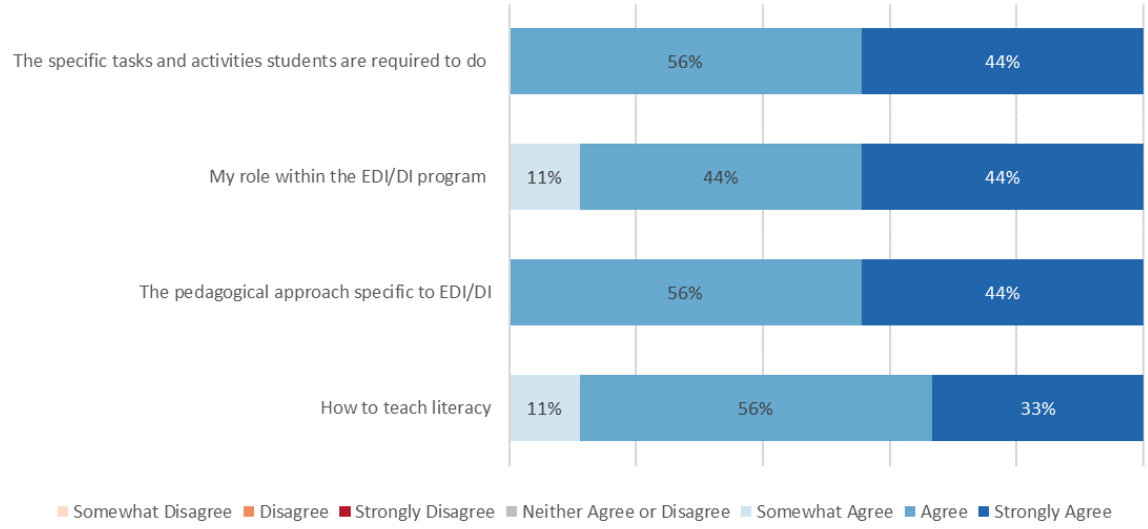


Figure 25. Teacher understanding of literacy and EDI/DI after participation in training (n=9).

Two respondents had not attended training for EDI/DI. This was due to working in a remote school with restricted travel during wet season and the training not having been provided since the respondent began their new role. The respondent indicated that in-school training was provided. Survey respondents were asked about implementation of EDI/DI after attending the training as shown in the figure overleaf. All respondents somewhat agreed, agreed or strongly agreed that the goals and potential benefits of EDI/DI were clear to them after the training. Most respondents were positive about the resources they were given to implement the program after training. Most respondents somewhat agreed to strongly agreed that they felt ready to implement the program in their classrooms, however one respondent disagreed with this statement.

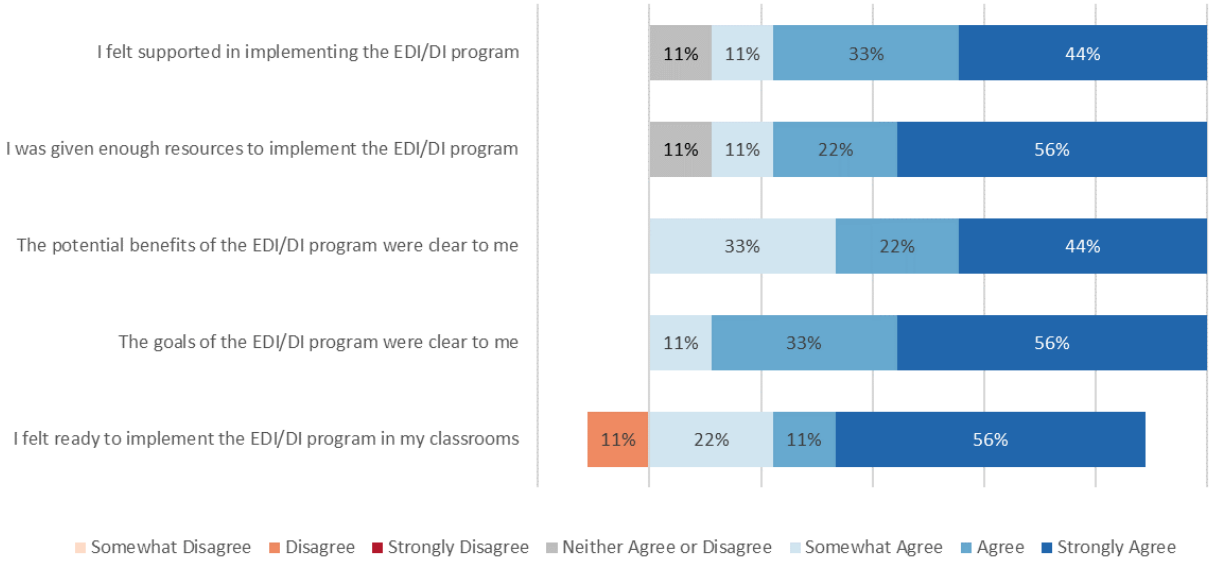


Figure 26. Teacher perceptions of implementing EDI/DI after training (n=9).

When asked how long it took to feel proficient at teaching literacy using EDI/DI, the response was varied ranging from one term to one and a half years. Years spent teaching using the EDI/DI approach ranged from one year to five years. Three of the ten respondents indicated that they taught other subjects using the EDI/DI approach. These subjects included maths, religious education, and integrated topics.

Overall summary of engagement with training and observations

- The number of observation sessions conducted had significantly decreased since the start of program implementation, the number of training sessions has remained relatively consistent
- Most survey respondents help positive views of their ability to implement the program after training

4.5.2. Leadership Turnover

School reporting tools record the current school staff, including leadership, involved in the implementation of the DI program. The number of principals identified as involved in the program can be used as a proxy measure for leadership turnover within the school. However, this measure is limited as principals are not necessarily involved in the implementation of the program.

The table below shows that nine schools indicated that two or three principals were involved with the FLFRPSP throughout its implementation at their respective schools. Six schools recorded one principal as engaged with the program during implementation; of these schools, three principals were recorded as engaged with 100 per cent of program implementation i.e. they were listed as involved in the implementation of the FLFRPSP in all of the schools SRTs. Accordingly, nine to twelve schools have experienced leadership turnover during the duration of the program.

Table 9. Number of Principals Recorded as Involved in the Implementation of the FLFRPSP

School	Number of Principals
31	1 (100%)
29	1 (100%)
39	1 (100%)
18	1
24	1
20	1
7	2
9	2
12	2
14	2
27	2
3	3
11	3
1	3
17	3

Overall summary of leadership turnover

- School reporting tools were used as a proxy measure for leadership turnover
- Based in school reporting tools, there is high leadership turnover across the program schools during the implementation of the FLFRPSP

4.5.3. School Fidelity

This section presents the school level data from the GGSA School Reporting Tools (SRT) for 2015, 2016, 2017 and 2018. Data for each of the 15 schools has been aggregated to provide an overall understanding of implementation, effectiveness and fidelity.

School Reporting Tools require school staff to record school fidelity measures. Overall school fidelity scores were calculated based on school's average scores on Teacher Readiness, Classroom Readiness, Resources Readiness, GGSA Readiness and Instructional Leadership. Fidelity scores for each domain were highest in 2016-2017 with only Teacher Readiness showing a gradual improvement from 2015 to 2018. The overall school fidelity reached a peak in 2017 before declining slightly at the end of the program in 2018.

