EVIDENCE FOR LEARNING

Thinking Maths

Learning Impact Fund Evaluation Report

A professional learning program supporting teachers to engage middle-school students in maths

Evaluation Report and Executive Summary

September 2018

ACER

Independent Evaluators:

Dr Katherine Dix, Dr Hilary Hollingsworth, Mr Toby Carslake

About the evaluator

The project was independently evaluated by a team from the Australian Council for Educational Research (ACER): Hilary Hollingsworth, Katherine Dix and Toby Carslake.

ACER is one of the world's leading educational research centres. As an international, independent, non-profit, non-government organisation, ACER generates its entire income through contracted research and development projects, and through developing and distributing products and services, with operating surplus directed back into research and development.

Contact details:

Hilary Hollingsworth
The Australian Council for Educational Research Ltd (ACER)
19 Prospect Hill Road, Camberwell, Victoria, 3124, Australia
P: +61 3 9277 5555
E: hilary.hollingsworth@acer.org

Acknowledgements

This evaluation was commissioned and funded by SVA Evidence for Learning. The Australian Council for Educational Research (ACER) wishes to thank Evidence for Learning for their ongoing support throughout the evaluation. In particular, we would like to acknowledge John Bush, Pauline Ho and Matthew Deeble for their steady guidance in this flagship project.

We would also like to recognise the pioneering work of Pauline Carter and Maureen Hegarty, along with the Thinking Maths team at the South Australian Department for Education. From the very beginning of this project, they have been incredibly helpful, open and supportive – a partnership that has contributed significantly to the success of the evaluation.

ACER would also like to extend their gratitude and thanks to the many school communities involved in the research. We received sustained cooperation and support from principals, teachers, and students in the schools who participated in the trial. We thank all of these communities for their efforts, without which the evaluation could not have proceeded.

Executive Summary

The project

Thinking Maths is a three-term structured professional learning program for Years 6-9 mathematics teachers to engage middle school students' mathematics learning. The Thinking Maths program has been developed by the South Australian Department for Education (the Department), based on its Teaching for Effective Learning (TfEL) Framework. The program aims to address a significant drop in students' mathematics performance in NAPLAN from Years 7 to Year 9.

Thinking Maths supports teachers to improve students' learning of mathematical content as outlined in the Australian Curriculum Mathematics during the transition between Primary and Secondary school[1] (currently Year 7 and Year 8 in South Australia). It focuses on three areas for better teaching and learning of mathematics: (a) using quality task design, (b) sequencing of conceptual development, and (c) using research-informed effective pedagogies. Years 6-9 teachers participate in five professional learning days at 4-5 week intervals over three school terms in an eight-month intervention period delivered and led by two facilitators from the Department. Teachers are expected to make a commitment to implement the strategies they learn after each professional learning day back in their classrooms to improve student engagement and achievement.

The evaluation of Thinking Maths was independently conducted by the Australian Council for Educational Research (ACER) during February to October 2017. It involved over 7068 students (Years 5-10)[2] in 158 government Primary and Secondary schools across South Australia. This efficacy evaluation[3] was a multi-site, two-armed (intervention and business-as-usual control) Randomised Control Trial (RCT), with randomisation at the school level. The primary research question was to identify the impact of the Thinking Maths program on the mathematics achievement of individual students. The trial evaluated student achievement by using data from the standardised ACER Progressive Achievement Tests in Mathematics (PATMaths), routinely collected by the Department in September each year (pre-test in 2016 and post-test in 2017).

---

[1] The terms 'primary' and 'secondary' will be capitalised when referring to schools or Year levels, and lowercase when referring to outcomes, in order to avoid confusion.

[2] A number of composite classes included students in Years 5 and 10. These students comprised 8% of the sample and were not excluded from the study in accordance with intent-to-treat protocol.

[3] An evaluation that tests if a program works under optimal conditions.

Key conclusions

1. The Thinking Maths program had a small positive effect, equivalent to one month of additional learning progress on Years 5-10 students' performance in the PATMaths achievement test, when compared to business-as-usual mathematics classes. These findings were not statistically significant[4].

2. Thinking Maths had a statistically significant impact equivalent to two months learning gain in Primary students' achievement on the PATMaths test. However, for Secondary students, there were two fewer months of learning progress.

3. Among a sub-sample of School Card[5] holders, the students (both Primary and Secondary) of Thinking Maths teachers had two additional months' progress in performance on the PATMaths test, however this finding was not statistically significant.

4. Thinking Maths had the largest statistically significant effect on mathematics teachers' pedagogical and content knowledge, as well as their professional identity and self-efficacy. The intervention also showed a small positive impact on teaching practices overall, with students reporting that Thinking Maths teachers were more likely to give extra help when needed, ask questions to check understanding and challenge their thinking. Findings showed similar gains on students' cognitive engagement, but no additional gains in metacognition. These results on student outcomes were not statistically significant. A small and statistically significant increase in students' mathematics anxiety was also found.

5. Teachers reported a number of benefits of this professional learning program including hands-on activities, expert modelling of metacognition strategies and teaching resources that supported teachers to directly transfer ideas to their classrooms. The process evaluation indicated that timetabled lessons, common tests, set text-books, and lack of time to plan were barriers to effective implementation in Secondary schools. Schools and program development should consider differences in learning contexts to better accommodate and support teachers to optimise implementation.

[4] Evidence for Learning will develop a plain English commentary on statistical significance to support readers in interpreting statistical results in our reports.

[5] The School Card scheme offers financial assistance to low-income families to assist with school fees for students attending government schools in South Australia.

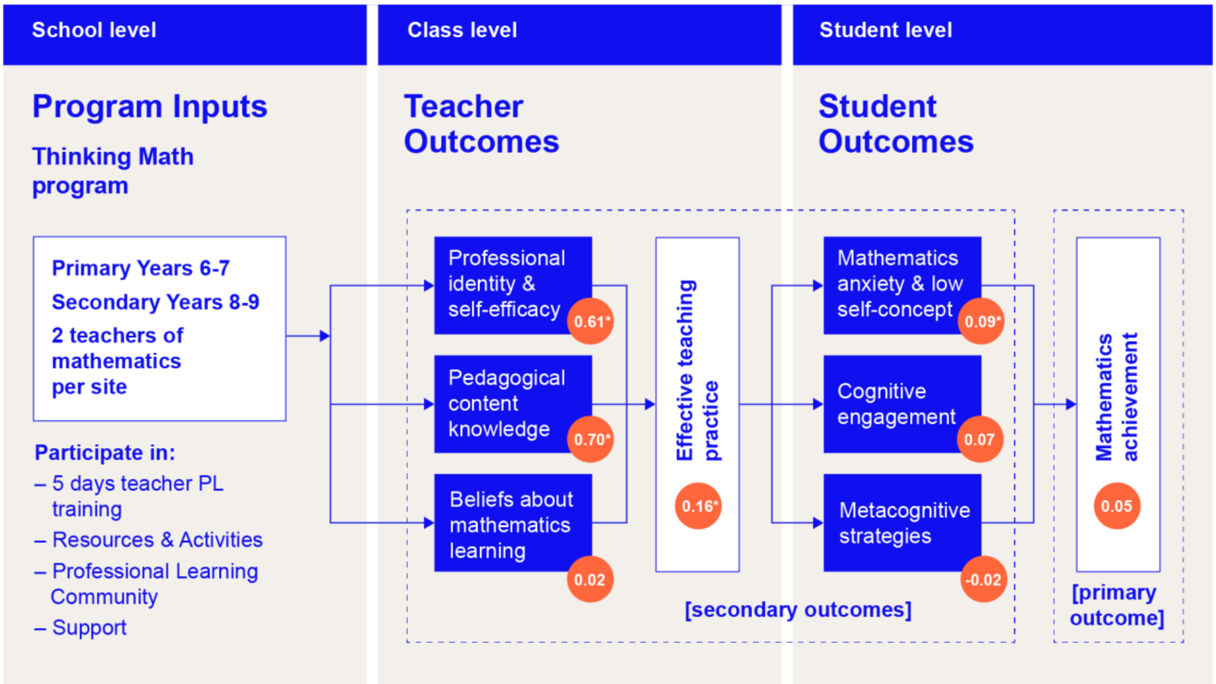
Main findings and impact

The evaluation found evidence of a small positive effect of the intervention overall. Students whose teachers attended the Thinking Maths program made more progress in mathematics than similar students in business-as-usual classrooms. The small positive effect is equivalent to one month of additional learning progress. However, this effect was not statistically significant. Across this cohort, there was also a small positive effect on students' cognitive engagement and no effect on metacognitive strategies, which were not statistically significant. Students also showed a small and statistically significant increase in their mathematics anxiety.

There is stronger confidence about the differences between Primary and Secondary students' achievement. Primary students (Years 5-7) of Thinking Maths teachers made a learning gain of an additional two months, however for Secondary students (Years 8-10), there were two fewer months of learning progress.

The program had a large positive impact on how teachers perceived their pedagogical content knowledge, particularly at the Primary school level. Teachers were directly involved in a professional learning (PL) program designed, primarily, to build capacity in this domain. The evaluation also found evidence of changed teaching practices. Teachers showed commitment to implementing their learnings in the classroom to the extent that students reported recognising a small improvement in effective teaching practice, more-so in the Primary context. Since students were indirectly involved in the program through their teacher, it was anticipated that the level of impact on their achievement would be less, particularly given the short post-test timeline that did not allow for changed teaching practices to have its full impact. In most schools, students were tested only two weeks after the last PL session. That there was a positive impact so shortly after the PL was completed is encouraging and may mean even greater gains in the future for students of these teachers.

The Thinking Maths evaluation logic model with impact evaluation effect size results is below.



* indicates statistically significant effect ($p<0.05$)

Thinking Maths teachers were highly positive about the program and advocated for its wide-spread rollout. The Primary and Secondary teachers reported largely similar barriers and enablers. However, what emerged by the end of the PL sessions, was that Primary teachers, more-so than Secondary teachers, reported the program had increased their mathematics understanding, their use of instructional strategies, and levels of student engagement. A correlation between student and teacher primary and secondary outcomes provided additional evidence that a stronger positive impact was experienced in the Primary schooling context. The Thinking Maths facilitators, in their role of providing consistent support across Primary and Secondary school teachers, identified the following factors that may have contributed to this difference:

- Dosage: Secondary students were only exposed to the 'treatment' of changed teaching for 3 hours per week as opposed to Primary students whose class teacher's shift in pedagogy was likely to impact more widely over the school day.
- Resources: A lack of concrete materials and equipment as well as shorter or inflexible lesson length may be a factor in the Secondary context.
- Flexibility: Fixed curriculum programming may not have allowed Secondary teachers the flexibility to trial tasks if they were off-topic.

Evidence for Learning’s security rating – How secure is this finding?

The primary finding has high security (see Appendix B for the security rating). This was an efficacy trial which tested whether the teacher professional learning intervention can work under developer-led conditions. It was a randomised controlled trial that included 167 schools at recruitment, with 158 schools that participated. Nine schools were dropped from the trial, equivalent to only 4.1% of students. This was mainly due to the non-response, where the test or survey was either not administered to students (e.g. absent on the day of testing) or the data could not be matched to a participant (e.g. they had moved school since the pre-test). There was good balance at baseline for the analysed sample. Pre-test surveys were carried out before randomisation; the evaluator, teachers and students were blind to the allocation.

The security of the trial was compromised by the re-randomisation process, which was driven by the number of teachers that needed to be in the intervention group (to meet the cap of 120) rather than on the randomisation of schools (that may have resulted in a number of teachers more or less than the exact 120 that would receive the intervention). There were some significant differences in the baseline characteristics of teachers in the intervention and control groups (e.g. years of experience). This introduces the risk that any difference in outcomes between the two groups is caused by the different composition of the groups, not by the impact of the intervention. Also, some randomised schools had other teachers previously receive training in the intervention prior to the study. Although none of the evaluation teachers had any prior exposure, the risk of possible contamination from colleagues was considered and accepted by all parties at protocol stage (ACER, the Department, and Evidence for Learning).

How much does it cost?

The cost of the Thinking Maths program is estimated at $149 AUD per student per year. This estimate includes training and materials ($1070 per teacher or $43 per student), and the significant cost of five Temporary Relief Teacher (TRT) days replacement ($2650 per teacher or $106 per student). Estimates are based on training being delivered to a group of 35 teachers with an average class size of 25 students, reaching 875 students. This amount is rated as *very low*, according to the Evidence for Learning Cost Rating approach, based on the approximate cost per student per year of implementing the intervention over three years (see Appendix A). As a Department-developed and delivered program, all costs were borne by the Department.

Research results

Intervention vs control	Effect size [95% CI]	Estimated months progress*	E4L security rating+	Number of students	P value	E4L cost rating
All students	0.05 [0.00 – 0.10]	+1	4 of 5 padlocks	7068 students in 158 schools	0.38	$ (1 of 5)
Primary Years 5-7	0.14 [0.08 – 0.19]	+2	N/A	5013 students in 119 schools	0.05	$ (1 of 5)
Secondary Years 8-10	-0.16 [-0.25 – -0.07]	-2	N/A	2055 students in 56 schools	0.05	$ (1 of 5)
School Card holders	0.11 [-0.04 – 0.27]	+1	N/A	666 students in 118 schools	0.21	$ (1 of 5)

* Refer to Appendix A, used to translate effect size into estimated months progress.
+ Refer to Appendix B, for E4L independent assessment of the security rating.

Table of Contents

List of Tables

List of Figures

1 Introduction

This report presents the findings of the outcome and process evaluation of the South Australian Department for Education's (the Department) Thinking Maths program, independently conducted by the Australian Council for Educational Research (ACER).

1.1 Intervention

The Thinking Maths program is designed to support teachers to improve students' learning of mathematical content, as outlined in the Australian Curriculum Mathematics, during the transition between Primary and Secondary school (currently Year 7 and Year 8 in South Australia). To promote improved teaching and learning in mathematics, the professional learning (PL) sessions used hands-on work in mathematics – demonstrated by the facilitators – to support teachers' pedagogical content knowledge in three areas: (a) using quality task design, b) sequencing of conceptual development and c) using research-informed effective pedagogies (details explained in Section 1.5).

The Thinking Maths program aims to address a significant drop in students' mathematics performance in NAPLAN[6] from Year 7 to Year 9. One main intention of the program is to encourage deeper learning through collaboration between Primary and Secondary teachers. For Primary teachers, it seeks to promote more conceptual mathematical understanding and intellectual stretch. For Secondary teachers, it seeks to promote alternative conceptual, concrete ways of thinking and encourage pedagogical shift. These professional conversations, over levels of schooling, are intended to provide more continuity of learning for students during the transition stage between Primary and Secondary school.

The intervention period began in Term 1 and finished in Term 3 of the 2017 school year, and included two mid-term breaks. During this period, students either attended 'Thinking Maths-enriched' maths classes or business-as-usual maths classes, based on their schools' randomised allocation. All students then completed a standardised mathematics achievement test (the primary outcome measure).

A pilot of Thinking Maths was undertaken in 2016 with two groups of 30 teachers, in preparation for the evaluation. While feedback from teachers who participated in the early stages of the program's development were promising, a rigorous evaluation was sought to understand its effectiveness in building teacher capability and improving student mathematics outcomes.

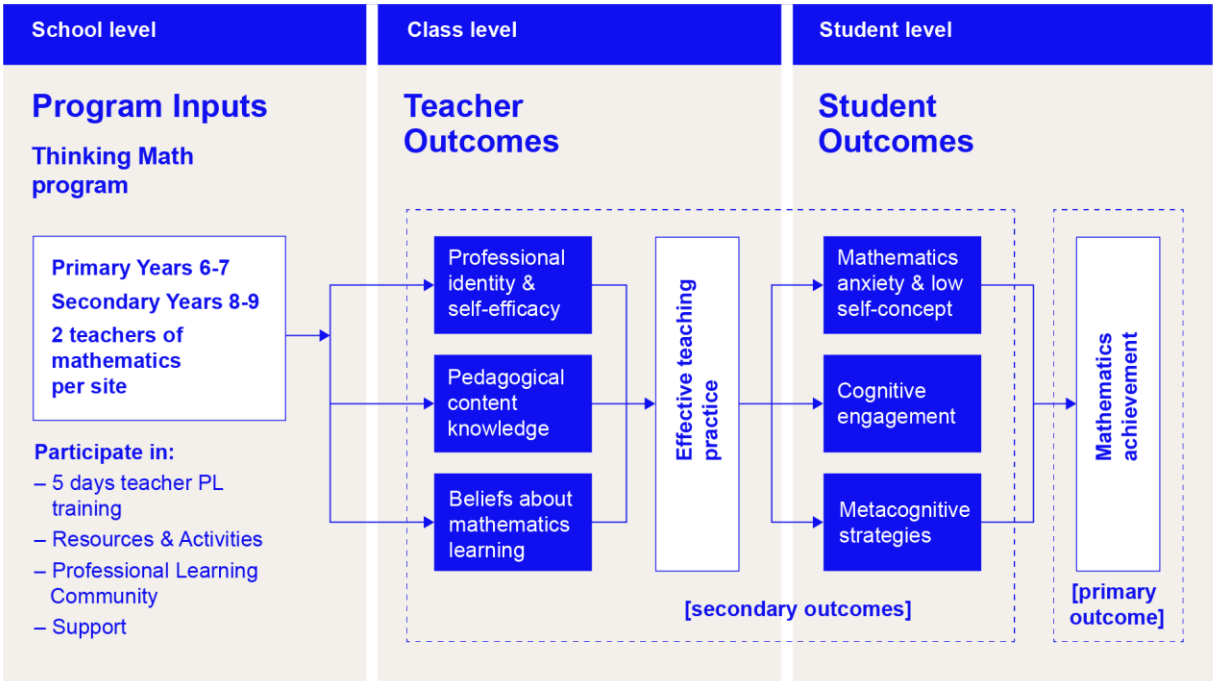
[6] The National Assessment Program – Literacy and Numeracy (NAPLAN) is an annual assessment for students in Years 3, 5, 7 and 9 in Australia.

1.2 Evaluation objectives

As detailed in the evaluation protocol and Statistical Analysis Plan (available from: evidenceforlearning.org.au/lif/our-projects/thinkingmaths/), the broad research question underpinning the evaluation was: *To what extent does the Department's Thinking Maths program improve student mathematics outcomes by improving teachers' capability to make mathematics learning deeper and more engaging?*

The key elements emerging from this guiding question, involved teacher capabilities and teaching practices, and student mathematics capabilities and achievement. These elements and their relational influences are shown in the evaluation logic model in Figure 1 and underpin the evaluation research questions. In essence, Year 6-9 teachers participate in professional learning with the commitment that between sessions they reflect on and apply program ideas in their mathematics lessons. The logic model reflects anticipated changes in teachers' professional identity and self-efficacy, pedagogical content knowledge and beliefs about mathematics teaching that result in changes in teaching practice. These in-turn influence students' mathematics self-efficacy, cognitive engagement in learning, and metacognitive strategies, with the outcome of improved learning.