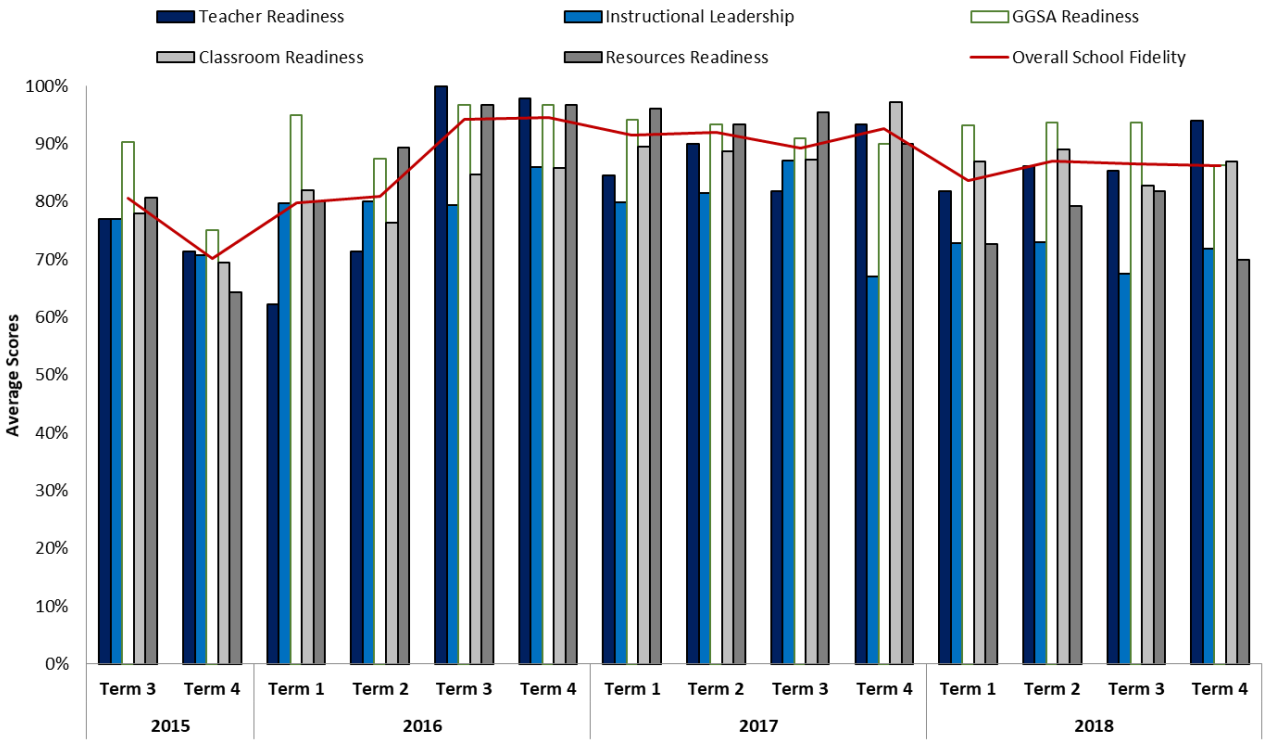


Figure 27. Average school fidelity from school reporting tool data.

Modifications to program implementation

The survey asked respondents if they had modified the EDI/DI program and half of survey respondents indicated that they had modified the program. When asked to expand on the ways they had modified the program, they referred to two key modifications that included reduced time for teaching literacy and the use of additional resources and activities not already used by the program.

A few of the survey respondents indicated they modified the EDI/DI program by reducing the number of hours they spent implementing the program each day:

> Having trouble to get 2.5 hours across the school, not enough staff, so we aim for one lesson per kid per day to get to students, rather than 2.5 hours. (School 9, Principal)

In addition to modifying time, a small number of survey respondents indicated that they provided additional activities and resources in literacy lessons, such as incorporating hand writing activities, providing resources students can manipulate or providing whiteboards:

> Implemented hands-on materials students can physically manipulate to assist in their learning and maintain focus. Presented other letter shapes alongside DI letter shapes. (School 18, Teacher 2)

Consistent with the survey, principals interviewed found the lack of interactive and hands-on activities a barrier to program implementation. Four principals commented on the lack of interactive and hands-on activities which limited the students' exposure to creative and imaginative learning opportunities. One principal expressed their view on the rigid structure of the program:

> […] because that's what they're wanting the kids to do, sit down and do all this rigid instructions, but early learning's all about play and discovery and oral language and phonics, and they need to be imaginative and creative and all those sorts of things. (School 18, Principal)

During interviews, principals were asked whether they had changed the program in any way. Most participants reported high fidelity, yet minor modifications to the program were made. Both schools that reported program adaptations and those that did not perceived the program had been beneficial to students or teachers. Most participants described aspects of the program amenable to modification or contextualisation, in particular, program content. One principal emphasised the importance of allowing flexibility to the program, for example in relation to the teaching strategies and methods that are used:

> I don't believe teachers should stick to the script, which I think should allow other things to come into it, if you know what I mean. (School 31, Principal)

Number of lessons per day

Teachers record the number of lessons they conduct for each program, each week. From this, the average number of lessons per day was calculated for all groups in the continuing schools and then grouped by year. There were several outliers in the lessons per day which were excluded.

The benchmark for number of lessons per day for implementation of the FLFRPSP programs is one. Across the lifetime of the program, none of the three literacy programs reached this benchmark. On average, language program lessons were consistently below 0.5 per day, reading program lessons were consistently below 0.6 per day and spelling program lessons were consistently below 0.7 per day.

Overall, spelling programs had the largest percentage of lessons completed, followed by reading and language. For each program the number of lessons per day has decrease slightly from the beginning of the program in 2015 to 2018.

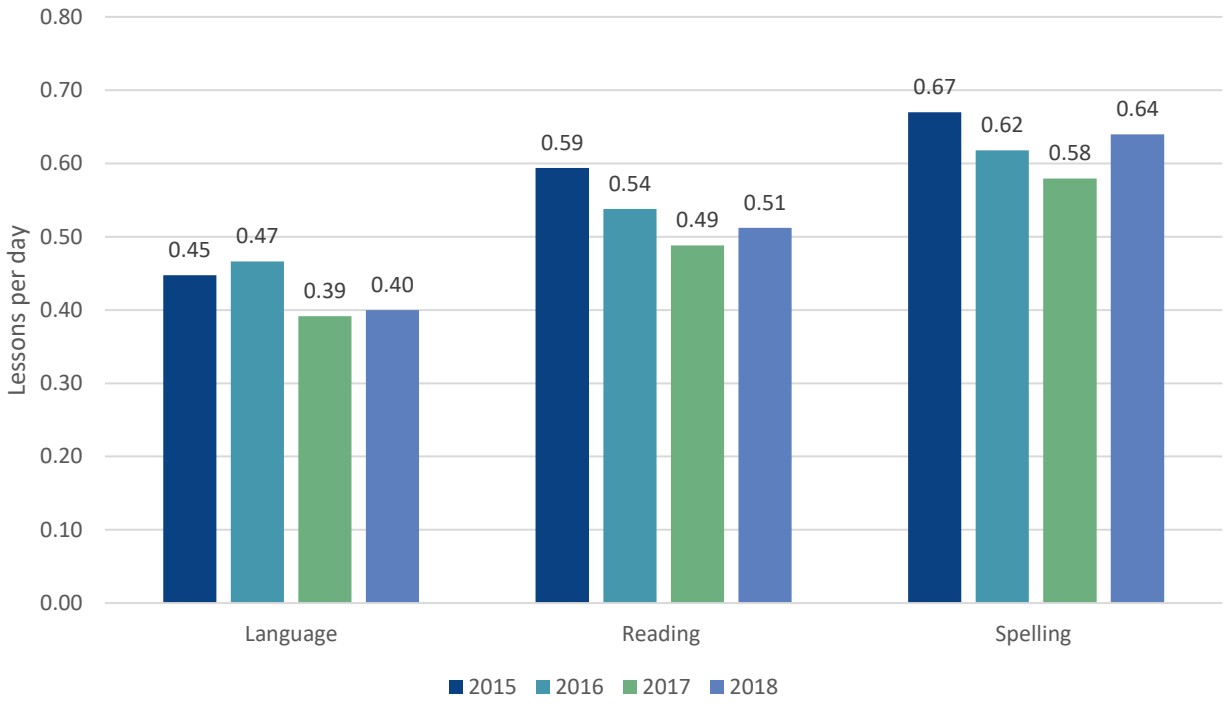


Figure 28. Average number of lessons per day

The overall average number of lessons per day for language, reading and spelling were split by school. Average lessons per day were not recorded consistently across years or the three programs, School 11 did not record any lessons for spelling. The figure overleaf shows that lessons per day were variable within and between schools though the duration of the FLFRPSP. In four years of program implementation, only school 24 had an average number of lessons per day which met the benchmark of one.

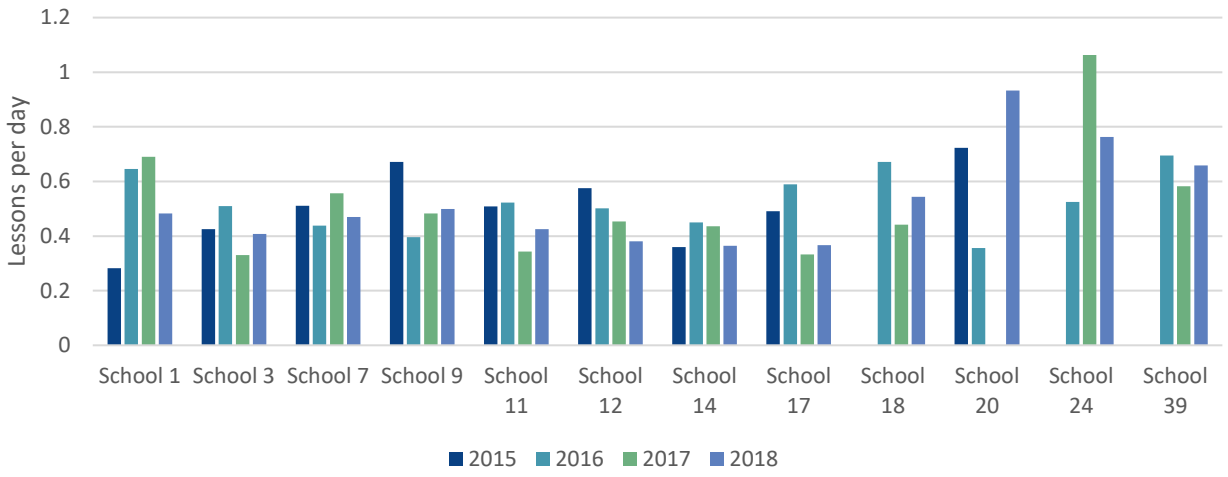


***Figure 29.**. Average lessons per day by school*

Time available for implementation

The survey asked respondents if they had enough time to cover all the EDI/DI content in the time available. Half of survey respondents indicated they did not have enough time to cover all the EDI/DI content in the time available. In the open-ended most of the respondents who indicated they did not have enough time to cover all the EDI/DI material, highlighted limited staff as the reason for this:

> Lack of teaching staff - we are 3 teachers down. (School 3, Teacher 1)

One survey respondent felt they were unable to cover all the content in the time available as a composite class structure meant there was a wide age and ability range in their class resulting in too many ability groups. A small number of survey respondents also commented on other contextual factors which impacted program implementation and the ability to cover all of the material in the time available, including student behaviour issues and school contextual factors that interrupted lessons such as having all school masses.

Overall summary of school fidelity

- Overall school fidelity scores peaked in 2016-2017 and have been declining since 2018
- Schools have been unable to meet the benchmark of one lesson per day across the duration of the program, this is consistent with the view of survey respondents that there is not enough time to cover all the EDI/DI content in the time avaliable
- Several survey and interview participants indicated they have modified the program to make it more suitable to their context

4.5.4. Facilitators and Barriers to Program Implementation

All staff survey respondents identified what they found easiest about implementing the program and what they found most difficult about implementing the program. Factors that facilitated program implementation included a reduction in time spent planning lessons, the provision of organised resources and leadership support. One survey respondent found nothing easy about the program. Barriers to implementation were more varied and included poor student attention, poor student attendance, staff satisfaction and turnover, and program management.

Resources to support implementation

Survey respondents identified a number of key enablers that facilitated program delivery, including a reduction in the time needed by teachers to plan, along with the provision of resources.

The most common theme to emerge in what survey respondents found easiest about implementing the program was a reduction in time spent planning lessons as the lesson plan and materials are provided as a part of the program. The majority of survey respondents commented on this reduction in time spent planning lessons:

> The program is provided. A teacher can pick it up and go with it. No planning. No prep required. (School 18, Teacher 2)

Similar to reduced planning time, a few respondents specifically commented on the provision of organised resources as a factor that made FLFRPSP easy to implement:

> When the books, manuals and paperwork is organised correctly, DI works well. (School 20, Teacher 1)

This was consistent with principal interviews, with two principals appreciating the external support given from the program, including external assessors to keep the teachers on track, and funding to purchase the resources. One principal explained the importance of the external support to implementing the program:

> The fact that it is supported by the Department, I think that would be very difficult to do I think, well if we didn't have the external support. The fact that it's being funded by the Department… like not have to buy the resources after helps, I don't think our school would be able to afford to do it. (School 12, Principal)

Some survey respondents found the provision of unorganised or unnecessary resources a factor that made the FLFRPSP difficult to implement:

> Middle management and the variation of unnecessary forms that cannot be altered. A total waste of paper. (School 20, Teacher 2)

This included one survey respondent who had identified organised resources as a factor that made the FLFRPSP easy to implement also identified it as a factor that made the program difficult to implement:

> The resources, books, teachers' manuals can be a nightmare to try to organise. (School 20, Teacher 1)

Program structure as a facilitator and barrier to implementation

A small number of survey respondents mentioned specific components of the program which facilitated implementation, such as the consistent approach of teaching, from planning to delivery:

> Consistent approach. Materials organised. Planning completed. (School 18, Teacher 1)

The analysis of interview participants' responses reveals a tension between benefits associated with program structure (consistency across teachers, whole school approach to literacy, positive impact on student behaviours) and limitations of such program structure. In particular, respondents described limited opportunities for students to develop teacher-student relationships, and for teachers to adapt the program, or indeed differentiate their teaching methods to cater for students with additional learning or sensorial needs. Whilst the structure of the program affords benefits for teachers and students, it can also limit classroom dynamics (and teacher-student interactions in particular) and student-content interactions (eg limited opportunities to engage in hands-on activities or develop writing skills). One of the principals expressed their frustration towards the rigid structure:

> Do I think it's right that kids don't start doing any reading until they've finished 40 lessons? I would say no. … This is a real constraint. (School 39, Principal)

Principal and teacher buy-in

Both leadership buy-in and teacher buy-in were identified by staff survey respondents as enablers of implementation, where the absence of buy-in is a barrier to implementation.