Figure 1. Thinking Maths evaluation logic model



The primary research focus was to identify the impact on student mathematics achievement, due to teachers eight-month involvement in the Thinking Maths program, compared to teachers in 'business-as-usual' classrooms.

Specifically, the primary research question asked:

1. Did the Thinking Maths program enable middle-school students to improve their mathematics achievement above typical learning growth?

A further set of secondary research questions were developed that assessed the impact of the Thinking Maths program on other elements in the logic model (see Figure 1).

2. Did Thinking Maths develop middle-school students as powerful learners of mathematics in terms of a) mathematics self-efficacy, b) cognitive engagement in learning, and c) metacognition?
3. Did Thinking Maths shift teachers' mathematics teaching practice towards a more inclusive, student-centred learning approach? Did changes in teachers' practices due to Thinking Maths influence students' mathematics outcomes?
4. Did Thinking Maths build the capacity of teachers in terms of a) professional identity, b) pedagogical and content knowledge, and c) beliefs about mathematics teaching and learning?
5. How cost-effective is the Thinking Maths program?

In addition to the outcome evaluation, a process evaluation was also undertaken (Humphrey et al. 2016). Six process evaluation research questions were developed. These focussed on the quality of delivery of the four program inputs (professional learning sessions, resources and activities, professional learning community, and support), as well as the implementation of the Thinking Maths ideals in the classroom. The process evaluation questions are discussed in the Methods section.

1.3 Background evidence

The pedagogical principles of the program are research-based and draw on the work of leading educational experts including Sullivan (2011; 2013), Dweck (2000), Claxton (2012), Boaler (2002; 2015), Boaler and Humphreys (2005), and Meyer (2016). The learning resources are drawn from the Teaching for Effective Learning (TfEL) Framework (DECS 2010) and Scootle (2016), as well as organisations such as NRICH (2016), You Cubed (2016) and Estimation 180 (2016). The Thinking Maths program showcases these and other freely available online resources with the intention that teachers incorporate the resources in their learning design to deliver the Australian Curriculum Mathematics in differentiated ways responsive to individual student's needs, interests and dispositions.

The research-informed strategies that the program asks teachers to implement are summarised as:

- Establish a culture of learning;
- Encourage metacognition and conceptual understanding;
- Engage and challenge students in their learning; and
- Professional reflection and networking.

For example, regularly sharing teaching experiences and discussing what works and doesn't work with colleagues, supports *professional reflection and networking*. This in-turn can improve teaching practice and the *learning culture*, as does the opportunity for teachers to become the learner to increase the visibility of learning from the students' perspective (Miller 2009). Studen*ts' metacognition and conceptual understanding* can be improved when students' low self-efficacy is first addressed before trying to raise their achievement (Miller 2009). Teachers are also encouraged to adopt a 'Growth Mindset' to *engage and challenge students* (Dweck 2000; Dweck, Walton & Cohen 2014). Dweck's work highlights links between confidence and self-efficacy in mathematics, mastery of problems, and building resilience, when teachers and students work together. A further strategy involves the program facilitators modelling effective practice over an extended period of time in order to support teachers as they develop the theoretical understanding and tools that will enable them to take a self-regulated inquiry approach to their everyday practice (Timperley 2008). One such modelling is in the nature of questioning, in order to improve teachers' questioning skills (Redfield & Rousseau 1981). Another is to encourage teachers to do problems that require them to apply previously learned knowledge and skills, by using physical manipulatives and working together (Killian 2015; Claxton 2012).

As stated by the program developers:

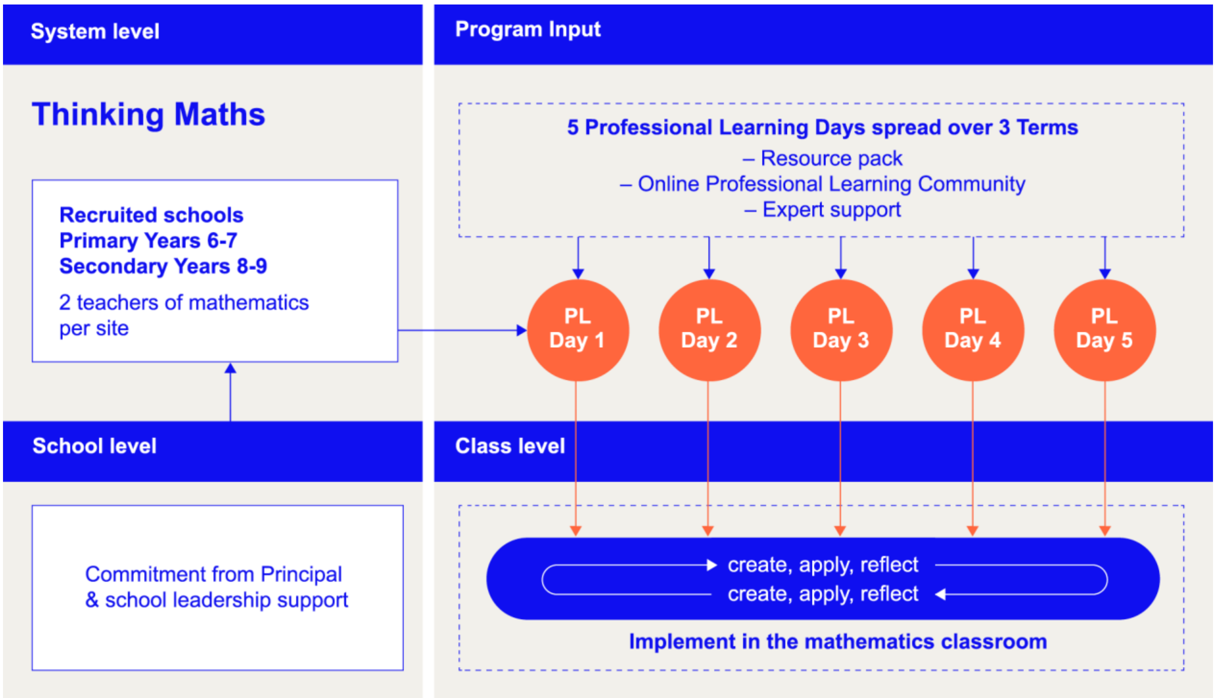
The Thinking Maths program was piloted in 2015, but no assessment of the impact of the program on students' mathematics achievement was undertaken. During the pilot, participating teachers were surveyed to provide their responses to the professional learning experience, including any consequent shifts in pedagogy and any observed changes in student engagement and achievement. Anecdotal evidence through teacher feedback collected from the pilot sessions during 2015 indicated diverse and comprehensive shifts in pedagogical practice and beliefs and attitudes about the teaching and learning of mathematics as a result of their participation in the professional learning program.

Given that no evaluation of the impact of the program on students' mathematics achievement had yet occurred, this evaluation constituted an efficacy trial to assess the beneficial effects of the program under optimal conditions of delivery (Flay et al. 2005). Student achievement was assessed using the Progressive Achievement Tests in Mathematics (PATMaths), routinely collected by the Department in 2016 and 2017.

1.4 What does the program involve

The program comprises four key inputs: the professional learning days, the activities and resources, the professional learning community, and the professional support. Figure 2 presents the sequencing of the four key inputs of the Thinking Maths program. A description of each input follows.

Figure 2. Key inputs of the Thinking Maths program



Professional learning days

Teachers participate in a cycle of five PL days, scheduled 4-5 weeks apart, generally over a period of three school terms. In total, the program involves 30 hours of face-to-face professional learning, with an additional expectation of engagement in professional reading, journaling, and presenting to the group (see table 25 for details). During the sessions 'research-informed strategies' are explicitly demonstrated by the program facilitators with collaborative activities that involve reflecting, sharing, modelling, applying, being the learner/practitioner, and accessing professional resources (see Addendum for Day 1 outline).

During the PL sessions, teachers are challenged to 'wear two hats' – one as a 'learner', as they tackle tasks and experience emotions by putting themselves in the place of the learners in their class, and the other as a 'reflective practitioner', needing to note the facilitators' learning design, questioning, responsiveness, and how that impacts the learner in a task.

The two Thinking Maths facilitators guide the sessions with a mix of whole-room instruction and hands-on group or individual work. Participants (30 teachers at a time) are grouped around six tables with materials (paper, blocks, texts, scissors, etc.) for the various tasks. Teacher-pairs sit together but facilitators ensure that each table has a mix of Primary and Secondary teachers. During group activities, the facilitators circulate around the tables modelling good questioning technique and pedagogical practice, as well as challenging the thinking and assumptions of teachers. Both the pedagogical and content components of each PL day, are designed to significantly extend teachers, both as a learner, and as a practitioner.

The approach gives teachers the rare opportunity to work together with professionals across the Primary and Secondary context and allows time to reflect on and question their professional practice away from the business of school.

Activities and resources

Teachers receive a set of resources, including a reference textbook (Van de Walle et al. 2014) to use during the four intervening periods between the PL days to reflect on and implement program ideas in their mathematics classes. This implementation process follows a cycle of Action, Reflection and Creation. Teachers complete journal entries, read relevant literature, use the resources, and share the evidence of changing behaviours and outcomes during subsequent PL days.

Professional learning community

Throughout the Thinking Maths program, teachers' improvement efforts are supported by telephone, email, and online discussion in a professional learning community. The online component provides teachers with the opportunity to access shared resources and contribute to forum discussions, and while it is not a mandated aspect of the program, it is hoped that teachers will engage with it to some extent during the periods between the PL days.

Professional support

Ongoing support is provided to teachers by the Thinking Maths facilitators, their school leadership, and their colleagues. Thinking Maths facilitators reach out to teachers via email and phone on a regular basis during the intervening periods where teachers are expected to implement the newly learnt research-informed strategies back in their classroom to improve student achievement and engagement.

1.5 Teaching strategies

The program's approach focuses on the following three areas for better teaching and learning of mathematics:

- Using quality task design. For example, clear learning intentions referenced to the Australian Curriculum and delivering engaging lessons with multiple entry and exit points to support students in linking mathematical ideas to solve problems
- Sequencing of conceptual development. For example, encouraging metacognition and growth mindset through the use of effective questioning and differentiating the curriculum by presenting tasks with multiple entry and exit points to cater for students with a wide range of mathematical experience and dispositions.
- Using research-informed effective pedagogies. For example, establishing a culture of learning, encouraging metacognition and conceptual understanding, engaging and challenging students in their learning, and opportunities for professional reflection and networking.

Each of the PL days are structured to include specific content and *pedagogical* elements.

Day 1: Patterns and Generalisation; Differentiating learning

Day 2: Space and Measurement; Effective questioning

Day 3: Geometry; Active and collaborative learning

Day 4: Statistics; Personalising and connecting learning

Day 5: Location, Directed Number and Geometry; Teaching for understanding

As well as being the developers of the program, the facilitators are experienced mathematics educators and professional presenters. The Thinking Maths facilitators aim to model rigorous teaching and learning processes, provide individualised support, and identify current useful mathematics learning resources that meet the needs of participants.

The broad behaviours, actions, and practices teachers are expected to engage with and implement as a result of participating in Thinking Maths are summarised below. Importantly, these research-informed strategies are modelled by the two Thinking Maths facilitators during the PL days.

Establish a culture of learning

- Build student self-efficacy through a positive disposition to maths and a belief that everyone can learn maths.
- Create a safe environment for learning where everyone's thinking is heard and valued. Value mistakes and reward good thinking rather than only the right answer. Ensure there is an entry point for all learners and invite guesses and estimates.
- Promote resilience and have a 'growth' rather than a 'fixed' mindset.
- Foster the belief that all students can learn maths and need opportunities to tackle hard problems.

Encourage metacognition and conceptual understanding through the use of effective questioning

- Change from 'telling students' to 'asking students', encourage students to talk about their thinking and develop their reasoning skills through purposeful questioning. Rather than re-explaining a concept, use questioning to get an insight into the nature of their misconceptions, guide them to expose an inconsistency and allow them to self-correct.
- Provide students with the opportunity to connect with and build on their prior knowledge.

Engage and challenge students in their learning

- Evoke curiosity and wonder, ask them to guess or estimate, allow students to pose their own questions. Make learning active, hands on and experimental.
- Provide opportunities for students to learn from each other.
- Always consider the level of student thinking required by a task. Ensure students of all levels of experience and knowledge know what productive struggle is and are supported to experience it.

Be mindful and observant of

- Their own and their students' attitudes to and beliefs about mathematics.
- The types of questions they ask learners.
- The level of student thinking required in their tasks.
- Whether tasks have multiple entry and exit points appropriate to their students.
- What they reward and value in their classroom.

Undertake the following activities outside of the professional learning days

- Complete required professional reading.
- Utilise the reference textbook by Van de Walle et al. (2014) where appropriate in their learning design.
- Participate in the online discussions.
- Trial strategies and tasks in their own classroom.
- Share resources and upload to PLC (professional learning community) forum.
- Collect and analyse student work samples.
- Keep a professional journal to support their share back at next PL day.
- Share learning with co-participant, colleagues at their site and in their partnership.

1.6 Ethical review and trial registration

In accordance with ACER's Code of Practice, ethics approval to conduct the evaluation was granted by ACER (30/06/2016, Ref no. 544883) and the Department (20/12/2016, CS/16/00075-1.16). Permission was obtained from all participating school principals to seek opt-in consent from nominated teachers. Plain language statements about the evaluation were provided to all participants (principals, teachers, students and parents/carers) and a student opt-out approach was used. Two weeks were allowed to elapse after which consent was assumed. The schools, teachers, parents and students were blind to their random allocation to the treatment or control (wait-list) groups until after the consent and pre-survey period.

A multi-stage Data Access and Transfer procedure set by the Department Business Intelligence Unit provided class lists along with student pre/post PATMaths achievement data and background data (gender, age, Aboriginal or Torres Strait Islander background, School Card, Disability). It was of critical importance to adhere to strict data security protocols to ensure participant privacy.

The Thinking Maths Protocol is registered with the Australian New Zealand Clinical Trials Registry: ACTRN12618000437268, www.ANZCTR.org.au/ACTRN12618000437268.aspx.

1.7 Project team

The program delivery team was led by Ken Lountain, Executive Leader, Strategic Design, in the Department's Learning Improvement Division. The Thinking Maths facilitators, Pauline Carter (Department Professional Officer, Critical and Creative Thinking) and Maureen Hegarty (Mathematics Consultant), co-developed and co-presented the Thinking Maths program, with assistance from Kath Ireland (Project Officer, Primary Learners Directorate). In addition to the delivery of the program, the team recruited schools, co-facilitated the Thinking Maths Briefing Event, and provided evaluation support.

The independent evaluation team at ACER's Educational Monitoring and Research Division, was led by Hilary Hollingsworth (Principal Research Fellow, Melbourne). Katherine Dix (Senior Research Fellow, Adelaide) undertook the impact and process evaluations with assistance from Toby Carslake (Research Officer, Adelaide). Katherine Dix was the trial statistician and led the writing of progress reports, the Statistical Analysis Plan (SAP), and the final evaluation report and Addendum.

2 Methods

In order to achieve robust findings that were representative of government schools in South Australia, the trial design, sample size and statistical approach, detailed in this section, were selected to detect an effect that was sufficiently large to be of educational significance at the student level, accounting for the nestedness of data and prior performance.

2.1 Trial design

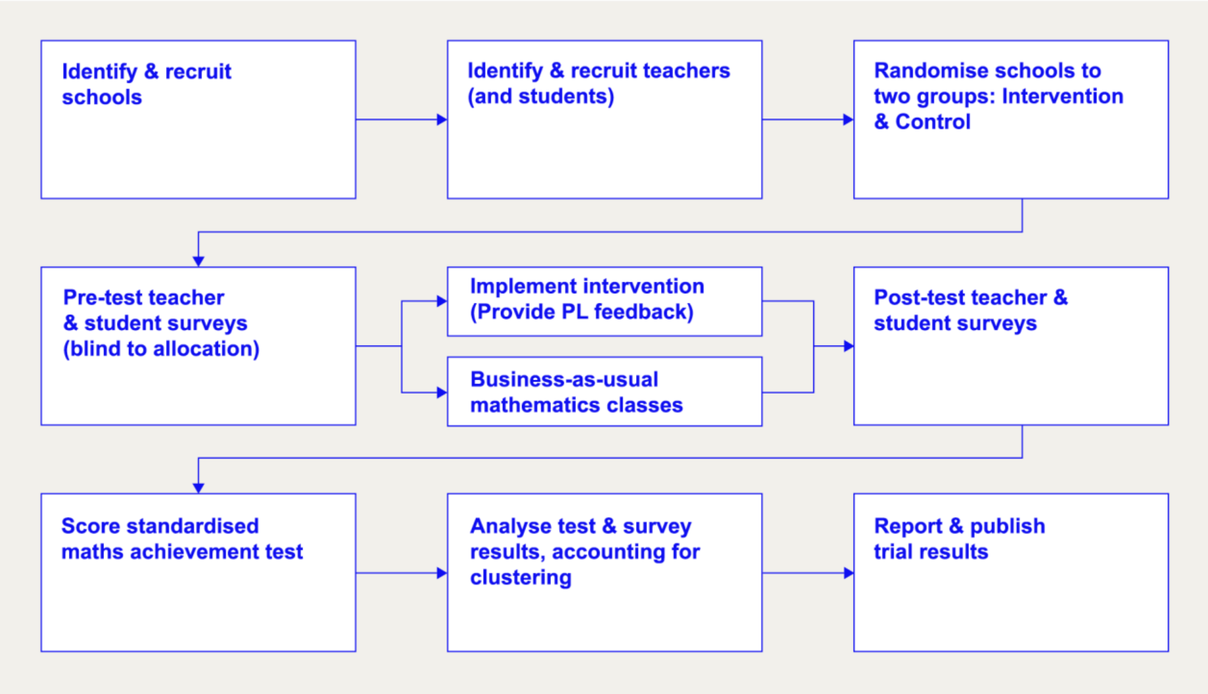
This evaluation was a multi-site, two-arm, parallel group, efficacy trial in South Australia, involving quantitative pre and post data collection. Randomisation was done at the school level. This approach was responsive to the program's design, which involved training two teachers from a school to take their learnings back into their intact classrooms. Since the program directly involved teachers and not students, it was not appropriate to randomise at the student level. A total of 167 schools were recruited through a self-selection process by submitting an Expression of Interest (EOI), along with the nomination of two teachers of students in Years 6-9 mathematics classes in each school. Teachers were approached by their principal or an announcement was made and teachers volunteered. Some small schools only nominated one teacher.