One respondent commented on the leadership support in their school as a factor that assisted in easy implementation of the program. Conversely, one teacher found nothing easy about the program to implement because of the challenge to get buy in. This was in contrast to the response from a teacher from the same school:

> Nothing, rather difficult to get everyone on board. (School 20, Teacher 2)

Flexibility for context

Staff survey respondents and interview participants identified a number of school contextual factors as barriers to implementation. This included student attendance and behaviour, and class composition. There was a perceived mismatch between the specific school context in which the program was implemented, and the America-focused content of the program resources, which was identified as a barrier for student engagement with the program.

Survey responses to what school staff found most difficult about implementing the program were varied. Half of the respondents reiterated contextual factors which they had previously identified which made implementation difficult and affected student literacy gain, including low student attendance, student behaviour and attention, and student overall satisfaction with the program:

> The transient nature of the students and the inconsistent attendance. (School 18, Teacher 3)

> Too long (3 hours). Students became disengaged when not with teacher. Not allowing for varied learning dispositions. (School 3, Teacher 2)

A few respondents commented on the composition of classes as the most difficult factor in implementing the program. With one respondent indicating that having too many students in one group made implementation difficult, and one respondent indicating that the age range in the class meant there were too many groups:

> The age range and having so many groups to get through. (School 1, Teacher)

A few respondents believed staff attitude and lack of commitment was the biggest challenge to implementing the FLFRPSP, and negatively impacted implementation of the program:

> Commitment from all staff. (School 9, Principal)

A recurrent theme in principal interviews related to the use of an American context in program resources, with five principals saying that it was limiting the benefit of the program because due to the low applicability of the materials to the school context, and particularly given students could not relate to the content of the resources. One principal illustrated this limitation in relation to their school context:

> Yes, the American context does not work in our remote Indigenous context. When you're talking about things like putting your slippers on and getting the pipe out and those sorts of things, my kids don't wear any shoes up here. (School 18, Principal)

Other limitations mentioned by at least two principals included not aiding the development of teacher-student relationship, no time for teaching other subjects, the very prescriptive structure which restricted ability to deliver to classes of varying ages, groups and abilities, and not user-friendly resources.

Factors influencing no literacy gain

When asked whether there were students in their school whose literacy did not seem to be improving, the majority of survey respondents said yes. Reasons for limited literacy improvement included student factors such as students with additional needs, low attendance and poor behaviour, as well as the program overall.

Low Attendance

Three respondents highlighted poor attendance as the reason for small literacy gain in some students:

> They are not at school often enough. (School 9, Principal)

Student additional emotional and behavioural needs

Four respondents mentioned additional needs for which students do not receive support and poor behaviour as the reasons for why some students' literacy was not improving:

> Attendance, behaviour, forced program. (School 20, Teacher 2)

Program choice

Two respondents believed that it was the specific program or pedagogy that impacted the literacy gain of students and meant some students have not improved:

> Not correct pedagogy. (School 3, Teacher 2)

Reasons for program withdrawal

The responses of interviewed principals from recently withdrawn schools were analysed separately from other principal interview data to avoid bias in the overall results. Both principals from recently withdrawn schools identified the reason for their school withdrawing from the program as limited student progress, in literacy overall and also in writing which is not covered by the program:

I think there was a lack of opportunity for the kids to participate in writing, or improve in their writing skills. (School 11, Principal)

Overall summary of perception of facilitators and barriers to implementation

- There has been less engagement with surveys and interviews compared to previous years, showing a waning interest in engagement with evaluation
- Overall there was a high level of variation in teacher and principal perception of the enablers and barriers to program implementation
- School and community context were considered important factors for implementation; several respondents noted the need for the adaptation of DI for a rural Australian context (eg changing American language, considerations for students with differing needs)

5. Discussion

The foundational aims of the FLSRPSP were to:

1. Improve students' literacy abilities and results, and
2. Increase teacher pedagogical skills in literacy development

The purpose of this final evaluation report is to provide a summative assessment of the impact of the FLFRPSP against these aims. Since the last report, additional NAPLAN, program, principal and school staff perception data have been collected. With the inclusion of this additional data, the following results of the program in relation to the aims have been observed, as described below.

5.1. Aim 1: Improving Students Literacy

Literacy skills have been assessed using all available data. Specifically, all jurisdictions provided NAPLAN data for the majority of participating schools. However, due to the decreased sample size compared to the previous report, ELYND data for WA Catholic Schools has not been included.

NAPLAN has been acknowledged as limited as it is a single outcome measure of literacy progress because of its measurement properties and its content. NAPLAN is inadequate in isolation to capture literacy progress because:

- Not all students in the program are represented at any one time – only those in Years 3 and 5 of the relevant evaluation years
- The high turnover of students in program schools mean that students individual progress is not possible to track in comparison years
- As mentioned, NAPLAN is a relevant but more distal measure of skills taught in the program and does not measure directly whether students have mastered component reading skills such as phonics or phonemic awareness which are more proximal to program experiences.

5.1.1. Literacy Progress

Noting the limitations of NAPLAN above, there was varied mean growth in all NAPLAN domains observed, however overall most program schools showed positive literacy growth. The greatest mean difference was observed in the NAPLAN domain Spelling. There were no statistically significant differences between program and control schools in any domain. Consistent with NAPLAN data, there was varied student mastery across language, reading and spelling.

5.1.2. Perception of Program Impact on Students

In addition to the literacy outcomes assessed through NAPLAN, staff perceptions of the impact of the FLFRPSP on student literacy outcomes were assessed through interviews with principals of participating, or recently withdrawn program schools and through staff surveys. In total nine principals were interviewed and ten staff surveys, representing five different schools, were completed.

Overall, principals held varied views of the progress, growth and achievement of students due to their FLFRPSP participation. Consistent with the limitations of NAPLAN as a measure of literacy progress, two principals felt unable to determine whether there was evidence of program impact on literacy using standardised testing measures. Three other principals reported varied and year level-specific improvements, however the remaining three principals who were interviewed, reported an overall decline of student performance because of the DI model.

In contrast, all survey respondents agreed that the FLFRPSP had helped improve student literacy, and when asked if the program had been successful overall, 80 per cent of respondents agreed. Survey respondents indicated that there were students in their class whose literacy skills were not improving; and poor attendance and poor behaviour (as factors that negatively impact the teacher's ability to deliver the content of the program) were considered the main reasons for the lack of improvement in these students. Aim 2: Increase Teacher Pedagogical Skills

The implementation of any new pedagogical approach requires a shift in classroom teacher practice. In the FLFRPSP, indicators of change in teacher skills and practice included measures of teacher effectiveness, derived from the program itself were examined, and these were organised into three indices: Classroom Organisation, Instructional Delivery and Behaviour Management and Engagement.

5.1.3. Program Data

School reporting tools show that overall, there has been varied quality program implementation across the life of the FLFRPSP. High implementation was observed in the 2016-2017 period, with the largest number of observations conducted and high fidelity and teacher implementation effectiveness scores observed. However, since 2016, there has been a significant decrease in the number of observations conducted since 2016, and further, there has been a decrease in teacher implementation effectiveness and fidelity both in instructional leadership and overall fidelity. Across the life of the program almost all schools have not met the benchmark of one lesson per day.

5.1.4. Perception of Program Impact on Teacher Pedagogy

In addition to program data, the impact of the FLFRPSP on teacher pedagogical skills was assessed through interviews with principals of participating or recently withdrawn schools, and finally, staff surveys. In total, eight principals were interviewed and ten staff surveys, representing five different schools, were completed.

Overall, principals held varying views of teachers' professional growth as a result of the FLFRPSP. Principals indicated that there was limited engagement with the program from teachers during initial implementation. For some schools, despite this, teacher attitudes towards the program shifted, inspiring more teacher accountability. However, in contrast, other principals perceived the scripted nature of DI as reducing accountability and preventing teachers from implementing their preferred pedagogies, negatively affecting teacher attitudes.

Consistent with program data, principals indicated there had been a reduction in participation in training, which may have negative impacts upon support and resources for teachers, along with their understanding of the goals of program and readiness to implement the FLFRPSP.

Overall, the key findings of the evaluation in relation to program aims provides a picture of high variance in implementation and outcomes of the FLFRPSP both on students and teachers, however it is evident that there has been a reduction in implementation (degree, fidelity and effectiveness) since 2016. The following section provides a discussion to explain why these outcomes have occurred, drawing on previous reports and relevant literature where appropriate.

5.2. Relationship Between Implementation and Mediating Factors

A similar story to the previous reports (Clinton, Dawson, McLaren, & Koelle, 2017; Dawson, Clinton, Koelle, & McLaren, 2018) has emerged from the 2018 data collection. There remain clusters of success in literacy changes across rural and remote primary schools as a consequence of the FLFRPSP, as well as gains in spelling and writing.

Further, there is principal and teacher perception of change and impact across several 'promising' schools. However, there are several schools that have withdrawn from the program who held a critical view of the program, particularly from a principal's perspective:

> No, we felt the kids were going backwards, if anything. (School 19, Principal)

5.2.1. Teacher Pedagogical Preference and Program Theory of Change

These views are in part, complicated by the nature of teaching along with historical reforms. Irrespective of the reform or program, pedagogical preference remains a dividing facet, the classroom is a domain where the teacher is an autonomous professional and hence, pedagogical preference can be considered an individual choice. This nature of teaching is bought to the fore in the FLFRPSP because programs with DI have a fixed structure.

5.2.2. School Structure and Leadership to Support Sustained Implementation

Further, the FLFRPSP implementation and experiences of all those who participated, and particularly among schools with lower levels of implementation and change in literacy outcomes for students, highlight the importance of structures-in school, support of school leadership to sustainably implement the FLFRPSP. School principal leadership that supports the program and establishes infrastructure to support program implementation has also been found to be essential in every year of the evaluation- it is a key factor to program success.

5.2.3. Relationship Between Quality Implementation and Program Effectiveness

The findings from the 2018 data collection also clearly highlight the *direct relationship* between implementation and program effectiveness. We describe implementation as those constructs that relate to the degree of implementation, the fidelity of implementation, the level of adaption and the overall quality of delivery. As we have suggested it is

difficult to provide exact and systematic measures of all these components due to the challenge in capturing and monitoring fidelity and adherence in the classroom environment. Davis (2018) suggests that assessing effectiveness is not feasible in these complex contexts where there are measurement challenges, therefore it may be inaccurate to conclude on effectiveness. However, this evaluation has considered these complex environments by triangulating and synthesising a series of data sources including program data with interviews and survey results in order to provide some insight on program implementation and effectiveness.

The degree and fidelity of the program has been variable throughout the three years of the program and more so in 2018. There is some suggestion that adaptation of the program is quite high in some areas, for example, teachers making their own resources to fit context. It also seems that often the elements of the program are handed to support staff within schools. It is important to note that this may be the case in schools where there has been change in literacy levels, which teachers perceive relates to program. Considering that across the evaluation years there has been growth in some areas despite high levels of differential implementation, the questions arises as to what impact the program could have with great levels of quality implementation.

5.2.4. Implementation Challenges in Rural and Remote School Contexts: Supporting Structural Integrity

The contexts in which the program is being implemented have been described as difficult. Teacher and principal turnover are high and therefore the need to obtain buy-in continues throughout implementation, and there is a continual need to build competency and support the development of positive attitudes towards the program through building knowledge and awareness. Traditionally in most social programs, this process is fixed or predominantly occurs early in the implementation of an intervention model, but with such high turnover and other contextual factors the FLFRPSP has continually faced a need to continue to gain buy-in and raise knowledge and awareness. Interestingly one of the recommendations from the Independent Review into Regional, Rural and Remote Education was to ensure challenges, and opportunities, of working in rural and remote schools and contexts are explicitly included in the selection processes for teacher education and initial appointment processes (Halsey, 2018).

When context is considered, the structural integrity of the school to implement must be estimated. Structural integrity refers to capacity of the infrastructure to carry the load over time (Clinton, 2018). For a program to be scalable, it must be able to withstand the fluctuations of the complex dynamic system that is education. This includes social, cultural, economic and policy change. In this case, remote schools face a challenging context with many factors that are often beyond their control. Any intervention implemented in this context must include processes, and ideally structures to enable implementation to occur with high levels of quality and be sustained. The data gathered in the evaluation demonstrates that, at commencement, the FLFRPSP design had high levels of integrity; however, this has become variable according to many leaders with in school.

While the context of schools participating in the program is challenging, so too is the workforce stability and the variable structures of the education system that schools and the provider of the FLFRPSP face. Changing policies, support and access to data across the three states and territory has been variable and is considered a barrier to implementation. This would be true for any multi-state educational program.

5.2.5. Evaluation Evidence

Further, and perhaps consequently, evaluation data has not been readily available from the systems, schools, or the provider. For example, targeted literacy data to determine literacy change has not been made available, which has made it difficult to determine the influence of the implementation of FLFRPSP on individual students over the three-year period. This lack of forthcoming data from systems, schools or the provider is an indicator of lack of engagement in the evaluation process.

The mediating factors highlighted explain the implementation process, and outcomes observed in the evaluation. They also point to, a need to improve the structural integrity of the FLFRPSP to be *implementable* within the specific remote context. It should also be noted, that there is also a need to support remote schools to improve their own structural integrity for the purposes of implementing the FLFRPSP, which could also support other reforms and interventions occurring in the school. The diagram overleaf highlights the relationship between the mediating factors and the FLFRPSP.