Schools were randomly assigned to the intervention (63 schools) or the control (104 schools) and met the cap of 120 places available in the PL sessions. The intervention schools commenced in Term 1 2017, while 'business-as-usual' schools acted as a wait list control group, receiving the intervention after trial completion in Term 4 2017.

Quantitative pre and post data collection included an assessment of mathematics achievement using PATMaths tests (pre-test in September 2016, post-test in September 2017), along with teacher and student pre and post online surveys (see Outcome Measures section). Thinking Maths intervention teachers also completed five Professional Learning Feedback Forms (see Addendum). This data, collected by ACER, was augmented with existing student background data provided by the Department (e.g. student name, EDID, date of birth, gender, disability, ATSI, School Card status). The overarching approach, adapted from Torgerson and Torgerson (2013), is shown in Figure 3.

Apart from the number of schools recruited to the project, which was increased from 120 to 167 schools due to an overwhelming response during recruitment, and that block randomisation was not used, the design of the trial remained consistent with the design outlined in the published protocol.

Figure 3. Key steps for school-based RCT



2.2 Participant recruitment and selection

Recruitment through an EOI was undertaken at the school level by the Department in order to attract at least 150 sites, where 63 sites received the intervention and remaining sites acted as control. The control group included additional schools to allow for control attrition, in addition to any overflow due to popularity. In total, 167 schools were recruited. It was preferable to have two teachers from each school receiving the intervention (120 teachers), but not to the exclusion of small schools, or schools with only one mathematics teacher. In addition to an emailed flyer and webpage, the Department also advertised the program through school network partnerships, the South Australian Primary Principals Association, the South Australian Secondary Principals Association, and social media.

Eligible schools met the following criteria:

- Government school located in South Australia.
- School catered for students in Years 6-7 and/or Years 8-9 (K-12 Area schools were counted as one site).
- The teacher preferably taught a Year 6, 7, 8 and/or 9 class in mathematics, but not to the exclusion of small or remote schools. Teachers of Year 5 and Year 10 students were included. Although most teachers involved in the evaluation taught Years 6 to 9 (91% of students), based on teacher availability, Year levels extended from Year 5 (7.8%) to Year 10 (0.6%).
- The teacher had not previously participated in the Thinking Maths intervention.

In order to participate in the program and the evaluation, schools agreed to:

- Participate in a briefing session, prior to random assignment and data collection.
- Be randomly assigned to have the Thinking Maths program in either a) the 2017 school year or b) late 2017 – 2018.
- Cooperate with the evaluation team to provide teacher contact details and student class lists, support consent procedures, and allocate time for online surveys.
- Students were eligible if they belonged to the class of a participating teacher and had a PATMaths 2016 score and a PATMaths 2017 score.

2.3 Sample size

In order to detect an effect that was sufficiently large to be of statistically practical significance at the student level (i.e. above 0.2), and given that teachers were clustered within schools, the following recommendations about sample size were provided at the initial design stage.

The desired alpha was 0.05 and power was 0.8, with a minimum detectable effect size (MDES) of small (Cohen's d = 0.2). We also needed to take into account the design effect of clustering by including an estimate for the intra-cluster correlations (ICC). This accounts for students in one school being more like each other compared to students in another school (Hutchison & Styles, 2010; Eldridge et al., 2006) when the sample is not a simple random sample, resulting in a net loss of information. In other words, from a statistical perspective, similarities between students in the same class effectively reduce the number of participants in the intervention (Torgerson & Torgerson 2013). The 'design effect' was used to estimate the extent to which the sample size should be inflated to accommodate for the homogeneity in the clustered data. In similar studies in Australia, Zopluoglu's (2012) recommended an Australian ICC coefficient range of 0.2-0.3 (p.264) and the PISA 2012 Technical Report used an Australian ICC for mathematics of 0.28 (OECD 2014, p.439). Taking a conservative approach, we adopted an initial ICC coefficient of $\rho = 0.3$.

In order to minimise sample size and achieve the desired Minimum Detectable Effect Size (MDES) of 0.2, the Bloom MDES formula with both level-1 and level-2 covariates was used (Bloom, Richburg-Hayes & Black 2007), which increases the power of a cluster-level RCT by including pre-post test correlation. The hierarchical model controls for the majority of variance, which is known to be explained by prior achievement, both at the school level and the student level. The remaining variance, therefore, is more sensitive to explaining the impact by teacher participation (or not) in the intervention.

Accordingly, a minimum sample of 120 schools (60 intervention, 60 control) was needed to achieve a MDES of 0.2 with covariates that accommodate design effects and provide allowances for participant attrition and missing data. Through the recruitment process, a sample of 167 schools was achieved – 63 intervention, 104 control – well over the minimum number of 120 schools needed. Table 1 presents the MDES at different stages of the study.

Table 1. Minimum detectable effect size and Intra-class Correlations at different stages

Stage	N Schools n=treat; n=control	Correlation between pre & post test	ICC	Stratification	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	150 (60; 90)	-	0.30	NA	80%	0.05	0.18
Randomisation	167 (63; 104)	0.78	0.20	NA	80%	0.05	0.15
Analysis	158 (63; 95)	0.78	0.21	NA	80%	0.05	0.15

Pre and post test data was not available at the protocol stage, and as described above, ICC was estimated to be 0.30. At the randomisation stage, once pre-test data was available, ICC was calculated to be 0.20, and this was recalculated to 0.21 in final analysis. Given that the calculation of MDES involves ICC, the MDES values were recalculated at each stage.

The sample achieved through the recruitment, resulted in an unequal allocation of school clusters to each treatment arm. In effect, the control arm was over-sampled by two-thirds at randomisation and by a half at the point of analysis (Hayes & Moulton, 2017). With cluster sizes ranging from 8 to 92 students, the average cluster size for the 95 schools comprising the control group was 44.9 (*sd* =14.3) students. The average cluster size for the 63 schools comprising the intervention group was 44.5 (*sd* =13.7) with cluster sizes ranging from 15 to 75. These were close to the value of 50 students per cluster assumed in the protocol. Note that in South Australian government schools, classes have a starting cap of 24 students but may be as many as 30 students by the end of the year. In some schools, team teaching can result in a double-class resulting clusters of twice the assumed value. In this study, there were seven double-classes in the control group and six double-classes in the intervention group.

The method by Eldridge et al. (2006, p.1292) was used to “judge the possible effect of unequal cluster size”, based on the relationship between the design effect and coefficient of variation of cluster size, as the ratio of standard deviation of cluster sizes to the mean cluster size. The results indicate the effect of unequal cluster size is trivial (within a margin of 10%) when the coefficient of variation is less than 0.33. The coefficient of variation for the control and intervention groups were 0.32 and 0.31, respectively. Importantly, with an effect size of ES=0.02, there was no differences the distribution of clusters between the control and intervention groups.

2.4 Randomisation

This study used concealed randomisation[7] so that there was no foreknowledge of the randomised allocation (Torgerson & Torgerson 2013). Randomisation at the school level was done after teachers had been recruited and consented to participate in the study. This occurred after a Briefing Event and once they had completed the Teacher pre-survey. Accordingly, all participants, including teachers, schools and the Department (the recruiters and program implementers) did not know which group the schools were randomised into until their baseline pre-survey had been submitted.

In order to maintain independence and concealment from the program implementers (the Department) and the evaluation funders (Evidence for Learning), the Department provided the sampling frame of participating schools to ACER, and ACER undertook a simple random sample of cases on the de-identified list of schools in which only a school ID number and the number of participating teachers was known. Specifically, the 'Select cases: random sample' dialog box in SPSS 22 was used to select a specified number of schools from the total list. As described next, several independent attempts were necessary. The trial statistician performed the random sample once a colleague, independent to the project, had de-identified the list and assigned the school ID number. Accordingly, the evaluation statistician was blind to the schools being randomly allocated to either the treatment or control groups during the randomisation procedure.

Because the Department developers budgeted for an allocation of 120 teachers to undertake the Thinking Maths Professional Learning sessions in 2017, it was necessary to achieve a random sample of schools that resulted in 120 teacher places. However, while most of the 167 participating schools nominated two teachers, as requested, a small number of schools only nominated one teacher, and two K-12 schools nominated four teachers (two in the Primary year-levels and two in Secondary). A total of 318 teacher places were initially indicated by schools. The first sampling attempt specified 60 schools to be randomly drawn in SPSS, but yielded fewer than 120 teacher places and was rejected. A simple random sample was redrawn, but with 63 schools specified, at which point the desired number of 120 teacher places was achieved. This group of 63 randomly sampled schools and their 120 teachers formed the intervention group. The remaining 104 schools and their teachers formed the control group. The number of students was not known at the time of randomisation, however, once pre-test data and class lists were provided, this involved 2922 students in the treatment group and 4445 students in the control group.

It should be noted that while no nominated teachers had previously undertaken Thinking Maths (in accordance with the selection criteria), 26 participating schools had one or two other teachers complete Thinking Maths in a previous year. Nine of these schools were in the intervention group and 17 schools were in the control group. The risk of 'contamination' was considered early on at the protocol stage by all parties – Evidence for Learning, the Department, and ACER – and was deemed to be negligible. Moreover, it was considered unfair to exclude schools (and all eligible teachers) from the evaluation based on the participation of one or two other teachers in a previous year.

[7] The Protocol originally stated a block design, but given the larger recruited sample it was decided at the time of randomisation, not to use any stratification.

2.5 Outcome measures

In order to address the main research questions emerging from the evaluation logic model (see Figure 1), one primary outcome measure and seven secondary outcomes measures were selected or developed using pre-existing validated scales. This section briefly presents the outcome measures used in this evaluation.

Primary outcome and baseline testing

The primary outcome identified in this evaluation – the outcome that determined whether or not the intervention was effective – was improved student achievement in mathematics for learners in the intervention group compared to the control group. This was measured by the ACER Progressive Achievement Tests in Mathematics (ACER 2011).

PATMaths is a thoroughly researched, Australian test designed to provide objective, norm-referenced information to teachers about the level of achievement attained by their students in the skills and understanding of mathematics. All PATMaths tests have a common achievement Rasch scale, enabling results to be compared between different Year levels. The PATMaths Fourth Edition tests cover six mathematics strands, namely, Number, Algebra, Geometry, Measurement, Statistics, and Probability. Each test comprises at least five items for each of the strands it covers with a total of 40-50 items depending on the Year level. Within a test, the items are ordered from easiest to most difficult. The test is completed by students online within a 40 minute timeframe, as instructed by the classroom teacher. Further information about ACER PATMaths is available at www.acer.org/pat/tests/mathematics.

The test is scored instantaneously through the ACER Test Scoring and Analysis software. The PAT raw score is the number of correct answers on a test. The PAT scale score is the test raw score converted to the relevant PAT scale. Based on analysis of the data using the Rasch model, this scale enables student achievement and question difficulties to be located on the same scale across Year levels. The standardised PAT scale score for middle-school cohorts generally range between 50 to 200 scale units and is the primary outcome measure used in this evaluation. A positive pre-post coefficient difference indicates learning growth.

Secondary outcomes

The secondary outcomes identified in this evaluation were captured through two surveys (discussed below) using attitudinal responses to sets of conceptually similar items on five-point Likert scales. Confirmatory factor analysis and Item reliability analysis were conducted in order to summarise the items and derive meaningful constructs in the form of mean scores. Reliabilities of 0.8 or more are described as high; between 0.7 and 0.8 as moderate; and between 0.6 and 0.7 as low.

The four secondary outcomes collected through the Student Survey (see Addendum) were:

- **Students' mathematics anxiety and low self-concept** (SASE): Mean score of 10 items (e.g. *I get nervous doing maths problems*; *I am just not good at maths*) measured on a five-point Likert scale of Strongly disagree (1) to Strongly agree (5), with high internal reliability (α=0.89).
- **Students' cognitive engagement** (SCOG): Mean score of five items (e.g. *I know what my teacher expects*; *My teacher believes all students can be good at maths*) measured on a five-point Likert scale of Strongly disagree (1) to Strongly agree (5), with high internal reliability (α=0.81).

- **Students' metacognitive strategies** (SMET): Mean score of five items (e.g. *I check my maths school work for mistakes*; *I try to connect the things I am learning in maths with what I already know*) measured on a five-point Likert scale of Strongly disagree (1) to Strongly agree (5), with moderate internal reliability (α=0.76).
- **Students' learning through effective teaching practice** (SETL): Mean score of 16 items (e.g. *My teacher asks me to explain my answers*; *We work in groups to come up with joint solutions to a problem*) measured on a five-point Likert scale of Never (1) to Always (5), with high internal reliability (α=0.89).

The three secondary outcomes collected through the Teacher Survey (see Addendum) were:

- **Teacher professional identity and self-efficacy** (TPID): Mean score of seven items measured on a five-point Likert scale of Not at all (1) to A great deal (5), with high internal reliability (α=0.89). For example, teachers were asked to what extent they could *motivate students who show low interest in maths*, and *create opportunities for all students to experience productive struggle*.
- **Teacher pedagogical content knowledge** (TPCK): Mean score of 10 items measured on a five-point Likert scale of Not at all (1) to A great deal (5), with high internal reliability (α=0.91). For example, teachers were asked how confident they were in *using questioning to develop students' conceptual understanding*, and in *identifying students' learning challenges*.
- **Teacher beliefs about mathematics learning** (TBEL): Mean score of three items (e.g. *I deeply believe that everyone can learn maths*) measured on a five-point Likert scale of Strongly disagree (1) to Strongly agree (5), with low internal reliability (α=0.68).

Data collection

For the collection of the **primary outcome measure**, students sat the pre-test during September 2016 and the post-test (the primary outcome) in September 2017, only two or three weeks after the final PL session. Note that the administration of PATMaths occurs state-wide at the same time each year, so the short timeframe between program completion and testing was unavoidable.

The Department provided ACER with the pre (2016) and post (2017) PATMaths scale scores for each participating student. The resulting database was coded and de-identified following data-linkage of the achievement data to student pre and post survey data. Reflecting the nested nature of the data, students were also linked to classes (teachers) and schools.

For the collection of the **secondary outcome measures**, two online surveys, the Student Survey (see Addendum) and the Teacher Survey (see Addendum), were purpose-designed for the evaluation to assess teachers' and students' beliefs, attitudes, and behaviours related to teaching and learning mathematics.

Scales and items were designed with pre-post capacity in mind and to be appropriate for participants in Primary or Secondary settings, as well as those either in the control or treatment groups. As part of the process of survey design, a review of the literature regarding attitudes and beliefs towards mathematics teaching and learning was conducted to source candidate scales and items for the instruments (e.g. EEC 2016; OECD 2014; Fredricks & McColskey 2012; Pintrich & DeGroot 1990; Daraganova, Edwards & Sipthorp 2013; Dix et al. 2010). Where necessary, items were modified or new items were developed to meet the specific needs of the evaluation.

The Teacher and Student Surveys were administered online (using ACER's platform) and were designed to take no more than 30 minutes to complete (well within a lesson time). Participating teachers were provided with links to the surveys through an email invitation with instructions to complete the Teacher Survey and administer the Student Survey. Survey administration occurred on two occasions. The pre-surveys were conducted in February 2017 before recruited schools were given their random allocation to the control or treatment groups. The post-surveys were conducted in October 2017 following the PATMaths administration period. Responses were automatically scored and collated into a secure downloadable database through the online survey hosting platform. Pre and post survey data were cleaned and, along with the PATMaths data, matched using student class lists, preserving the nestedness of students and teachers in classes in schools, at which point the data was de-identified.

Response rates for the PATMaths test (primary outcome measure) and the Student and Teacher surveys (secondary outcome measures) at baseline and post-test are presented in Table 2.

Table 2. Response rates for the primary and secondary outcome measures

Outcome	Instrument	N	Baseline	%	Post-test	%	Total
Primary	Student PATMaths	7367	7367	100.0	7068	95.9	14435
Secondary	Student Survey	5951	5930	99.6	4606	77.4	10536
	Teacher Survey	304	300	98.7	264	86.8	564

2.6 Analysis

A statistical analysis plan (SAP) was developed (see evidenceforlearning.org.au/lif/our-projects/thinkingmaths/). The statistical approach and sample size used in the evaluation were selected to detect an effect that was sufficiently large to be of educational significance at the student level, accounting for the nestedness of data and prior performance. The research design and statistical approach used methodology to achieve findings that were representative of government schools in South Australia.

The primary aim of the analysis was to assess whether the Thinking Maths program had a significant impact on students' mathematics achievement, as measured by the post-intervention test scores and controlling for prior attainment in the form PATMaths test scores. Analysis was conducted in HLM Version 6 using the principles of intention-to-treat. That is, the original random assignment to treatment and control group was reflected in the analysis, regardless of whether the student actually received the intended intervention (not withstanding missing data). All the preparatory data processing was performed in the statistics tool SPSS 22.

The primary student outcome measure, student mathematics achievement (PATM17), was analysed using a hierarchical linear model (HLM) to reflect the nested nature of the data and the method of treatment, with students nested within classes, within schools. The student model included individual student's prior attainment PATM16 score. Intervention and control groups were compared by including an intervention indicator at the school level, where Intervention = 1 and Control = 0.

The secondary analysis looked at the outcomes from the teacher and student survey questionnaires (see Addendum). Equivalent two-level models were tested for all student and teacher secondary outcomes using the derived variables (as explained in the Outcome Measures section above) from pre and post survey results.

Effect size

Statistical significance was assessed at the 5% level. Multi-level regression methods of analysis were used and 95% confidence intervals are reported (Hill et al. 2008). Effect sizes were calculated in accordance with E4L guidelines and are presented alongside 95% confidence intervals.

Although the evaluation design suggests that the effect size should be calculated taking into account unequal cluster sample sizes, Hedges (2007, p.346) argued that “the results become considerably more complicated when cluster sample size are unequal – sufficiently complicated that it is difficult to obtain much insight from examining the formulas in the case of the unequal cluster same size”. He goes further to suggest that the effect size formulas for equal cluster sample sizes provides a good approximation and “close to the exact values” (p.351). This approach, used by others (e.g. Shriver 2009; Cunningham 2010), avoids the complexity of the unequal cluster sample size formula, risking the likelihood of misleading results. Cunningham (2010, p.19) reported that, “The difference in cluster sizes is often ignored because there are very few appropriate and easy-to-use formulas.” As stated by Hedges (2007, p. 352), “The use of cluster means as the unit of analysis is a common approach”. The cluster means are a suitable approximation that can then be used in the effect size formula for unequal cluster sample sizes. When determining the effect size, we used the total variance, rather than the residual variance from the clustered model, noting that variations in a post-test outcome due to different sources must be fully accounted for in a statistical model (Tymms 2004; Hedges & Hedburg 2007; Hedges 2007). Such a method (rather than using Hedges’ g) is necessary due to the use of multilevel models in the analyses (Torgerson et al. 2016). Further details are provided in the Statistical Analysis Plan.

Imbalance at baseline

At the time of school assignment, only school-level information and the number of nominated teachers in each school were available. Imbalance in student prior achievement (PATM16 scores) was assessed in a HLM model. The direct effect of Treatment on PATM16 (i.e. intercept) tested if there was a significant difference in the maths achievement at baseline in the two groups of schools.

Missing data

As with any data collection process, missing data arose for several reasons. In a very small number of cases, items appeared to be inadvertently missed or the participant chose not to answer. While most surveys were completed, a small number were only partially completed. Where 20% or less of items were missing in a scale, the scale score was derived based on the remaining items, thus avoiding the need to impute missing data. For scales with more than 20% of items missing, the value was coded as missing (-999).

The main form of missing data was non-response, where the test or survey was either not administered (e.g. absent on the day of testing) or the data could not be matched to a participant (e.g. they had moved school since the pre-test). This included primary outcome data that was not provided by the Department for the 299 students of 14 teachers from nine schools who withdrew early in the evaluation period (see the CONSORT diagram, Figure 4). The small amount (4.1%) of missing data in the primary outcome (PATM17) was within acceptable ranges, and Little’s (1988) MCAR test returned a non-significant results, indicating that missing values were completely at random between control and treatment groups (chi-squared =1.49, p=0.22). A non-response bias analysis was undertaken at the school, teacher and student levels to identify to what extent the recruited sample was representative and comparable.

Given that the HLM analysis required no missing data, analysis was undertaken using listwise deletion of any participants with incomplete information.

Subgroup analyses

School Card: Subgroup analysis was conducted for the population of School Card students (Student School Card holder = 1). The School Card scheme offers financial assistance to low-income families to assist with school fees for students attending government schools in South Australia. Typically, these students belong to low socioeconomic communities. For this analysis, the primary and secondary outcome analysis models were re-estimated but only using the data limited to this sample (n=666).

Cohort: Similarly, subgroup analysis was conducted for the population of students and teachers in the Primary years of schooling (Years 5-7) and in the Secondary years of schooling (Years 8-10) using data limited to each sub-sample.

Additional analysis

A Year-level analysis was undertaken by comparing student learning gain in the control and treatment groups. Year 5 and Year 10 student results were not included due to insufficient sample size. A cohort-level analysis was also conducted to test for any relationships between student and teacher outcomes.

2.7 Implementation and process evaluation methods

The main purpose of the process evaluation was to understand the implementation of the Thinking Maths program and to identify elements of successful delivery, along with areas for refinement. Typically, findings can be used formatively (e.g. providing feedback that helps developers refine their intervention), summatively (e.g. helping to explain impact, or lack thereof), and for knowledge generation (e.g. improving our understanding of how an intervention works). The findings provide for triangulation, improve understanding, and give context to the results found from the primary and secondary outcomes. With this in mind, the following six process research questions were developed.

1. What are the critical elements of the Thinking Maths program, in terms of quality of delivery, fidelity and dosage?
2. How applicable and useful is the Thinking Maths approach (PL, online community, support, resources) in primary and secondary school settings?
3. What are the barriers and facilitators to the effective implementation of Thinking Maths in middle-school classrooms in different contexts?
4. To what extent do teachers engage with the Thinking Maths program?
5. How can the Thinking Maths program be improved?
6. What are the risks and challenges in expanding the Thinking Maths program to scale?

A cross-sectional design was conducted in three stages, which involved observation of the delivery of the program, a Briefing Event to support evaluation quality, and the Thinking Maths feedback survey.

Observation of professional learning days

ACER took the opportunity during 2016 to observe the five days of professional learning provided in the Thinking Maths program to gain a comprehensive understanding of the approach and strategies used, in the absence of documentation, and to inform the development of the evaluation tools and processes. The 2016 program involved two groups of 30 teachers of Years 7 and 8 students from Primary and Secondary schools. The 2016 PL days, held at the Education Development Centre, were delivered by the same Thinking Maths facilitators between June (Term 2) and November (Term 4) 2016 and followed the same time-frame and content as the 2017 program under evaluation. As an example of the type of information collected, a lesson map for Day 1 is provided in the Addendum. The key learnings from the observations supported the development of the Teacher and Student surveys (for impact data) and the PL Feedback survey (for process data), resulting in effective instruments that met the needs of the evaluation.

Briefing Event to support evaluation quality

In the lead up to the evaluation and prior to randomisation in early February 2017, a Briefing Event was conducted to ensure that participants in the evaluation were informed about the purpose, structure, content and timing, and their role in the evaluation. This event was seen as important for securing participant cooperation and supporting RCT quality, particularly to strengthen the buy-in from control schools, once revealed. Two live sessions were streamed to webinar, with over 250 people attending in person and 60 sites/people attending online.

Process data collection

In addition to the Student and Teacher Surveys developed to address the impact evaluation (see Addendum), a teacher Professional Learning Feedback Form (see Addendum) was specifically designed to address the seven process evaluation questions. The feedback form was administered online at the end of each of the five PL days and also served as a measure of compliance.

The response rates for the teacher Professional Learning Feedback Form administered on five occasions are provided in Table 3. Some 520 forms were received from the 117 teachers participating in Thinking Maths, resulting in an average response rate of almost 90%. The very high level of survey participation was due, in the main, to the strong encouragement and support given by the Thinking Maths facilitators.

Table 3. Thinking Maths Professional Learning Feedback Form response rates

PL sessions	Day 1	Day 2	Day 3	Day 4	Day 5
Number of responses received	116	117	102	92	93
Response rate (n=117)	99.1%	100.0%	87.2%	78.6%	79.5%

Compliance

Given the complexity of implementing a program in schools, it was anticipated that some teachers would engage more readily than others with the Thinking Maths program, and in turn, would be better able to effect change in the classroom. An assessment of compliance with intervention was conducted. Because the Thinking Maths program was enacted through the participating teachers, it was these treatment group teachers that were identified as the 'compliers'. Our assessment assumed that:

1. Randomisation worked – the number of controls who would have been 'non-complier' with the treatment (if they had been offered it), was the same as the number of non-compliers with the treatment who were offered it;
2. Non-compliers who were offered the treatment had the same treatment as the controls (i.e. business-as-usual); and
3. Simply being offered, the treatment didn't affect the outcome.

Compliance was represented as a numerical score, based on the number of PL days each teacher (or their replacement) had attended (scored 0 to 5). The responses to the PL Feedback Form, collected at the end of each PL day during the evaluation period, was used to verify the participation data (sign-on sheets). An acceptable threshold for compliance was the attendance to at least three training days, providing they did not occur consecutively on days four and five.

2.8 Collecting cost data

Prior to commencing the efficacy trial, ACER and the Department discussed the need to collect and document the costs associated with running the Thinking Maths program (Hummel-Rossi & Ashdown 2002; Levin 1995). The Department maintained detailed accounts of the start-up and running costs and then derived estimates based on a group of 35 teachers. While the evaluation involved cohorts of 30 teachers, the facilitators felt that the group could feasibly be expanded to accommodate 35 teachers without loss of quality. The 'ingredients method' (Chambers & Parrish 1994; Levin 1995) was followed, which accounts for the costs of the resources required to implement the educational intervention being evaluated, rather than focusing on a budget. The approach involved "three distinct phases: (a) identification of ingredients; (b) determination of the value or cost of the ingredients and the overall costs of an intervention; and (c) an analysis of the costs in an appropriate decision-oriented framework" (Levin 1995, p.383). All aspects of the program were costed, including TRT time, facilitator costs, materials, venue hire, and administration.

The tabulated cost data was provided to ACER at the completion of the evaluation period. The Department project team were interviewed to discuss their costings and clarify any uncertainties.

The cost per student per year was calculated on the cost estimates of one group of 35 teachers with an average class size of 25 students, equating to 875 students per year.

2.9 Timeline

The Thinking Maths evaluation timeline is presented in Table 4. There was no substantive deviation from the original proposed timeline.

Table 4. Thinking Maths evaluation timeline

Date	Activity	Team
2016		
Jun-Nov	Thinking Maths Pilot: observations of the five PL days	ACER
Sep	PATMaths student achievement (Weeks 7-10 Term 3): pre-test	SA Department
Sep-Oct	School recruitment: Expression of Interest (EOI), nomination of teachers, class-lists submitted	SA Department
2017		
Feb	Thinking Maths Evaluation Briefing event	ACER/The Department
Feb	Teacher and Student pre-survey administration period	ACER
Mar	Randomisation of schools into control and intervention groups	ACER
Mar-Sep	Thinking Maths program delivery period: 5 PL days, support, classroom implementation	SA Department
Mar-Sep	PL Feedback Form administration period: Intervention teachers submit an online form at the end of each training day	ACER
Sep	PATMaths student achievement (Weeks 7-10 Term 3): post-test	SA Department
Sep-Nov	Teacher and Student post-survey administration period	ACER
Nov-Dec	Preliminary data linkage, cleaning and analysis	ACER
Dec	School Reports sent to participating teachers	ACER
Dec	PATMaths 2016 & 2017 student data provided to ACER	SA Department
2018		
Jan-Jun	Outcomes analysis undertaken; Impact and process evaluation data analysed; Draft report prepared	ACER
Sep	Final E4L/ACER Evaluation Report released	E4L/ACER

3 Impact evaluation

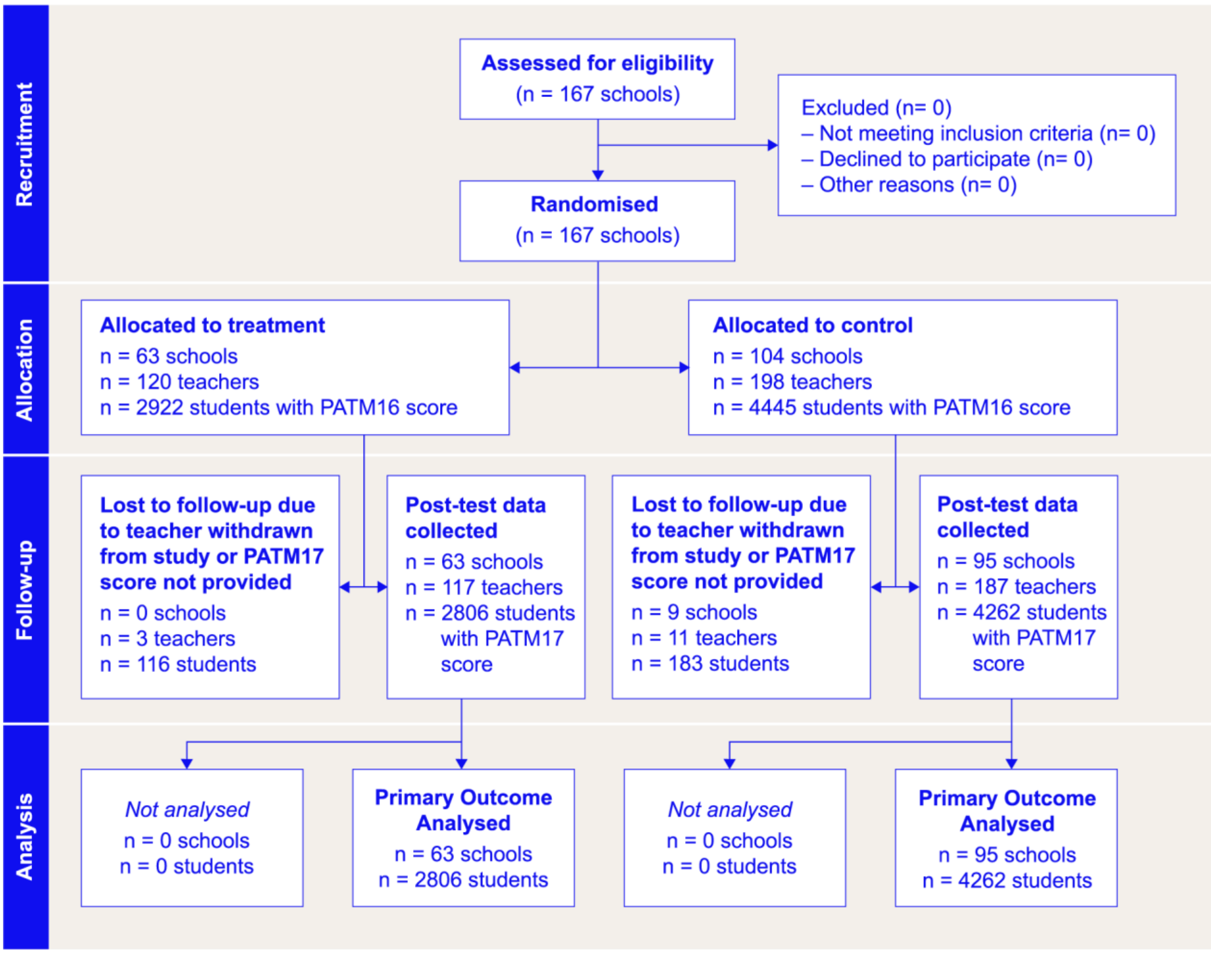
3.1 Participants

The success of recruiting schools, engaging participants and maintaining their engagement throughout the eight-month evaluation period was a result of the strong working relationship between ACER and the Department, and their commitment in supporting a high-quality evaluation, as summarised in the following counts.

- 167 schools were recruited, 158 participated
- 318 teachers were nominated, 304 were in the final analysis
- 7367 students were in the target classes, 7068 were in the final analysis

Figure 4 shows a CONSORT diagram of participant and school flow through the trial (Moher et al. 2010). Within the 167 schools, 318 mathematics teachers and their classes were nominated for participation, however early in the evaluation, 14 teachers withdrew from the program due to role/class reassignment or long-service-leave. This included a small number of schools in the control group that made the decision to withdraw from the program because the delayed timing competed with other planned activities. In these schools, the PATMaths student achievement data (the primary outcome) was not provided. In addition, student PATMaths data was unable to be matched for one class in the treatment group.

Figure 4. Participant flow CONSORT diagram



The CONSORT diagram also shows the unequal randomisation of schools, resulting in a larger control group to accommodate the strong interest and any control group attrition without the loss of statistical power. Of the students in the 304 evaluation classrooms across 158 schools, 7068 students had pre-test (PATM16) and post-test (PATM17) mathematics achievement scores. This group was used in the analyses on an intent-to-treat basis.

School characteristics

The Department received 167 expressions of interest from South Australian government schools to participate in the Thinking Maths project. This is 32% of government schools in the state. The fact that these schools expressed their interest, however, potentially made them different from schools that did not express their interest. In order to assess the representativeness of the participating sample, a non-response bias analysis was undertaken using ACER's database of all schools in South Australia (August 2015), to compare recruited schools with all other non-recruited government schools. Table 5 presents a comparison of school type, location, and socio economic index (SEIFA: Socio-Economic Indexes for Areas).

Table 5. Characteristics of recruited schools in comparison to non-recruited government schools in South Australia

SA government schools (N=523)	Recruited through EOI (n=167)		% Non-recruit	Bias	Chi-square p-value
	% Withdrawn	% Participated			
School type	n=9	n=158	n=356		0.002
Primary (K-7)	88.9	62.0	72.5	-10.4	
Combined (K-12)	0.0	16.5	18.3	-1.8	
Secondary (8-12)	11.1	21.5	9.3	12.2	
School location	n=9	n=158	n=356		0.001
Metro	55.6	63.3	43.8	19.5	
Rural	44.4	29.7	44.4	-14.6	
Remote	0.0	7.0	11.8	-4.8	
SEIFA index	n=9	n=158	n=356		0.053
Low (1-3)	55.6	31.8	43.8	-12.0	
Mid (4-7)	33.3	38.2	35.7	2.5	
High (8-10)	11.1	29.9	20.5	9.4	

Based on a comparison of the potential for bias among recruited (n=167) and non-recruited (n=356) schools, there were statistically significant differences in the Chi-squared test for school type ($p<0.05$) and location ($p<0.05$), but not SEIFA ($p>0.05$). The bias analysis indicates that participating schools (n=158) were under-represented in non-metro Primary settings and over represented in Secondary metro settings. The interpretation and generalisability of findings should be viewed with this in mind.

A baseline comparison of the randomly allocated intervention (n=63) and control (n=95) schools across a number of characteristics is provided in Table 6. The similar distribution, within ±10% bias, of schools by type, location, SES, size, and Aboriginal and Torres Strait Islander proportion, suggests that the profiles of the two groups of schools were comparable at baseline. This was supported by the results of the chi-tests and t-tests, which showed no significant difference ($p<0.05$). On this basis, schools randomly allocated to the control and intervention groups were considered comparable on these observable characteristics.

Table 6. Baseline comparison of school-level characteristics

School characteristics	Intervention Group		Control Group		Bias	p-value
School-level (categorical)	n/N (missing)	%	n/N (missing)	%		chi-square
School type						0.820
Primary	40/63 (0)	63.5	58/95(9)	61.1	2.4	
Combined (K-12)	11/63 (0)	17.5	15/95(9)	15.8	1.7	
Secondary	12/63 (0)	19.0	22/95(9)	23.2	-4.1	
School location						0.271
Metro	40/63 (0)	63.5	60/95(9)	63.2	0.3	
Rural	21/63 (0)	33.3	26/95(9)	27.4	6.0	
Remote	2/63 (0)	3.2	9/95(9)	9.5	-6.3	
SEIFA index						0.353
Low (1-3)	19/63 (0)	30.2	31/95(9)	32.6	-2.5	
Mid (4-7)	28/63 (0)	44.4	32/95(9)	33.7	9.2	
High (8-10)	16/63 (0)	25.4	32/95(9)	33.7	-6.7	
School-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)		t-test
School size	63 (0)	468.8(325.0)	95(9)	522.3(378.0)	-10.0	0.359
Aboriginal and Torres Strait Islander %	63 (0)	4.7(5.6)	95(9)	5.8(6.7)	1.1	0.309

Teacher characteristics

Teachers in recruited schools were randomised at the school-level into either the intervention (n=117 teachers) or control (n=187 teachers) groups, following their completion of the Teacher Survey, which included a number of background items. A baseline comparison of the characteristics of these teachers is presented in Table 7. Teachers, on average, had 12 years of teaching experience and two-thirds of teachers were female. Most teachers (93%) had at least a Bachelor's degree in education, mainly being in the field of Primary (45%) or Secondary (29%). While two in five teachers had studied mathematics post-schooling, only one in five identified it as their main area of specialisation. Two-thirds of the participating mathematics classes were Year 7 or Year 8.