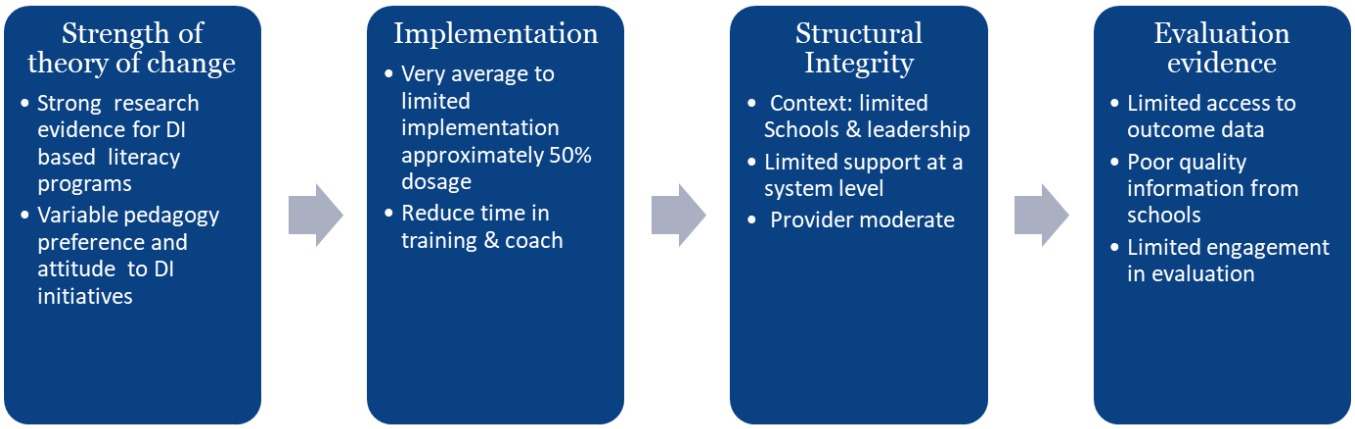


Figure 30. Understanding mediators and outcomes.

While there have been a number of factors that negatively mediated the implementation of the program, the evaluation findings point to some cases of success. The profiles of schools in which change has occurred, and in which the school and community have embraced the program, illustrate the need for the program. It is significant that some schools continue to have tremendous success with the program. These schools in partnership with the provider have been able to overcome the barriers to the program implementation in their complex contexts, with their implementation process contributing to the structural integrity of the program in their schools. The diagram overleaf illustrates the criteria derived from these profiles that support sustained implementation and effect.

As evident in the figure below, success in the implementation of the FLFRPSP depends upon a partnership between the school, system, community, and the providers of the FLFRPSP.

Figure 31. Success model.

In attempting to model the progress over time or the diffusion of the FLFRPSP to all participating remote primary schools, a life course approach was used to illustrate progress, model scalability, and predict sustained change. The mediating factors discussed in this section have been applied in this model to understand the principles of stability related to this program.

The 2017 Life Course Model of the FLFRPSP (Clinton, Dawson, McLaren, & Koelle, 2017) indicated that clusters of impact were apparent, particularly for Catholic Education WA schools. One year on, the 2018 report (Dawson, Clinton, Koelle, & McLaren, 2018) illustrated a diffused impact in Catholic Education WA schools, where the program had a greater and

more sustained impact than other jurisdictions. The 2018 report indicated a steady growth in NAPLAN results and differences between control and intervention groups in Writing and Spelling. It was therefore predicted that, based on those trends, the schools participating in the program in some areas were set to progress at a greater rate than control schools. However, this effect has not been observed in the data analysed.

Of course, it is not possible to make detailed or confirmatory conclusions about the impact of FLFRPSP on student literacy based on the NAPLAN data that was available for this report. Based on NAPLAN data it appears that, consistent with the 2018 report, the program is having large effects on Spelling, a large effect on Writing, a medium effect on Reading and a small effect on Grammar and Punctuation. However, despite these effect sizes, in contrast to previous years, control schools have recorded greater mean NAPLAN scores compared to schools participating in the program.

The growth of participating schools predicted in the 2017 report (Clinton, Dawson, McLaren, & Koelle, 2017) assumed that drivers such as training, and teacher mindset would influence teacher buy-in and attitudes, and once such buy-in and positive attitudes had been developed this would contribute to a positive growth trajectory of program implementation and impact.

The 2018 report (Dawson, Clinton, Koelle, & McLaren, 2018), which documented a diffused impact in Catholic Education WA schools where the program had had a greater and more sustained impact than other jurisdictions, revised the predicted trajectory of program implementation. The absence of drivers, such as systematic implementation, were found to negatively impact the progression of the program in all jurisdictions, resulting in a revised trajectory of program implementation.

This current report has, through further data collection, presented a richer and more sophisticated understanding of the contextual factors that impact the progression of the FLFRPSP. For example, the degree of teacher turnover causing new teachers being introduced to the program every year, effectively looping program implementation back to the beginning of the life course model every year, impacts the progression of the program.

The figure below documents the program progression trajectories through the life course model across each year of the evaluation and demonstrates how it has been revised as further evidence was gathered.

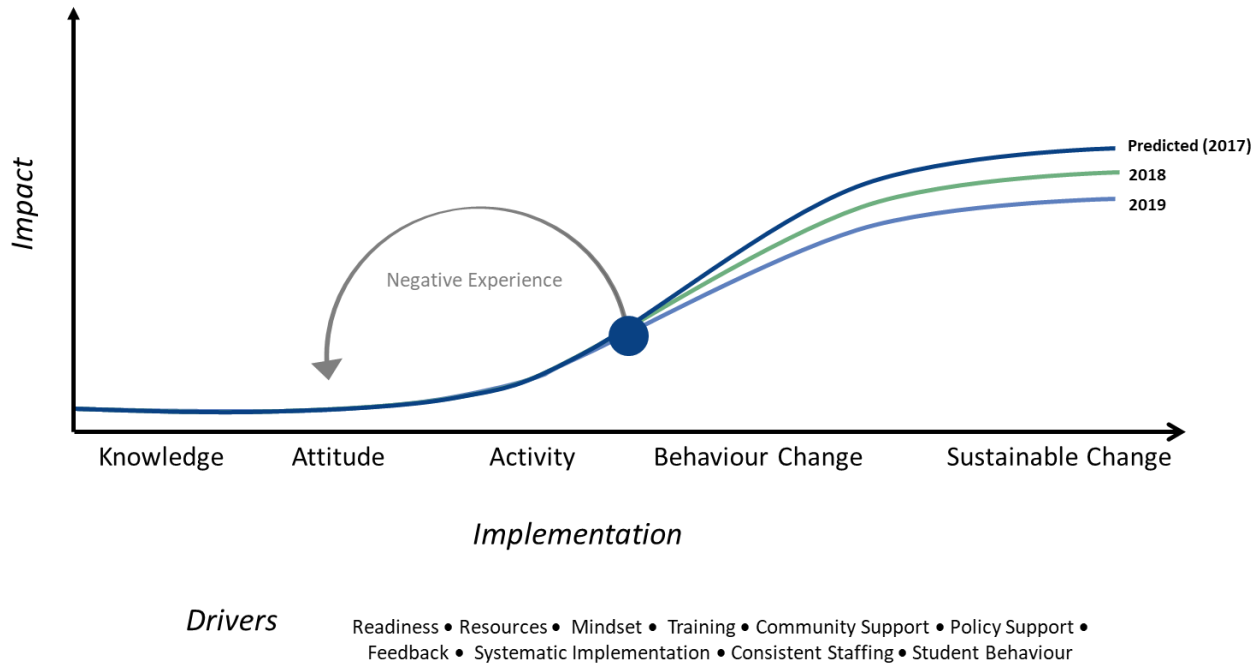


Figure 32. FLFRPSP Implementation model.