Table 7. Baseline comparison of teacher-level characteristics

Teacher characteristic	Intervention Group		Control Group		Bias	p-value
Teacher-level (categorical)	n/N (missing)	%	n/N (missing)	%		chi-squared
Gender (female teachers)	80/117 (0)	68.4	120/187 (0)	64.2	4.2	0.452
Highest level in mathematics studied						0.017
Year 12 or below	62/117 (0)	53.0	122/187 (0)	65.2	-12.2	
Diploma or Certificate	27/117 (0)	23.1	21/187 (0)	11.2	11.8	
Bachelor or above	28/117 (0)	23.9	44/187 (0)	23.5	0.4	
Highest teaching qualification completed						0.032
Diploma or Certificate	14/117 (0)	12.0	8/187 (0)	4.3	7.7	
Bachelor degree	91/117 (0)	77.8	163/187 (0)	87.2	-9.4	
Master's Degree	12/117 (0)	10.3	16/187 (0)	8.6	1.7	
Field of teaching qualification						0.183
Primary or below	60/117 (0)	51.3	83/187 (0)	44.4	6.9	
Middle-School	32/117 (0)	27.4	46/187 (0)	24.6	2.8	
Secondary	25/117 (0)	21.4	58/187 (0)	31.0	-9.6	
Mathematics specialisation	25/117 (0)	21.4	43/187 (0)	23.0		0.740
Teacher-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)		t-test
Years of teaching experience	117 (0)	13.73 (11.76)	187 (0)	10.09 (9.46)	3.6	0.003
Year-level of the mathematics class	117 (0)	7.12 (0.94)	187 (0)	7.12 (1.05)	0.0	0.987
Teacher professional identity and self-efficacy	117 (0)	4.21(0.60)	182 (5)	4.22 (0.65)	-0.01	0.059
Teacher pedagogical and content knowledge	117 (0)	3.78 (0.52)	182 (5)	3.66 (0.53)	0.12	0.061
Teacher beliefs about mathematics learning	117 (0)	3.60 (0.49)	182 (5)	3.47 (0.56)	0.13	0.092

There were some significant differences ($p<0.05$) in the characteristics of teachers in the intervention and the control groups. Thinking Maths teachers had studied mathematics to a higher level and had more teaching experience, but fewer had a Bachelor Degree. However, there was no significant difference between the treatment and control groups at baseline in the teachers' beliefs about mathematics learning ($p=0.09$), teachers' professional identity and self-efficacy ($p=0.06$), and teachers' pedagogical content knowledge ($p=0.06$).

Student characteristics

A baseline comparison of the characteristics of students allocated to the Thinking Maths intervention (n=2806) or the control group (n=4262) is presented in Table 8. These were the students of the one or two nominated teachers from each of the 158 participating schools, for which data was available. The mean age of students was 12.1 years, with 91% of students in the targeted grades of Years 6-9.

Table 8. Baseline comparison of student-level characteristics (in primary analysis)

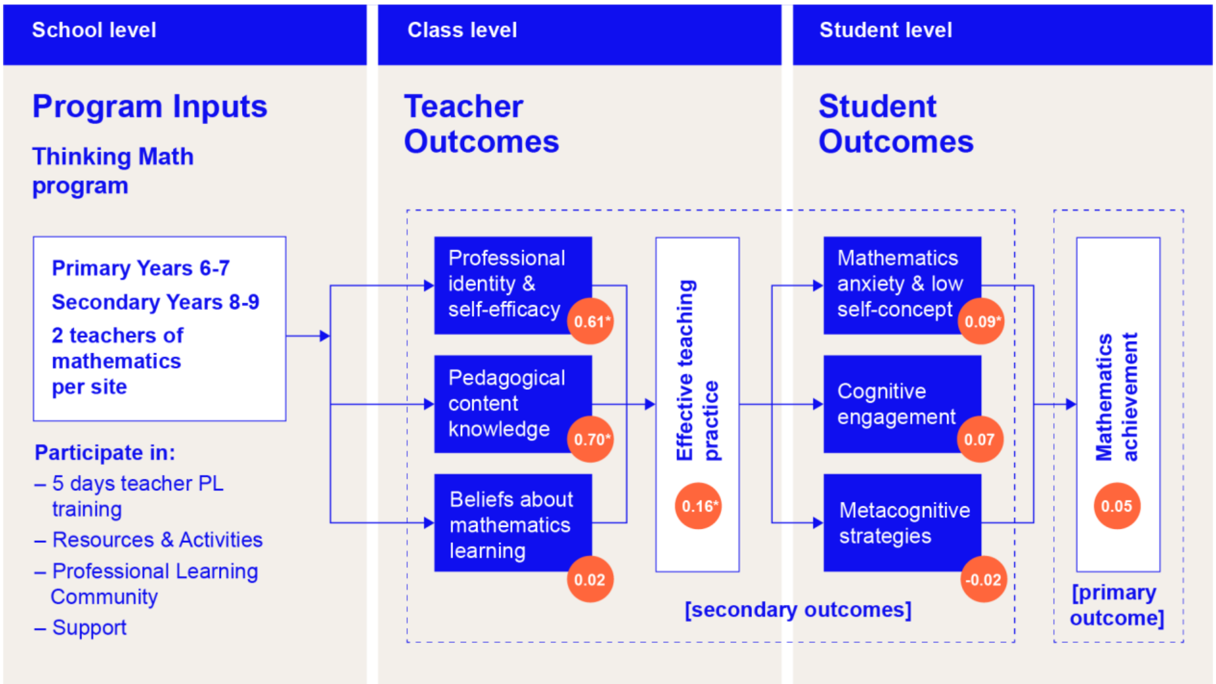
Student characteristic	Intervention Group		Control Group		Bias	p-value
Student-level (categorical)	n/N (missing)	%	n/N (missing)	%		t-test
Gender (female students)	1409/2806 (0)	50.2	2122/4262 (0)	49.8	0.4	0.727
Disability	218/2806 (0)	7.8	300/4262 (0)	7.0	0.7	0.249
Aboriginal and Torres Strait Islander background	124/2806 (0)	4.4	221/4262 (0)	5.2	-0.8	0.144
School Card holder	293/2806 (0)	10.4	373/4262 (0)	8.8	1.7	0.017
Year Level						0.431
Year 5	219/2806 (0)	7.8	343/4262 (0)	8.0	-0.2	
Year 6	856/2806 (0)	30.5	1284/4262 (0)	30.1	0.4	
Year 7	965/2806 (0)	34.4	1346/4262 (0)	31.6	2.8	
Year 8	479/2806 (0)	17.1	917/4262 (0)	21.5	-4.4	
Year 9	271/2806 (0)	9.7	314/4262 (0)	7.4	2.3	
Year 10	16/2806 (0)	0.6	58/4262 (0)	1.4	-0.8	
Student-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)		t-test
Age (as at 1/1/2017)	2806 (0)	12.1 (1.2)	4262 (0)	12.1 (1.2)	0.0	0.511
Effective teaching and learning	2450 (356)	3.5 (0.7)	3484 (778)	3.4 (0.7)	0.1	0.130
Maths anxiety and low self-efficacy	2450 (356)	2.7 (0.8)	3484 (778)	2.7 (0.8)	0.0	0.079
Cognitive engagement	2450 (356)	3.9 (0.7)	3484 (778)	3.9 (0.7)	0.0	0.237
Metacognitive strategies	2450 (356)	3.5 (0.7)	3484 (778)	3.5 (0.7)	0.0	0.421
PATMaths 2016 score (average prior attainment)	2806 (0)	124.7 (11.1)	4262 (0)	125.1 (11.6)	0.4	0.651

There was no significant difference ($p<0.05$) at baseline in the intervention and control groups in gender, disability, or Aboriginal and Torres Strait Islander background. There was a small difference ($p=0.02$) in the number of School Card holders, with greater representation in the intervention group. There was no significant difference ($p<0.05$) in the treatment and control groups at baseline in average prior attainment of students (PATMath16 scores), nor in students' beliefs about effective teaching and learning, their maths anxiety and low self-efficacy, their cognitive engagement, and their metacognitive strategies.

3.2 Outcomes evaluation results

The outcome evaluation was guided by four research questions that emerged from the logic model. Figure 5 presents the logic model with the inclusion of the effect sizes as a measure of the impact of the Thinking Maths program, all of which is described in detail through the research questions in this section.

Figure 5. Outcome evaluation effect size results



* indicates statistically significant effect (p<0.05)

Primary outcomes

1. Did the Thinking Maths program enable middle-school students to improve their mathematics achievement above typical learning growth?

The primary aim of the analysis was to assess whether the Thinking Maths intervention had an impact on students' mathematics achievement, as measured by the post-intervention PATM17 test scores, controlling for prior attainment in the form of PATM16 scores. Table 9 shows the results of the whole-group analysis and subgroup analyses, conducted on an intention-to-treat basis. See Addendum Table 1A for the results of the model specification, with prior attainment and the treatment indicator as the main explanatory variables, together with school-level clustering to account for students nested in schools. For the reduced subgroup samples, results should be interpreted in light of the reduction in power the approach brings.

After controlling for prior achievement, the multi-level statistical analysis involving the **full sample** (n=7068 students in 158 schools), showed that the Thinking Maths program had a small impact on the treatment group. Table 9 shows that the mean post-test mathematics achievement score for students above the typical learning gain of 3.65 score points was, on average, 0.38 score points higher in the treatment group than the score of students in the control group.

The effect size of ES=0.05 (CI: 0.00 – 0.10, p = 0.38) was not statistically significant for the full sample. This translates to one month of additional learning based on Evidence for Learning's conversion to months progress (see Appendix A).

Table 9. Effect of the Thinking Maths program on student primary outcomes

Primary outcome	Intervention Group		Control Group		Effect size (HLM)		
	n (miss)	Mean (SD) [95% CI]	N (miss)	Mean (SD) [95% CI]	ES	95% CI	p-value
Primary analysis							
PATMath17	2806 (116)	128.74 (11.46) [128.31-129.16]	4262 (183)	128.51 (12.24) [128.15-128.88]	**0.05**	0.00 – 0.10	0.38
Subgroup analyses							
Primary Years 5-7	2040 (87)	128.28 (11.87) [127.76-128.79]	2973 (92)	127.22 (12.41) [126.77-127.66]	**0.14**	0.08 – 0.19	0.05
Secondary Years 8-10	766 (29)	129.96 (10.17) [129.24-130.68]	1289 (91)	131.51 (11.30) [130.89-132.13]	**-0.16**	-0.25 – -0.07	0.05
School Card holders	293 (16)	125.39 (10.33) [124.21-126.58]	373 (28)	124.76 (11.14) [123.63-125.89]	**0.11**	-0.04 – 0.27	0.21

The subgroup analysis of **Primary Years 5-7** (n=5013 students in 119 schools), shows that the Thinking Maths program had a small positive impact on mathematics for this group in the treatment compared with those in the control group. The mean post-test mathematics achievement score was 1.02 score points higher in the treatment group than the score of students in the control group. The significant difference was equivalent to an effect of ES=0.14 (CI: 0.08 – 0.19, p = 0.05) or two months additional learning.

The subgroup analysis of **Secondary Years 8-10** (n = 2055 students in 56 schools), shows a small negative effect on mathematics for this group of students. The post-test mathematics achievement score was, on average, 1.18 score points lower in the treatment group than the score of students in the control group. The significant difference was equivalent to an effect of ES=-0.16 (CI: -0.25 – -0.07, p = 0.05) or two fewer months of learning progress.

The subgroup analysis of **School Card holders** (n=666 students in 118 schools), shows a small positive effect on mathematics for this group of students. The post-test mathematics achievement score was, on average, 0.83 score points higher in the treatment group than the score of students in the control group. The difference was equivalent to an effect of ES=0.11 (CI: -0.04 – 0.27, p = 0.21) or one month of learning gain.

Secondary outcomes

The Thinking Maths program directly involved teachers and, through their mathematics teaching, indirectly involved the students in their classrooms. As such, it was anticipated that the level of impact of the program may be greatest on teachers and their teaching practices, but less-so on students and their mathematics achievement, particularly given the short timeframe. This is evident in the evaluation logic model in Figure 5, which shows a progressive reduction in effect size. Working back along the path of impact from the primary outcome, we address the research questions and report the following secondary outcomes.

2. Did Thinking Maths develop middle-school students as powerful learners of mathematics in terms of a) mathematics self-efficacy, b) cognitive engagement in learning, and c) metacognition?

The three secondary outcome measures, presented in Figure 5 and described in the methods section, were used to assess the impact of Thinking Maths on students' mathematics anxiety and low self-concept, their cognitive engagement, and their metacognitive strategies. Table 10 presents the main results and subgroup analyses, including the effect size of Thinking Maths on each outcome, together with its 95% confidence interval and p-value. See Addendum Table A3 for the subgroup model specifications, with the treatment indicator as the explanatory variable and controlling for clustering of teachers and students within schools.

Table 10. Impact on students' beliefs and attitudes

Student secondary outcomes	Intervention	Control	Effect size (HLM)		
	Mean (SD) [95% CI]	Mean (SD) [95% CI]	ES	95% CI	p
Secondary analysis n(miss)	2467 (339)	3484 (778)			
Mathematics anxiety and low self-concept (SASE)	2.77 (0.77) [2.74-2.80]	2.71 (0.81) [2.68-2.73]	**0.09**	0.04 – 0.14	0.04
Cognitive engagement (SCOG)	3.90 (0.69) [3.87-3.93]	3.84 (0.73) [3.82-3.87]	**0.07**	0.02 – 0.12	0.26
Metacognitive strategies (SMET)	3.43 (0.67) [3.40-3.46]	3.44 (0.70) [3.42-3.47]	-0.02	-0.08 – 0.03	0.63
Subgroup: Years 5-7 n(miss)	1899 (141)	2536 (437)			
Mathematics anxiety and low self-concept (SASE)	2.71 (0.79) [2.68-2.75]	2.65 (0.81) [2.62-2.68]	**0.07**	0.01 – 0.13	0.14
Cognitive engagement (SCOG)	3.96 (0.68) [3.93-3.99]	3.90 (0.71) [3.87-3.93]	**0.09**	0.03 – 0.15	0.20
Metacognitive strategies (SMET)	3.48 (0.68) [3.45-3.51]	3.51 (0.71) [3.48-3.53]	-0.03	-0.09 – 0.03	0.58
Subgroup: Years 8-10 n(miss)	568 (198)	948 (341)			
Mathematics anxiety and low self-concept (SASE)	2.98 (0.69) [2.93-3.04]	2.85 (0.77) [2.80-2.90]	**0.20**	0.10 – 0.31	0.01
Cognitive engagement (SCOG)	3.70 (0.69) [3.64-3.75]	3.69 (0.78) [3.64-3.74]	-0.01	-0.12 – 0.09	0.90
Metacognitive strategies (SMET)	3.26 (0.59) [3.21-3.31]	3.27 (0.67) [3.23-3.31]	-0.03	-0.13 – 0.08	0.72
Subgroup: School Card n(miss)	243 (50)	275 (98)			
Mathematics anxiety and low self-concept (SASE)	2.91 (0.69) [2.83-3.00]	2.84 (0.78) [2.75-2.94]	**0.10**	-0.08 – 0.27	0.26
Cognitive engagement (SCOG)	3.68 (0.67) [3.59-3.76]	3.74 (0.78) [3.65-3.83]	-0.04	-0.21 – 0.14	0.73
Metacognitive strategies (SMET)	3.32 (0.59) [3.25-3.40]	3.37 (0.70) [3.29-3.46]	**-0.05**	-0.22 – 0.12	0.61

Students' mathematics anxiety and low self-concept: The main result shows that the post-test SASE score (reliability α=0.89) of students in the intervention group was, on average, 0.07 scale units higher than the score of students in the control group. This very small significant increase in mathematics anxiety was equivalent to an effect of ES=0.09 (CI: -0.14 – 0.32, p<0.05). For example, 10% of Thinking Maths students, compared to 8% of business-as-usual students strongly agreed that they *get nervous doing maths problems*, and 4% of students in both cohorts strongly agreed that they were *just not good at maths*.

The subgroup results indicate that Primary students in the treatment group were 0.05 scale units higher than the score of their peers in the control group. In comparison, Secondary students in the treatment group were 0.15 scale units higher than the score of their peers in the control group. The effect size for Primary students was ES=0.07 (CI: 0.01 – 0.13, p=0.14), compared to ES=0.20 (CI: 0.10 – 0.31, p=0.01) for Secondary students. School Card holders in the treatment group were, 0.07 scale units higher than the score of their peers in the control group, equivalent to an effect of ES=0.10 (CI: -0.08 – 0.27, p=0.26).

Students' cognitive engagement: The result shows that the post-test SCOG score (reliability α=0.81) of students in the intervention group was 0.05 scale units higher than the score of students in the control group. This very small increase in students' cognitive engagement in mathematics was equivalent to an effect of ES=0.07 (CI: -0.16 – 0.30, p=0.26). For example, similar numbers of students (26% control; 28% treatment) strongly agreed to the statement: *I know what my teacher expects me to do,* while half the students (49% control; 50% treatment) strongly agreed that, *My teacher believes all students can be good at maths*.

The subgroup results indicate that Primary students in the treatment group were 0.06 scale units higher than the score of their peers in the control group. In comparison, Secondary students in the treatment group were 0.01 scale units lower than the score of their peers in the control group. The effect size for Primary students was ES=0.09 (CI: 0.03 – 0.15, p=0.20), compared to ES=-0.01 (CI: -0.12 – 0.09, p=0.90) for Secondary students. School Card holders in the treatment group were, on average, 0.03 scale units lower than the score of their peers in the control group, equivalent to an effect of ES=-0.04 (CI: -0.21 – 0.14, p=0.73).

Students' metacognitive strategies: The result shows that the post-test SMET score (reliability α=0.76) of students in the intervention group was, on average, 0.02 scale units lower than the score of students in the control group. Equivalent to an effect size of ES=-0.02 (CI: -0.26 – 0.21, p=0.63), there was no difference in students' metacognitive strategies between the intervention and control groups. For example, there was no difference in the control or treatment groups in the number of students who strongly agreed that, *I check my maths school work for mistakes* (11%), or *try to connect the things I am learning in maths with what I already know* (20%).

The subgroup results indicate that Primary students in the treatment group were 0.02 scale units lower than the score of their peers in the control group. Similarly, Secondary students in the treatment group were 0.02 scale units lower than the score of their peers in the control group. The effect size for Primary students was ES=-0.03 (CI: -0.09 – 0.03, p=0.58), compared to ES=-0.03 (CI: -0.13 – 0.08, p=0.72) for Secondary students. School Card holders in the treatment group were 0.03 scale units lower than the score of their peers in the control group, equivalent to an effect of ES=-0.05 (CI: -0.22 – 0.12, p=0.61).

Based on these results and in response to Question 2, Thinking Maths:

- marginally raised students' mathematics anxiety and low self-concept, but more so in the Secondary Years 8-10 compared to the Primary Years 5-7.
- marginally raised students' cognitive engagement overall, but more so in Primary Years 5-7 and not Secondary Years 8-10.
- had no impact on students' metacognitive strategies overall, nor in Primary Years 5-7 or Secondary Years 8-10.
- marginally raised mathematics anxiety in School Card holding students, with no effect on their cognitive engagement and a marginal reduction in metacognitive strategies.

3. Did Thinking Maths shift teachers' mathematics teaching practice towards a more inclusive, student-centred learning approach? Did changes in teachers' practices due to Thinking Maths, influence students' mathematics outcomes?

Rather than having teachers assess a shift in their own mathematics teaching practice, a more sensitive approach was taken by having the secondary outcome assessed by students. The results are presented in Table 11.

Table 11. Impact on teaching practice according to students

Secondary outcome: Learning through effective teaching practice (SETL)	Intervention Group		Control Group		Effect size (HLM)		
	N (miss)	Mean (SD) [95% CI]	N (miss)	Mean (SD) [95% CI]	ES	95% CI	p
Full sample	2467 (455)	3.55 (0.66) [3.52-3.58]	3484 (961)	3.43 (0.73) [3.40-3.45]	**0.16**	0.11 – 0.22	0.01
Subgroup							
Years 5-7	1899 (199)	3.60 (0.65) [3.57-3.63]	2536 (437)	3.48 (0.70) [3.45-3.51]	**0.18**	0.12 – 0.24	0.01
Years 8-10	568 (256)	3.38 (0.65) [3.33-3.44]	948 (341)	3.28 (0.77) [3.23-3.33]	**0.09**	-0.01 – 0.20	0.42
School Card	243 (50)	3.38 (0.66) [3.29-3.46]	275 (98)	3.42 (0.74) [3.33-3.51]	-0.04	-0.21 – 0.13	0.72

Students' learning through effective teaching practice: The result shows that the mean post-test SETL score (reliability α=0.89) of students in the intervention group was 0.12 scale units higher than the score of students in the control group. This significant difference in improved teaching practices was equivalent to an effect of ES=0.16 (CI: -0.07 – 0.40, $p<0.05$). For example, 29% of business-as-usual students reported that *My teacher asks me to explain my answers* almost every lesson, whereas the same was true of 35% of Thinking Maths students. Similarly, in response to the statement, *My teacher asks me or my classmates to present our mathematical thinking*, 23% of business-as-usual students, compared to 27% of Thinking Maths students, had teachers who did this almost every lesson. This evidence, of a more student-centred teaching approach as reported by the students themselves, indicates that Thinking Maths did shift teachers' mathematics teaching practices that, in turn, may have influenced students' mathematics outcomes.

The subgroup results indicate that Primary students in the treatment group were, on average, 0.12 scale units higher than the score of their peers in the control group. In comparison, Secondary students in the treatment group were 0.07 scale units higher than the score of their peers in the control group. The effect size for Primary students was ES=0.18 (CI: 0.12 – 0.24, p=0.01), compared to ES=0.09 (CI: -0.01 – 0.20, p=0.42) for Secondary students. School Card holders in the treatment group were 0.03 scale units lower than the score of their peers in the control group, equivalent to an effect of ES=-0.04 (CI: -0.21 – 0.13, p=0.72).

Based on these results and in response to Question 3, Thinking Maths indirectly:

- improved students' learning through effective teaching practice, but more so in the Primary Years 5-7 compared to the Secondary Years 8-10.
- had little impact on School Card holding students with regard to their views on effective teaching practice.

4. Did Thinking Maths build the capacity of teachers in terms of a) pedagogical and content knowledge, b) beliefs about mathematics teaching and learning, and c) professional identity?

The Thinking Maths program was designed to have impact on student outcomes through teachers' engagement in professional learning, and subsequent improvement in their capacity to teach mathematics in the classroom. Being the participants with direct exposure to the program (see Figure 5), it was expected that any differences in outcomes would be greatest in the teachers. Three secondary outcome measures were used to assess the impact of Thinking Maths on teachers' professional identity and self-efficacy, their pedagogical and content knowledge, and their beliefs about mathematics learning. Table 12 presents the results. The model specifications, with the ITT indicator as the explanatory variable and controlling for clustering of teachers within schools, are given in Addendum Table A2.

Teachers' professional identity and self-efficacy: The result shows that the mean post-test TPID score (reliability α=0.89) of teachers in the treatment group was 0.33 scale units higher than the score of teachers in the control group. This significant difference in teachers' professional identity and self-efficacy was equivalent to an effect of ES=0.61 (CI: 0.37 – 0.85, p<0.01). For example, post-survey results found that 9% of business-as-usual teachers (control) compared to 23% of Thinking Maths teachers (treatment) felt that they could, to a great extent, *motivate students who show low interest in maths*. Similarly, while 18% of control-group teachers could, to a great extent, *create opportunities for all students to experience productive struggle*, the same was true for 34% of Thinking Maths teachers.

The subgroup results indicate that Primary teachers in the treatment group were 0.44 scale units higher than the score of their peers in the control group. In comparison, Secondary teachers in the treatment group were 0.06 scale units higher than the score of their peers in the control group. The effect size for Primary teachers was ES=0.80 (CI: 0.51 – 1.09, p<0.01), compared to ES=0.13 (CI: -0.30 – 0.55, p=0.54) for Secondary teachers.

Teachers' pedagogical and content knowledge: The result shows that the mean post-test TPCK (reliability α=0.91) score of teachers in the treatment group was, on average, 0.37 scale units higher than the score of teachers in the control group. This significant difference in knowledge was equivalent to an effect of ES=0.70 (CI: 0.47 – 0.94, p<0.01). For example, teachers were asked how confident they were in *using questioning to develop students' conceptual understanding*. Post-survey results found that 14% of business-as-usual teachers, compared to 35% of Thinking Maths teachers, were confident to a great extent.

The subgroup results indicate that Primary teachers in the treatment group were 0.46 scale units higher than the score of their peers in the control group. In comparison, Secondary teachers in the treatment group were 0.19 scale units higher than the score of their peers in the control group. The effect size for Primary teachers was ES=0.85 (CI: 0.57 – 1.14, p<0.01), compared to ES=0.38 (CI: -0.05 – 0.81, p=0.050) for Secondary teachers.

Table 12. Impact on teacher outcomes

Teacher secondary outcomes	**Intervention**	**Control**	**Effect size (HLM)**		
	Mean (SD) [95% CI]	**Mean (SD) [95% CI]**	**ES**	**95% CI**	**p**
Secondary analysis n(miss)	117 (0)	182 (5)			
Professional identity and self-efficacy (TPID)	4.05 (0.51) [3.96-4.15]	3.72 (0.57) [3.64-3.80]	**0.61**	0.37 – 0.85	0.00
Pedagogical content knowledge (TPCK)	4.02 (0.43) [3.94-4.10]	3.64 (0.59) [3.56-3.73]	**0.70**	0.47 – 0.94	0.00
Beliefs about mathematics learning (TBEL)	4.26 (0.59) [4.15-4.36]	4.25 (0.63) [4.16-4.34]	0.02	-0.21 – 0.25	0.87
Subgroup: Years 5-7 n(miss)	84 (0)	124 (3)			
Professional identity and self-efficacy (TPID)	4.15 (0.50) [4.04-4.26]	3.71 (0.58) [3.60-3.81]	**0.80**	0.51 – 1.09	0.00
Pedagogical content knowledge (TPCK)	4.04 (0.42) [3.95-4.13]	3.58 (0.60) [3.47-3.68]	**0.85**	0.57 – 1.14	0.00
Beliefs about mathematics learning (TBEL)	4.25 (0.59) [4.12-4.37]	4.30 (0.58) [4.20-4.40]	**-0.08**	-0.35 – 0.20	0.62
Subgroup: Years 8-10 n(miss)	33 (0)	58 (2)			
Professional identity and self-efficacy (TPID)	3.81 (0.47) [3.64-3.97]	3.74 (0.53) [3.60-3.88]	**0.13**	-0.30 – 0.55	0.54
Pedagogical content knowledge (TPCK)	3.98 (0.45) [3.82-4.14]	3.78 (0.53) [3.64-3.93]	**0.38**	-0.05 – 0.81	0.05
Beliefs about mathematics learning (TBEL)	4.28 (0.60) [4.07-4.49]	4.14 (0.73) [3.95-4.33]	**0.19**	-0.23 – 0.62	0.37

Teachers' beliefs about mathematics learning: The results show that the post-test TBEL (reliability α=0.68) scores of teachers in the treatment group was, on average, 0.01 scale units higher than the score of teachers in the control group. This difference in teachers' beliefs was equivalent to an effect of ES=0.2 (CI: -0.21 – 0.25, p=0.87). For example, 54% of business-as-usual teachers compared to 56% of Thinking Maths teachers strongly agreed that they *deeply believe that everyone can learn maths*.

The subgroup results indicate that Primary teachers in the treatment group were 0.04 scale units lower than the score of their peers in the control group. In comparison, Secondary teachers in the treatment group were 0.13 scale units higher than the score of their peers in the control group. The effect size for Primary teachers was ES=-0.08 (CI: -0.35 – 0.20, p=0.62), compared to ES=0.19 (CI: -0.23 – 0.62, p=0.37) for Secondary teachers.

Based on these results and in response to Question 4, Thinking Maths:

- improved teachers' professional identity and self-efficacy to a moderate effect, but more so for mathematics teachers in Primary Years 5-7 who reported a large effect, compared to teachers in Secondary Years 8-10 who reported a small effect.
- improved teachers' pedagogical content knowledge to a moderate effect, with a large effect experienced by mathematics teachers in Primary Years 5-7 and a moderate effect for teachers in Secondary Years 8-10.
- had no impact on teachers' beliefs about mathematics learning overall, although a small improvement for teachers in Secondary Years 8-10 was found.

Additional results

The results suggest that Thinking Maths did enable middle-school students, particularly those in the Primary years, to improve their mathematics achievement above typical learning growth. To gain a deeper understanding of the results in Primary and Secondary contexts, additional analysis was undertaken at the Year level.

Figure 6 presents a comparison of student pre (2016) and post (2017) PAT Maths scores in the control and treatment groups, after applying a baseline correction to remove the minor differences in pre-scores. Given that the target cohort is Years 6 to 9 students, the Year 5 and Year 10 student results are not presented due to insufficient sample size. Learning growth can be seen by the increase in pre-scores for each Year level group. At the transition from Primary school into Secondary school between Year 7 and Year 8, little growth is evident. The slope of each line represents one year's average learning gain for that cohort and the gap between the post scores of control and treatment students is interpreted as the impact of Thinking Maths. It shows the clear gains made in Years 6 and 7 and no difference in Year 8. Therefore, it is the decline in the Year 9 treatment group that is largely responsible for the fewer months of learning progress reported in the Secondary Years sub-group analysis. While learning gain is known to flatten with Year level, it is unclear why the Year 9 treatment cohort did not show at least some learning gain typical of other Year 9 students. Even if the Year 9 data is taken as anomalous, these results still suggest that Thinking Maths is more effective in the Primary years.

A further analysis was undertaken to test for any relationships between student and teacher outcomes in Primary and Secondary contexts using Pearson correlation. To do so, student-level data aggregated to the class-level was matched to teacher-level data.

Table 13 presents a correlation matrix of students and teachers in Primary (Years 5-7) and Secondary (Years 8-10) settings. Significant ($p<0.05$) correlations between teacher and student post-outcomes in the intervention group were mainly associated with the Primary context. It suggests that students were more likely to experience effective teaching practices if their teacher had strong beliefs about mathematics learning ($r=0.27$) and professional identity and self-efficacy ($r=0.25$). Similarly, students were more likely to have lower mathematics anxiety and higher self-concept ($r_{TPID}=-0.37$; $r_{TPCK}=-0.36$), greater cognitive engagement ($r_{TPID}=0.36$; $r_{TPCK}=0.34$), and metacognitive strategies ($r_{TPID}=0.38$; $r_{TPCK}=0.28$) if their mathematics teacher had strong professional identity and self-efficacy, as well as strong pedagogical and content knowledge. Students were more likely to gain higher post attainment mathematics outcomes in Primary classes where the teacher reported strong pedagogical content knowledge ($r=0.23$), and in Secondary classes where the teacher reported high professional identity & self-efficacy ($r=0.40$).

Figure 6. Baseline-corrected pre-post PAT Maths scores by Year level

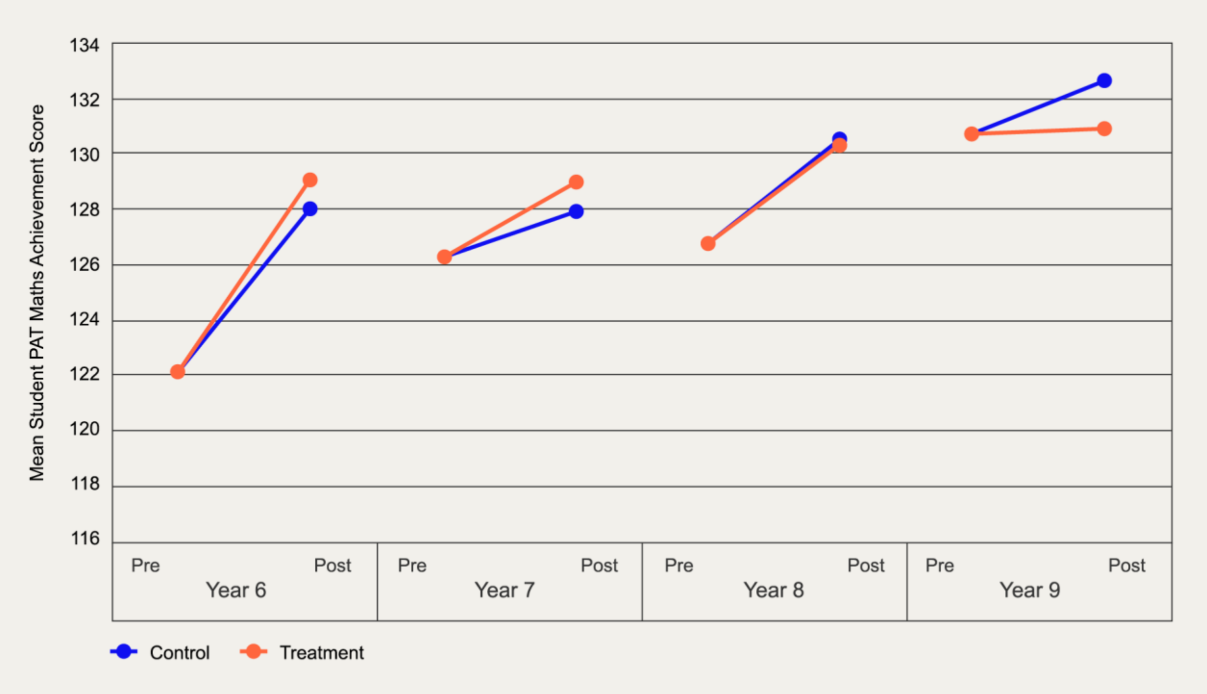


Table 13. Relationships between teacher and student outcomes in Primary and Secondary contexts in the treatment group

		TEACHER OUTCOMES					
Cohorts (n teachers/classes)		Prim (77)	Sec (31)	Prim (76)	Sec (30)	Prim (77)	Sec (31)
Pearson correlation (r)		Beliefs about maths learning (TBEL)		Professional identity & self-efficacy (TPID)		Pedagogical & content knowledge (TPCK)	
STUDENT OUTCOMES	Learning through effective teaching practice (SETL)	**0.27***	0.13	**0.25***	0.06	0.19	0.24
	Mathematics anxiety and low self-concept (SASE)	-0.16	-0.09	**-0.37****	0.02	**-0.36****	0.08
	Cognitive engagement (SCOG)	0.19	0.13	**0.36****	0.02	**0.34****	0.24
	Metacognitive strategies (SMET)	0.13	-0.06	**0.38****	0.03	**0.28***	0.14
	Post attainment (PATM17)	0.1	0.05	0.18	**0.40***	**0.23***	0.15

p<0.05; ** p<0.01. Note that SASE is negatively worded, so a negative correlation is desirable. The colour gradient of cells from red to blue indicate negative/undesirable or positive/desirable associations. Darker cells have stronger associations.

3.3 Cost

5. How cost-effective is the Thinking Maths program?

The cost estimates to set-up and run Thinking Maths, as provided by the Department (the program developers and funders), are detailed in Table 14 and Table 15.

These estimates were based on the costs associated with running the five-day professional learning program for one group of 35 teachers. It should be noted that all costs were borne by the Department, with no cost to the schools.

Table 14. Program set-up costs

Program set-up			Subtotal
Materials (based on 35 participants around 7 tables)			$1510
Pattern blocks (for 35)	$200		
2cm wooden blocks (7)	$140		
Dice	$100		
Pencil cases (7)	$150		
Counters	$70		
Pan balances (7)	$350		
Misc maths equipment (for 35)	$100		
Pouches and storage bags	$200		
Storage boxes (5)	$100		
Laminating sheets	$100		
Equipment for Presentations			$1495
Clax trolley	$495		
Technology (computer, storage devices)	$1000		
Preparation			$3750
Preparation of program materials	$1500/day	2 presenters/ 1 day	$3000
Initial set-up of online platform	$1500/day	1 presenter/ 0.5 days	$750
Total Program Set-Up Cost			$6755

Table 15. Program running costs for one group of 35 teachers

Program delivery			Subtotal
Administrative tasks			$3600
Program scheduling, liaison with schools, venue hire, copying, etc.	$600	6 days (1 prior, 5 during)	$3600
Consultancy fees			$26250
Presentation preparation	$1500/day	2 presenters/ 2.5 days	$7500
Presentation of 5 professional learning days	$1500/day	2 presenters/ 5 days	$15000
Facilitation of online learning platform	$1500/day	1 presenter/ 2.5 days	$3750
Professional learning days			$5000
Venue hire	$100/day	x 5 days	$500
Catering (35)	$900/day	x 5 days	$4500
Professional learning day consumables			$160
Brenex Squares (35)	$20		
Card (35)	$30		
Tape	$30		
Material replacements	$80		
Participant/Teacher materials costs			$170
Reference book	$90		
Folder	$20		
Printing	$60		
Participant/Teacher Replacement (program conducted in school time)			$92750
TRT replacement (35)	$530 x 35 = $18550	x 5 days	
Total Program Delivery Cost			$127930

If the current facilitators were to continue to deliver the program, the start-up costs would not be applicable. Therefore, the ongoing cost per student per year to deliver Thinking Maths under the current approach is $146. These figures are based on delivering the Thinking Maths program to one group of 35 teachers each with a class of 25 students, resulting in the program reaching 875 students.