Despite great successes in some schools, the high variability in implementation means the FLFRPSP does not appear to be ready for scale. Considering that the constant turnover in the school system and the changing policy of the complex education system are key drivers, it means that there is a constant requirement to train and convince teachers, leaders and systems of the worth of the program. Therefore, participating schools can be stuck in the early cycle of the diffusion of an innovation. Where there is more workforce stability, the program appears to work. This is

consistent with the findings of Bartanen et al. (2019), that schools who change principals have lower student achievement in maths and reading, and a high teacher turnover than schools who keep the same principal. As illustrated by the figure above, while program schools showed initial engagement with the program, without continued support and workforce stability implementation, progress and growth is halted.

Overall, this program can work well, and it could be a part of a cache of initiatives or programs that are accessible to schools. In this case, it should be understood that the program is fit-for-purpose, and it could be effectively implemented if teachers, leaders and community are aware, trained, and engaged with the idea of a direct instructional approach to supporting literacy development. It would be remiss not to note that the FLFRPSP could bring much needed change to the literacy development of children in remote primary schools. It is a significant resource for schools that can add value, and it seems ill-advised for schools to not have access to it. However, the current system of implementation does not appear to go far enough to suggest scale is feasible.

5.3. Recommendations

Based on the evaluation of the FLFRPSP, there are a number of recommendations, including the need to allow for program adaptations and flexibility, adaptations to training to mitigate teacher turnover, and providing system level support for the implementation of the program.

Adaptations to Program to Support Scale

There is a wide variation of context and need and as such the program should consider how it's implementing the program according these variations. In order to support scaling of the FLFRPSP, adapting the program accordingly is crucial. Consider the following areas for adaptations:

- To better meet the needs of schools, teachers and students, the program should incorporate adaptations and flexibility in its structure and delivery, for example, allowing flexibility in the time prescribed for implementation of literacy lessons (ie less than 2.5 hours of literacy classes)
- Adapting the program to align with and promote the use of Australian resources
- Provide more flexibility for teachers and itinerant students
- Consider diversity and inclusion of students from different backgrounds and cultures
- Adapt the program for vulnerable students with complex and learning needs

Adaptations to Training

Considering the impact of teacher and leadership turnover on the progress and impact of the program, a number of training and teacher support recommendations are put forward that include:

- Greater flexibility in the training of teachers (eg online, micro credentialing) would mitigate factors that prevent them from attending training such as distance from training and contextual factors such as the wet season
- Training refreshers for schools are important given the high rate of teacher and leadership turnover, to ensure the constant change in staff does not prevent progress of implementation

System Support

Wider system resourcing and structural support are important for progression of implementation for schools that have buy-in as well as schools that have yet to be engaged. Consider the following support:

- Before implementation, consider providing the program to schools that have buy-in. This would help mitigate negative attitudes that effect progression, for example teachers who have a preferred pedagogy.
- Leverage on established network of schools (as demonstrated in 2017) that are engaged in the FLFRPSP as a way to facilitate for schools to support each other.
- The FLFRPSP should be considered a part of a literacy tool kit that helps support the literacy gain of rural and remote students.

6. References

Aarons, G.A., Ehrhart, M.G., Farahnak, L.R., & Sklar, M. (2014). Aligning Leadership Across Systems and Organizations to Develop a Strategic Climate for Evidence-Based Practice Implementation. Annual Review of Public Health, 35: 255-274.

ACARA (Australian Curriculum Assessment and Reporting Authority). (2016). About. Last Accessed: 18.05.2019. Retrieved From: https://www.nap.edu.au/about

ACARA (Australian Curriculum Assessment and Reporting Authority). (2017). ICSEA 2017: Technical Report. Last Accessed: 18.05.2019. Retrieved From: https://www.myschool.edu.au/media/1635/icsea_2017_technical_report.pdf

ACARA (Australian Curriculum Assessment and Reporting Authority). (2015). What Does the ICSEA Value Mean? Last Accessed: 18.05.2019. Retrieved From: http://docs.acara.edu.au/resources/About_icsea_2014.pdf

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

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

Australian Government Department of Education. (2018). *Flexible Literacy for Remote Primary Schools*. Last Accessed: 26.05.2018. Retrieved From: https://www.education.gov.au/flexible-literacy-remote-primary-schools-1

Australian Government Department of Education and Training [DET]. (2017). Non-government Reform Support Fund Guidelines. Retrieved From: https://docs.education.gov.au/documents/non-government-reform-support-fund-guidelines

Bartanen, B., Grissom, J.A., & Rogers, L.K. (2019). Impacts of Principal Turnover. Educational Evaluation and Policy Analysis, 41(3), 350-374.

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

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

Bereiter, C., & Engelmann, S. (1966). Teaching disadvantaged children in the preschool. Englewood Cliffs, New Jersey: Prentice-Hall.

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

Bodkin-Andrews, G. & Carlson, B. (2016) The legacy of racism and Indigenous Australian identity within education. Race Ethnicity and Education, 19(4). 784-807. DOI: 10.1080/13613324.2014.969224

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

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

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

Centres for Disease Control and Prevention (CDC). (1999). Framework for Program Evaluation in Public Health. Atlanta, Georgia: Author.

Clarke, S., Wildy, H. (2011). Improving the small rural or remote school: The role of the district. Australian Journal of Education, 55(1). 24-36.

Clinton, J., Au, L.K., Mclaren, P., Dawson, G & Arifin, D. (2016). Evaluation of the Flexible Literacy for Remote Schools Programme: Formative Report 1 – Part A. May 2016. Centre for Program Evaluation, The University of Melbourne

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

Dataworks. (2019). *Explicit Direct Instruction (EDI)*. Last Accessed: 27.05.2019. Retrieved From: https://dataworks-ed.com/research-edi/

Davis, A. (2018). Evidence-based approaches to education: Direct instruction, anyone? Management in Education, 32(3). 135-138.

Dawson, G. K., Clinton, J., Koelle, M., & McLaren, P. (2018). Evaluation of the Flexible Literacy for Remote Primary Schools Program: Main Report. Centre for Program Evaluation, University of Melbourne: Author.

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

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

Denton, C. A., Vaughn, S., & Fletcher, J. M. (2003). Bringing Research-Based Practice in Reading Intervention to Scale. Learning Disabilities Research & Practice (Wiley-Blackwell), 18(3), 201–211.

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

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

Ewing, B. F. (2011). Direct instruction in mathematics: Issues for schools with high indigenous enrolments: A literature review. Australian Journal of Teacher Education, 36(5), 63-91.

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

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

Godinho, S., Woolley, M., Scholes, M., Sutton, G. (2017). Literacies for remote schools: Looking beyond a one size fits all approach. Literacy Learning: the Middle Years, 25(1). 28-40.

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

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

Hattie, J. (2008). Visible Learning: A synthesis of over 800 meta-analysis relating to achievement. Abingdon, London: Routledge.