Table 16 shows the cost per teacher and per student for starting-up and running the Thinking Maths program. The once-off start-up cost of $6755 covers the purchase of re-usable materials and equipment to run the sessions. It would be the costs associated with setting up a new Thinking Maths delivery team. While the estimate below of $8 per student is based on one cohort of 875 students, this start-up cost in the evaluation was actually shared across four concurrent groups of 30 teachers, reaching approximately 3000 students – a small amount at approximately $2 per student per year.

Table 16. Start-up and delivery costs

Cost ($AUD)	per group (35 teachers)	per teacher	per student
Start-up costs (once-off)	$6755	$193	$8
5 TRT days replacement per year	$92750	$2650	$106
Delivery costs per year	$35180	$1005	$40
Total cost in first year	$134685	$3848	$154

Using Evidence for Learning's Cost Rating approach (Evidence for Learning, 2017), the approximate cost per student per year of implementing the intervention over three years was calculated, as shown in Table 17. Accordingly, the cost of the Thinking Maths program was estimated at $149 per student per year. This estimate includes training and materials ($1070 per teacher or $43 per student), and the significant cost of five TRT days replacement ($2650 per teacher or $106 per student). Estimates are based on training being delivered to a group of 35 teachers with an average class size of 25 students. This amount per student is rated as *very low (<$160)*, according to Evidence for Learning's Cost Rating guidelines (see Appendix A).

Table 17. Cost per year over multiple years ($AUD)

Number of years using program	Cumulative cost per student	Average cost per student per year \ (cumulative cost/number of years)
1 year	$154	$154
2 years	$300	$150
3 years	$446	$149

4 Process evaluation results

The purpose of the process evaluation was to identify features contributing to successful implementation and to understand participants' experiences of the intervention. This chapter presents the results of the process evaluation, based on observation, surveys and interviews with the Thinking Maths facilitators. The implementation of the program and the fidelity with which it was delivered are addressed through the process research questions.

4.1 Implementation

Implementation of the Thinking Maths program by the Department with four groups of approximately 30 teachers, occurred during an eight-month period from February to September 2017 in the school calendar year. Broadly speaking, this involved engaging teachers by conducting the five PL days, providing resources and activities, providing support, and running the online professional learning community. This section explores the implementation of these four elements – what was actually delivered to teachers and in schools – by examining the barriers and facilitators to successful implementation and how the program was received by participating teachers.

Elements of the program

1. What are the critical elements of the Thinking Maths program in terms of quality of delivery, fidelity and dosage?

The program-design was underpinned by well-established educational theory and used evidence-based research-informed strategies. Moreover, it was developed by local experts in the teaching of middle-school mathematics, with a thorough knowledge of the Australian Curriculum Mathematics and years of in-classroom experience. The developers of Thinking Maths were also the program facilitators, and accordingly, knew the program intimately. The facilitators were professional presenters, able to hold and maintain the engagement of participants throughout the five days of PL, in addition to providing professional support in-between PL days.

Professional learning days: Based on teacher ratings of the PL days, the quality of delivery was very high. Almost all teachers (98% on average) agreed or strongly agreed that the presenters were well organised and were engaging. Most teachers (97%) also agreed that the presenters were respectful and inclusive of different views, and encouraged active participation that supported learning. These results (see Addendum Figure A8) support the observations of the PL days conducted the previous year.

Almost all teachers (98%) reported to a moderate or great extent that the information was relevant and useful to their role, and that their understanding of maths teaching had improved because of the sessions. The majority of teachers (97%) felt that the activities and discussions supported their learning. Nine in 10 teachers felt more confident to a moderate or great extent about their capacity to make mathematics learning deeper (88%) and more engaging (90%). Most teachers (95%) were very motivated to learn more about improving students' mathematics outcomes, and also felt that the sessions, overall, met their needs and expectations. The reports of participating teachers about the professional learning (see Addendum Figure 9A) suggested that the quality of the program was very high.

Activities and resources: As part of the Thinking Maths program, teachers were expected to engage in a number of activities and use various mathematics resources, such as the set readings and reflective journaling (see Addendum Figure 10A). Teachers were asked which of the four Thinking Maths activities they had engaged in during the previous month (the 4-5 week periods between PL days). Three-quarters of teachers (76% on average) completed the readings, while 57% reported that they kept a reflective journal. Four in five teachers (80%) used the resources or materials provided by Thinking Maths, and three in five teachers (62%) shared evidence of classwork with colleagues based on something from Thinking Maths.

Professional learning community: Thinking Maths teachers were also expected to participate in the online professional learning community (Thinking Maths Moodle). Teachers were asked in what ways they had participated in the online community in the previous month. Results suggest that this element of the program was less effective (see Addendum Figure 11A). On average, two-thirds of teachers (67%) reported that they had logged into the Thinking Maths Moodle. Six in 10 teachers (59%) had downloaded resources, and one in 10 teachers (9%) had contributed to forum discussions. Overall, almost one-quarter of teachers (23%) reported that they did not participate in the online community. The main online activity, which was slow to start but did increase over time, was the downloading of resources.

Support: An important aspect of the Thinking Maths program and the successful engagement of teachers with it, was the support provided by the Thinking Maths facilitators, school leadership, and colleagues. Teachers were asked on five occasions, to what extent they felt well-supported over the last month to implement Thinking Maths strategies and activities in their classroom (see Addendum Figure 12A). On average, four in five teachers (81%) felt moderately or greatly supported by the Thinking Maths facilitators, while three in five teachers felt supported by colleagues (65%) and school leadership (60%).

In addition to indicating their level of engagement with the various Thinking Maths activities, teachers were asked which activities had been particularly useful. Over 230 comments were provided over the five feedback occasions. The majority of teachers felt that the Thinking Maths *resources and materials* were the most useful to them and the readings and textbook were also popular among teachers. Sharing their experiences with colleagues was also frequently nominated by teachers as being a useful activity, as was the reflective journal. A small number of teachers also provided additional comments about the value of the Thinking Maths Moodle, however, some teachers had difficulty logging in. Table 18 presents a selection of exemplar comments from Primary and Secondary teachers.

Based on these findings about the main elements of the program, it would be fair to say that the efficacy trial was conducted under optimal conditions, with the highest quality of delivery, fidelity and dosage provided by the Thinking Maths team, received by a highly motivated and receptive group of evaluation teachers (at the very least, evidenced by the 80% plus survey response rates). Under these conditions, all existing elements were considered to be important, although the online professional learning community much less so.

Table 18. Teachers' views about the most useful elements of the Thinking Maths program

Themes	Exemplar comments from Primary	Exemplar comments from Secondary
Hands-on activities and resources	*The resources and lesson ideas provided an immediately useful resource.* *Every single activity! This is because there are multiple entry points regardless of year levels and ability.*	*I found the activity on measuring the perimeter and area of your feet useful with my students to help understand these concepts better.* *Having task examples that I have experienced in the sessions gives me confidence to explore them with my students.*
Readings	*This month the readings have been particularly useful as I am about to teach that unit of work and the resources given at the latest session will be extremely useful and I am sure they will engage the students.*	*I found the readings to be particularly useful in my term planning as it provided me with a big picture understanding of the concepts I aimed to cover.*
Shared experience	*Sharing evidence enabled me to expand on the things I tried with ideas from my colleagues plus listen to new ideas that I can use.* *I think sharing evidence with my colleagues. Until you share you don't realise how many people are in your position worried to teach maths.*	*Being accountable for trying things in the class.*
Weekly journal	*Referring to journal weekly, to maintain … focus will be useful.*	*Writing in the reflective journal.*
Moodle	*The moodle resources have been great and I have been using them in the morning as warm ups.*	*Being able to download the resources from moodle to use in the future will be a great resource.*

Different contexts

2. How applicable and useful is the Thinking Maths approach in Primary and Secondary school settings?

Indicators for Support (3 items), Online professional learning community (3 items), Activities and resources (4 items), and the PL days (7 items) were constructed (mean item response), to investigate differences in how applicable and useful the Thinking Maths approach was in Primary and Secondary school settings. The standardised results presented in Table 19 indicate very little difference on the four indicators between teachers in Primary and Secondary contexts, although the views held by Primary teachers were consistently higher, albeit marginally, than the views held by Secondary teachers.

Table 19. Comparison of program elements in Primary and Secondary school settings

Thinking Maths program elements	Years 5-7 (n= 82)	Years 8-10 (n=35)	Difference	Effect Size Cohen's d	p-value t-test
PL sessions	92.0%	90.7%	1.3%	0.11	0.269
Support	77.7%	75.0%	2.7%	0.13	0.203
Activities & resources	70.7%	66.3%	4.4%	0.16	0.106
Professional learning community	46.8%	41.0%	5.9%	0.20	0.042

The most useful element, equally valued by over 90% of teachers in Primary (Years 5-7) and Secondary (Years 8-10) schools, was the PL days. Teachers consistently found the information delivered in the five PL days relevant and useful, supporting their learning and building their confidence about teaching maths.

The support provided, chiefly by the Thinking Maths facilitators, was marginally more valued by the Primary Years 5-7 teachers (78%) compared to the Secondary Years 8-10 teachers (75%). Similarly, the various activities and resources (e.g. readings, reflective journal, materials, sharing opportunities) were reasonably valued by 71% of Primary teachers and fewer Secondary teachers (66%).

The least valued element of the Thinking Maths program was the online professional learning community. Less than half of the Primary teachers (47%) and only two in five Secondary teachers (41%) found this element useful. This is reflected by 23% of teachers who reported that they had not participated in the online community. Unlike the other elements, the difference between Primary and Secondary teachers' views was statistically significant ($p<0.05$) to a moderate effect ($d=0.20$).

Barriers and facilitators

3. What are the barriers and facilitators to the effective implementation of Thinking Maths in middle-school classrooms in different contexts?

Thinking Maths teachers were asked through the PL Feedback Form, what aspects assisted and hindered them to implement the Thinking Maths strategies with their class during the intervening periods between the PL days. They were also asked to provide examples of how these enablers or barriers had impacted their practice. It should be noted that the general nature of the questions and the fact that teachers could only reflect within their specific context, meant that any differences identified between contexts, particularly with regard to Primary or Secondary contexts, were not explicit and needed to emerge from the thematic analysis. It was only through the additional data collected during the post-evaluation interview with the Thinking Maths facilitators that a clearer understanding of any differences could emerge. In their roles, the facilitators were not confined to a specific context and were in the unique position to provide support across all contexts, allowing them to establish behavioural norms and identify trends.

Enablers to implementation

Many teachers felt that the Thinking Maths resources were the most useful enablers in implementing the Thinking Maths program in their class. Teachers also identified specific parts of the program that helped them to successfully deliver the program in their class. These aspects were the questioning techniques, hands on practice, Thinking Maths strategies and activities, Thinking Maths online resources, the textbook, the session notes and the journal, and the engagement and support from the facilitators.

Teachers also felt that the questioning techniques taught in the program helped them implement Thinking Maths in their classroom. Teachers felt that these techniques improved their practice and encouraged student engagement and deep thinking.

Teachers thought that the shared experience of attending the PL days with a colleague helped them integrate the program at their site. They felt that this was achieved by being able to share ideas about teaching and also to plan with their colleagues. Having time to engage with the activities outside of the busyness of school and then having supported time to take the ideas and implement them back at school was important.

Teachers also felt that the hands-on practice components of the program helped with their confidence in the classroom and improved their practice. This in-turn improved student engagement in lessons and helped teachers to be more motivated and enthusiastic.

During the thematic analysis of teachers' comments, no obvious differences in Primary and Secondary contexts emerged in the results. There was a broader coverage of comments from Primary teachers, as expected, given the greater number compared to Secondary teachers. A summary of the implementation enablers that emerged from the thematic analysis, presented in Table 20, compares exemplar comments provided by Primary and Secondary teachers.

Table 20. Teachers' views about implementation enablers

Themes	Exemplar comments from Primary	Exemplar comments from Secondary
Resources and tasks	*Bringing the great tasks from the workshops back to my students - they look forward to them as much as I do!*	*Short thinking activities used as lesson starter of finisher when students tire of specified class work and need to reengage students to maths.*
Quality instruction	*I tried the perimeter and area of foot task due to the excellent facilitators' instruction- they made it clear to put the task into practice straight away.*	*The facilitators took great care to always give us experiences that modelled excellent teaching and learning opportunities.*
Teaching strategies	*I have got strategies and resources ready with me now. I have to look up the activities for the topics I am teaching now in class to make it more productive.* *Giving me an insight into the learning my students do and how to approach it.*	*The way I now pitch lessons to create more wonder and thinking by students rather than the teacher doing all the thinking.* *Repeated lessons over 5 training days have helped consolidate the "how to implement and question effectively".*
Good questioning technique	*Questioning more often and not jumping in so quickly to help students out.* *The facilitators have helped me focus on student aversion to problem solving and the need to work on questioning and being 'less helpful'.*	*Modelled questioning strategies have increased my strategy to support and scaffold students in the classroom.* *They have made me think about questioning techniques.*
Support	*The support and excellent motivation by facilitators to have ago was beneficial.*	*Enthusiasm that has been generated by the presenters.*
Paired-learning	*Having a colleague completing Thinking Maths with me and team teaching has help facilitate the implementation. We are able to bounce ideas off each other to successfully teacher activities we learnt over the program.*	*Having a colleague attend the training days with me and therefore keeping each other accountable.* *Having a colleague do the course as well, this means constant sharing of ideas and how things went*
Hands-on experience	*Participating in all the hands on activities has been amazing.*	*Enthusiasm and fun, understanding of content has increased, hands on-doing maths.*
Gaining confidence	*I have tried nearly all of them in my class and am gaining in confidence to implement these.* *I found the 'hexagon chain' exercise useful, as having something I could use straight away in the classroom made me feel confident.* *Sharing resources has boosted my confidence as well as some of my colleagues as we now have many engaging resources.*	*Seeing and doing the activities before then doing them with my class. This increases my confidence and understanding greatly!*
Engaged students	*It is the reaction of the students that prompts me to find new and interesting ways of delivering topics in the classroom.* *Engaging, hands-on practice. Improved student engagement has encouraged me to continue to develop less text-oriented maths lessons.*	*Having students who have such a passion for maths that they carry us on the journey with their enthusiasm so we keep going.*

Barriers to implementation

Teachers participating in the Thinking Maths program were also asked to describe any barriers that had hindered their implementation of Thinking Maths activities with their class during the intervening periods between the PL days.

Many teachers felt that they did not experience any barriers in implementing the program. However, of those that did, the main themes emerging from their responses included lack of time and resources, conflicts with other school activities, and conflicts between their lesson planning and the timing of the topics in the Thinking Maths program.

The majority of teachers felt that not having enough time was the biggest barrier to engaging with the program. Some of the ways that this was expressed included not having enough time to plan or to organise, to purchase or create the Thinking Maths teaching resources, or to implement the Thinking Maths strategies properly.

Teachers also reported that they struggled with the competing demands of other school activities. There were challenges for some when the curriculum area being taught, did not align to the topics taught in the Thinking Maths sessions. This was particularly evident in the Secondary context, where there was less flexibility about the sequence of topics.

A summary of the implementation barriers that emerged from the thematic analysis, presented in Table 21, compares exemplar comments provided by Primary and Secondary teachers.

Table 21. Teachers' views about implementation barriers

Themes	Exemplar comments from Primary	Exemplar comments from Secondary
No barriers	*No barriers experienced.* *No barriers to implementing activities.*	*No barriers. We have been allowed to spend money & have time to implement the program in our lessons.*
Time	*Having time to prepare the resources.* *Lack of time to prepare resources, kits etc.* *As a group of Year 6 and 7 teachers we don't have time to meet to look at developing lessons and resources for any subjects, including Maths.* *All the other projects underway and NAPLAN and meetings, interviews with parents and NEPS to be written up, just finding good quality time to focus on the curriculum.* *School based barriers have meant life at school is frantic. All are barriers to trialling new ideas in the classroom.*	*Secondary class - Tight timeline to teach a program - unable to swap topics around to trial Thinking Maths activities in class.* *Time to buy resources to implement activities. The general time pressures of teaching limit the time available to be creative when preparing new lessons, and implementation of the lessons in the classroom. These limitations are present and are not a result of Thinking Maths.* *Our time capacity to share good practice with colleagues is a limiting factor.*
Access to resources	*Not having enough concrete resources to implement some of the activities.* *Resourcing of maths equipment at our site.*	*It would be great to have more of these hands-on resources in high school without a teacher having to buy them for students out of their pay check.* *As a high school our resource cupboard is not as loaded with materials compared to primary.*
Topic misalignment	*Resources/topics covered have been different to those covered in the classroom.*	*The topics being taught are not aligned with Thinking Maths. eg: Algebra is put off until next year.*
Professional learning community Moodle access	*The barrier to accessing the online resources are time and passwords. This is a constant frustration!* *The Department internet consistently not working. Unable to access Moodle and other websites. These impact my practise because I can't teach my planned lesson and I feel unorganised.*	

During interview, Thinking Maths facilitators provided further insight about the possible differences in barriers experienced by teachers in the Primary and Secondary contexts. These insights are displayed in Table 22.

Table 22. Thinking Maths facilitators' views about differences in implementation barriers for Primary and Secondary teachers

Themes	Exemplar comments from Thinking Maths facilitators
Different levels of exposure	*Secondary students are only exposed to the 'treatment' for 3 hours per week as opposed to Primary students whose class teacher's shift in pedagogy was likely to impact more widely over the school day.*
Lack of resources	*The lack of concrete materials and equipment and shorter or inflexible lesson length may be a factor in the Secondary context.*
Lack of lesson flexibility	*Inflexible common programs which did not allow Secondary teachers to trial tasks as that topic had already been taught.*
Moodle	*We swapped over from Edmodo to Moodle this year and it didn't seem to be as widely used by teachers.*

4.2 Fidelity

From the collective voices of 117 teachers on up to five occasions (520 completed Feedback forms), along with regular updates from the Thinking Maths facilitators, the implementation of the Thinking Maths program was delivered as intended and went as planned. Moreover, the results suggest that the program was perceived by teachers as being of high quality, delivered by expert facilitators, and was found to significantly improve their confidence and understanding in teaching mathematics.