Hollingsworth, J. R., & Ybarra, S. E. (2017). Explicit direct instruction (EDI): The power of the well-crafted, well-taught lesson. Thousand Oaks, California: Corwin Press.

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

Joyce, B. R., & Showers, B. (2002). Student achievement through staff development. Association for Supervision and Curriculum Development.

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

Klimes-Dougan, B., August, G. J., Lee, C.-Y. S., Realmuto, G. M., Bloomquist, M. L., Horowitz, J. L., & Eisenberg, T. L. (2009). Practitioner and site characteristics that relate to fidelity of implementation: the early risers prevention program in a going-to-scale intervention trial. Professional Psychology, Research and Practice, (5), 467.

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

Lock, G., Budgen, F., Lunay, R., Oakley, G. (2012). Welcome to the Outback: The paradoxes of living and teaching in remote Western Australian schools. Australian and International Journal of Rural Education, 22(3). 117-134.

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

Luke, A. (2014). On explicit and direct instruction. Australian Literacy Association Hot Topics, 1-4.

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

Lyon, A.R. (2018). Implementation Science and Practice in the Education Sector. Retrieved From: https://education.uw.edu/sites/default/files/Implementation%20Science%20Issue%20Brief%20072617.pdf

Maher, E. J., Jackson, L.J., Pecora, P.J., Schultz, D.J., Chandra, A., & Barnes-Proby, D.S. (2009). Overcoming challenges to implementing and evaluating evidence-based interventions in child welfare: A matter of necessity. Children and Youth Services Review 31(5): 555-562.

McRae, D., G. Ainsworth, J. Cumming, P. Hughs, Y. Mackay, K. Price, M. Rowland, J. Warhurst, D. Woods, and V. Zbar. (2000). What Works? Explorations in Improving Outcomes for Indigenous Students. Canberra: National Curriculum Services and the Australian Curriculum Studies Association

Ministers for the Department of Education and Training. (2018). Helping improve Indigenous literacy rates [Press Release]. Retrieved from https://ministers.education.gov.au/tehan/helping-improve-indigenous-literacy-rates

Moir, T. (2018). Why is Implementation Science Important for Intervention Design and Evaluation within Educational Settings. Frontiers in Education, 3:61.

National Institute for Direct Instruction [NIFDI]. (2011). A bibliography of the direct instruction curriculum and studies examining its efficacy. Eugene, Oregon: Author.

National Institute for Direct Instruction [NIFDI]. (2015). Basic Philosophy of Direct Instruction (DI). Retrieved From: https://www.nifdi.org/what-is-di/basic-philosophy.html

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

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

Sharplin, E. (2002). Rural retreat or outback hell: Expectations of rural and remote teaching. Issues in Educational Research, 12(1). 49-63.

Shippen, M. E., Houchins, D. E., Steventon, C., & Sartor, D. (2005). A Comparison of Two Direct Instruction Reading Programs for Urban Middle School Students. Remedial and Special Education, 26(3), 175–182. https://doi.org/10.1177/07419325050260030501

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

Stockard, J., Wood, T.W., Coughlin, C., & Khoury, C.R. (2018). The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research. Review of Educational Research, 88(4), 479-507.

Stufflebeam, D. L. (2000). The CIPP Model for Evaluation In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), Evaluation Models: Viewpoints on Educational and Human Services Evaluation (2 ed.). Dordrecht, Netherlands: Springer.

Stufflebeam, D. L., & Coryn, C. L. (2014). Evaluation theory, models, and applications (Vol. 50). San Francisco, CA: Jossey-Bass.

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

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

Appendices

Appendix 1: Semi-structure interview protocol

Professor Janet Clinton
Centre for Program Evaluation
Melbourne Graduate School of Education, The University of Melbourne
Level 4, 100 Leicester Street, Victoria 3010 Australia
T: +61 3 9035 3697 **E:** jclinton@unimelb.edu.au

Appendix 1: Semi-structures interview protocol

Interviews with principals will be semi-structured, and focus primarily on the implementation of the program in their school, expectations and outcomes (particularly student and teacher), barriers and enablers, limitations and positive achievements, and attitudes towards the program.

The following are example items that will be used to guide the interviews:

Opening questions:

- How does FLRPS fit into your school's literacy curriculum/strategy?
- What was the justification for using FLRPS
- Do you participate in analysis and discussion of FLRPS data?
- Do you feel as though you outlined a clear goal that would be achieved by introducing FLRPS to the school?
- What is your role in relation to FLRPS?
- Have you attended any training on FLRPS?
- Who is responsible for overseeing the implementation of FLRPS in your school?
- How many staff are responsible for delivering FLRPS sessions in your school (If more than 1 are students taught by different teachers)?

Perceptions of the program

- What do you think about the structure of the programme?
    - *is there anything that does not work in your context or that you have changed or want to change?*
    - *Is there component that is more difficult to implement?*
- How engaged are you/the school with the program?
- What do you think about the FLRPS resources? (GGSA Training and Support)
    - *is there anything that does not work in your context or that you have changed or want to change?)*
- Are there any aspects of the programme that you have found particularly useful?
    - *is there any component of the program or anything about the way it is taught that you really value*
- What do the **pupils** think about FLRPS?
- What do **staff** think about FLRPS?
- What factors do you think are most critical to the success of FLRPS?
- What factors do you think are most to limit the success of FLRPS?

Perceptions of student outcomes

- What do the pupils think about FLRPS?
    - *is there any component of the program or anything about the way it is taught that they really value/engage with*
- Has FLRPS made a difference for your pupils (All? Some particularly?- examples)
    - *Has student literacy improved because of the program?*
    - *Are there any students for whom FLRPS is inappropriate or whom do not seem to improve from FLFRPS*
- In your opinion how much (if any) of the change observed in student outcomes to date has occurred as a result of FLRPS?

Perceptions of teacher outcomes

- Do you think the program has improved teacher pedagogy in your school?
    - *If so, how?*
    - *If not, why?*
- What has been most helpful to your teachers?

Program implementation (classroom level)

- Have you/your teachers changed the order of tasks around at all? / Is the sequence of learning tasks appropriate?
    - *Are any of the tasks inappropriate for your students/context*
- Do you feel you/your teachers have enough non-teaching time for record-keeping and lesson preparation
    - *How long do you spend in completing work related to FLRPS*
- What is your opinion of FLRPS program model?
- What factors do you think are most critical to the success of FLRPS? *(Focus on classroom level)*
- What factors do you think are most to limit the success of FLRPS? *(Focus on classroom level)*

Perceptions of program implementation (school level)

- Do you believe your school is better able to collect data to understand the impact of particular teaching strategies?
- Please list any factors that limited the success of FLRPS in yours school
- Please list any factors that were necessary to support the successful implementation of FLRPS in the school
- Do school leaders participate in teacher learning and development?
- Did the school leaders clearly articulate how FLRPS complimented other school programs or objectives?
- Do you feel the pace of implementation was appropriate?
- Were the outcomes of the FLRPS project were in line with expectations?
- How have you supported staff in implementing the FLRPS Program?
- Does your school need any additional help/coaching in implementing the FLRPS program.

Other

- Is there anything else you would like to add about the FLRPS program that you do not think has been covered already?