The Thinking Maths program involved teachers in a structured PL program, delivered by the program developers with measurable fidelity at the program level. However, the extent to which teachers took the new strategies and implemented them back in their mathematics classrooms was not prescribed beyond having a commitment to do so, making it difficult to assess fidelity of implementation at the classroom level.

Compliance and engagement

4. To what extent do teachers engage with the Thinking Maths program?

Apart from the three teachers who indicated early on their withdrawal from the study (non-compliers), the other 117 treatment teachers gave no indication that their class wasn't participating or otherwise receiving, at least to some extent, their learnings from the Thinking Maths program. The following attendance was recorded:

- 105 teachers (87%) attended all five days
- 12 teachers (10%) attended four days
- 1 teacher (1%) attended three days (missed days were non-consecutive), and
- 3 teachers (2%) did not attend and withdrew from the study before PL days commenced, due to a change in role.

Evidenced from the process evaluation results, the majority of teachers were highly engaged with the PL days, activities, and support. What was less clear was the extent to which teachers implemented their learnings back in their mathematics classrooms, although comments from teachers, such as, "I have since applied concepts I learned during the Thinking Math workshops to enhance my own activities (Primary teacher), are supportive.

Students also provided evidence of improved teaching practice. Table 23 provides several emergent themes from students' insights gathered in the post-survey about what they liked best about maths.

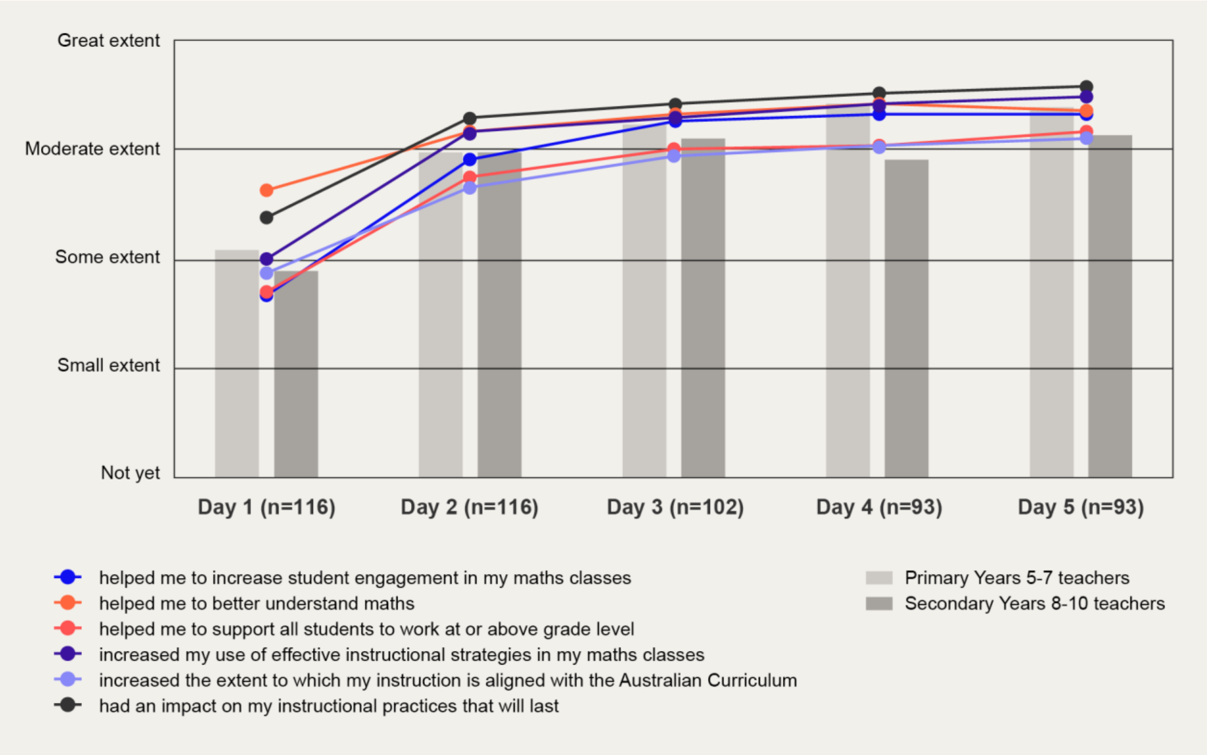
Table 23. Students' views on what they now like best about maths

Themes	Exemplar comments from Primary	Exemplar comments from Secondary
Problem solving and group work	*I like Thinking Maths because the problem-solving tasks help me with my maths. I also like teaching other children maths because I can hear what they think and how they solve the problems. I like making games and play them with other children, and I also like working in groups as it also helps me understand the maths that we are working on. (Year 6)*	*I like it when he challenges us with maths problems and we get to work in pairs to figure it out. (Year 8)*
Improved understanding and confidence	*He gives me and other students help when we need it and he helps us understand the problem a different way if our way of working out the problem isn't working for us and just getting us really confused. (Year 7)*	*The thing I like about maths at the moment is that I didn't understand it at the start but then got taught by my teacher and now I'm going really well. (Year 8)* *My teacher helped me gain more confidence when dealing with difficult maths problems because he helped my class and I go through it and explain step by step, so he makes sure that we all understand clearly. (Year 9)*
Purposeful questioning	*I like how my teacher makes math fun and he gives everyone a chance to express their thinking to the class if they see it a different way. (Year 7)*	
Engaging	*He also involves us students into the maths problems so he will use our names and make the problem sound funny so we laugh and do maths at the same time. (Year 7)*	*She makes maths really fun instead of just worksheets. (Year 8)*

Teachers responded to six items in the PL Feedback form to gauge the impact of their engagement with Thinking Maths. Figure 7 presents the results, which suggest that teachers engaged strongly with the program and this engagement improved throughout the program. By the final day, most teachers (92%) reported to a moderate or great extent that *Thinking Maths had an impact on their teaching practice that would last*. Most teachers also felt that their *understanding of mathematics had improved* (86%) and that Thinking Maths had *increased their use of effective instructional strategies in mathematics lessons* (91%). Similarly, 87% of teachers also believed that Thinking Maths had *helped them to increase student engagement in mathematics classes*.

Figure 7 also presents (as columns) the averaged response across the items as an overall measure of program impact. While, on average, all teachers indicated improvements in their mathematics teaching practices due to participating in Thinking Maths, teachers of Years 5-7 students reported great improvement by the fourth and fifth days in comparison to the Year 8-10 teachers. The difference by the last two days was statistically significant ($p<0.01$) and of medium effect size (Hedges' $g=0.50$). This suggests that the Thinking Maths program was more effective for Primary teachers than Secondary teachers.

Figure 7. Primary and Secondary school teachers' perceived impact of Thinking Maths



Suggestions for improvement

5. How can the Thinking Maths program be improved?

Teachers participating in the Thinking Maths program were asked how the program could be improved. Thematic analysis of over 239 comments collected at the completions of each PL day was undertaken. Ideas about PL improvement mostly came after the first two days. While the majority of teachers did not think that the PL could be improved, several other emergent themes are summarised in Table 24.

The small number of suggestions for improvement, included that the PL could be more focused on the Primary level, and that facilitators could circulate more. There was suggestion to scale the program to all teachers and that it should be 'compulsory'. Teachers also felt that extra time for planning, reflection, and discussion with colleagues both in the session and back at the school would be beneficial, as would extending the program beyond five days.

The results suggest that the overarching structure, logic and dosage of the program should not be altered, however there are avenues for adaptation of the program to best meet the changing and diverse needs of schools. Secondary teachers reported challenges in implementing Thinking Maths strategies when learning structures were less flexible, and when faced with competing demands of the curriculum. Given these results, there is scope for the program developers to consider the content and delivery that better supports Secondary teachers. Within the existing structure of the five PL days, there is scope to re-order or change the content focus of each day. There was also interest in expanding the days to include other content, for example, about assessment. Moreover, as new online resources and research-informed strategies become available, the content of the Thinking Maths program should be regularly refreshed. Content may also need to be tailored to the needs of specific contexts such as schools in communities with high proportions of Aboriginal or Torres Strait Islander populations.

Table 24. Teachers' views about areas for program improvement

Themes	Exemplar comments from Primary	Exemplar comments from Secondary
No change	*I don't know how it could be improved as the presenters use a variety of techniques and methods to implement their lessons, give us time to swap ideas, cover the curriculum and are generous in their handouts. The whole session is relevant to my role as a teacher. Thank you!* *I can't think of any ways for this program to be improved. I love attending these trainings because I always leave with something new to implement into my current practice.* *It has been one of the best, most effective and practical programs I have been involved in and it has really got me motivated and excited not only for student learning but my own. Never too late to learn new ways of approaching maths and it is so practical and easy to apply.*	*They couldn't really be improved from an understanding and in-depth activities. The presenters were outstanding.* *It has honestly been exceptional so far, I have enjoyed being challenged and adapting my own thinking and pedagogy around teaching math.*
Target Years	*Maybe doing some activities focused on the Year 5 and 6 work - seemed targeted towards high school more.* *I feel that the activities were more aimed at high school learning, so I guess a bit more explanation as to how it could be made appropriate for primary school.*	*Maybe having small groups that are secondary teachers, then year 6-7 and then junior primary etc.*
Pace	*When presenters were conferencing with table groups, others were left for a long time. So perhaps table groups could confer with each other.*	*There are times when the session could be moved on, especially when tasks have been completed by groups. I do like that the presenters give us challenges though as early finishers.*
Driving home the messages	*Finding a way to make it clear that category 1 students can demonstrate success with the right guidance. In many cases these are the students who are disengaged, they need engaging learning experiences and meaningful tasks. Having taught in a category 1 school for 7 years (and now category 5) I often find it difficult hearing teachers saying things like "my students will never be able to do that" before they even try... when in fact this is simply not true. I'm sure this should be a natural result of the course if everything goes to plan.*	*It reinforced the planning and delivery of mathematics that my teaching colleague and I are engaged in with our classes.*
Additional support from schools	*Schools to be more supportive of teachers attending - i.e maybe matching the day with a TRT day - so that teachers are able return and share with other classes. Time for spreading the information/knew findings.* *The program has been some of the best professional learning I have done. It would be worth keeping and having run over a longer period of time to cover more topics.* *Great program, I would come every month and gain more knowledge and skill every time.*	*My site releasing me and my colleague to follow up on strategies and plan feedback to other maths colleagues, as well as time to plan activities and trial them with students.*
Expand to all teachers	*If we really want to improve outcomes for students then it has to go beyond our group. I would put my hand up to assist in any way possible.* *I would love for this programme to be available to every teacher in SA, so that we can all be "expected" to present maths to students in this manner.* *The whole program should be videotaped and all teachers should have access to the videos as online PD. It would make an awesome course for teachers of all ages to do. High quality PD that is pitched at stretch and understanding which is just what we need in SA schools.*	*Offering it to as many maths teachers (secondary) as possible. Obviously cost prohibitive.*

6. What are the risks and challenges in expanding the Thinking Maths program to scale?

The main challenge in expanding the Thinking Maths program to scale relates to the face-to-face training. The current evaluation of Thinking Maths occurred under ideal conditions with the original developers delivering the program. The evaluation demonstrated that the concurrent running of four groups of 30 teachers was manageable. The program facilitators anticipate that this could be expanded to 35 without compromising quality and potential impact. Expanding the scale of the program beyond the training of 140 teachers per year, would require the training of new Thinking Maths facilitators. The Department currently does not have in place (to our knowledge) the documents or the processes to embark on a training model for new Thinking Maths facilitators. Under these more 'real-world' conditions, an evidence evaluation would be appropriate to ascertain the practical impact of the Thinking Maths program.

4.3 Control group activity

While 104 schools were initially randomised to control group, nine schools withdrew early on because the delayed-start competed with other planned activities. No data was available from these schools. The control group involved 187 teachers and the response rate to the baseline survey was 98% and to the post-survey was 86%.

Because randomisation occurred at the school level, there was no interaction between teachers in treatment schools (n=117) with teachers in control schools (n=187), notwithstanding the inadvertent interactions that may occur through other professional regional or sectorial activities. It should also be noted that a small number of schools in the control group may have had exposure to Thinking Maths as pilot schools in the previous years.

Business-as-usual was taken to be the myriad of normal activities and approaches that teachers use when teaching the Australian Curriculum Mathematics in South Australian government Primary and Secondary school classrooms. An indicative breakdown of the exposure of teachers and students to the Thinking Maths program compared to business-as-usual (based on Department policy) is presented in Table 25. It should be noted that it was expected that students in the control group had the same exposure to mathematics but under business-as-usual conditions.

Table 25. Potential exposure to Thinking Maths compared to 'business-as-usual'

Participant	Activities	Thinking Maths	Business as usual
Teacher	PL sessions (5 x 6hrs)	30 hrs	-
	Lesson preparation (per week)	2 hrs/week	2 hrs/week
	Presentation to the group (once)	2-5 hrs	-
	One reading per session with reflection (5 x 2-3 hours)	10-15 hrs	-
	Participating in online community (voluntary)	varies	-
Student	Primary students learning numeracy and maths (per week)	5 hrs/week	5 hrs/week
	Secondary students in maths class (per week)	3 hrs/week	3 hrs/week

5 Interpretation

It has been one of the best, most effective and practical programs I have been involved in and it has really got me motivated and excited not only for student learning, but my own. Never too late to learn new ways of approaching maths and it is so practical and easy to apply. (Primary teacher)

Key conclusions
4. The Thinking Maths program had a small positive effect, equivalent to one month of additional learning progress on Years 5-10 students' performance in the PATMaths achievement test, when compared to business-as-usual mathematics classes. These findings were not statistically significant[8].
5. Thinking Maths had a statistically significant impact equivalent to two months learning gain in Primary students' achievement on the PATMaths test. However, for Secondary students, there were two fewer months of learning progress.
6. Among a sub-sample of School Card[9] holders, the students (both Primary and Secondary) of Thinking Maths teachers had two additional months' progress in performance on the PATMaths test, however this finding was not statistically significant.
7. Thinking Maths had the largest statistically significant effect on mathematics teachers' pedagogical and content knowledge, as well as their professional identity and self-efficacy. The intervention also showed a small positive impact on teaching practices overall, with students reporting that Thinking Maths teachers were more likely to give extra help when needed, ask questions to check understanding and challenge their thinking. Findings showed similar gains on students' cognitive engagement, but no additional gains in metacognition. These results on student outcomes were not statistically significant. A small and statistically significant increase in students' mathematics anxiety was also found.
8. Teachers reported a number of benefits of this professional learning program including hands-on activities, expert modelling of metacognition strategies and teaching resources that supported teachers to directly transfer ideas to their classrooms. The process evaluation indicated that timetabled lessons, common tests, set text-books, and lack of time to plan were barriers to effective implementation in Secondary schools. Schools and program development should consider differences in learning contexts to better accommodate and support teachers to optimise implementation.

[8] Evidence for Learning will develop a plain English commentary on statistical significance to support readers in interpreting statistical results in our reports.

[9] The School Card scheme offers financial assistance to low-income families to assist with school fees for students attending government schools in South Australia.

This report presents the results of a large pragmatic randomised controlled trial of the *Thinking Maths* program, with an embedded process evaluation. The trial was robust and followed CONSORT standards, and the process evaluation used observation and examined the perspectives of teachers, students and program-delivery facilitators.

5.1 Impact evaluation

The impact evaluation found evidence of a small positive effect of the intervention. Students whose teachers attended the Thinking Maths program made more progress in mathematics than similar students in business-as-usual classrooms. The small positive effect can be estimated as equivalent to one month of additional learning progress. However, the effect was not statistically significant.

Across this whole cohort, there was a small positive effect on students' cognitive engagement and metacognitive strategies, which were not statistically significant. Students also showed a small and statistically significant increase in their mathematics anxiety.

Evidence also shows differences in students' achievement for Primary and Secondary school levels. Primary students (Years 5-7) of Thinking Maths teachers made a learning gain equivalent to two months while Secondary students (Years 8-10) had two fewer months of learning progress. Whilst there is no evidence of impact in Year 8, it is the decline in the Year 9 treatment group that is largely responsible for the few months of learning progress reported in the Secondary Years.

The program had a large positive impact on how teachers perceived their pedagogical content knowledge, particularly at the Primary school levels. The large effect reflects that teachers were directly involved in a professional learning program designed, primarily, to build capacity in this domain. The evaluation also found evidence of changing teaching practices. Teachers were deeply engaged and committed to implementing their learnings in the classroom to the extent that students reported a small improvement in effective teaching practice, but more-so in the Primary context. Since students were indirectly involved in the program through their teacher, it was anticipated that the level of impact on their achievement would be less, particularly given the short post-test timeline that did not allow changing teaching practices to have their full impact. In most schools, students were tested only two weeks after the last professional learning session. That there was a positive impact so shortly after the professional learning was completed is encouraging and may mean even greater gains in the future for students of these teachers.

5.2 Process evaluation

The process evaluation found that Thinking Maths – the professional learning days, the activities and resources, the professional learning community, and the professional support – provided a high-quality professional development experience for teachers, delivered as intended. The Thinking Maths facilitators were expert in their craft, and their choice to sequence intense, shared learning for a day with supported intervals in between was based on good educational practice and was highly effective. It gave teachers time to take-in and process the learning, reflect on their practice, and then implement and trial new practices in the classroom.

Thinking Maths teachers were highly positive about the program and advocated for its wide-spread rollout. There was no obvious difference in the types of barriers and enablers reported by Primary and Secondary teachers, nor in their engagement with the program, that might shed light on the differences in outcomes at the student level. However, significant differences did emerge by the end of training, more-so for Primary teachers than Secondary, about the extent that the program had increased their mathematics understanding, their use of instructional strategies, and levels of student engagement. A correlation between student and teacher primary and secondary outcomes provided additional evidence that a stronger positive impact was experienced in the Primary schooling context.

The Thinking Maths facilitators identified the following possible causal factors, in their unique position of providing support across all contexts allowing them to establish behavioural norms.

- Dosage: Secondary students were only exposed to the 'treatment' of changed teaching for 3 hours per week as opposed to Primary students whose class teacher's shift in pedagogy was likely to impact more widely over the school day.
- Resources: The lack of concrete materials and equipment, as well as shorter or inflexible lesson length may be a factor in the Secondary context.
- Flexibility: Fixed curriculum programming may not have allowed Secondary teachers the flexibility to trial tasks if they were off-topic.

5.3 Discussion

The primary analysis for this evaluation was to determine the impact of the Thinking Maths program on middle-school students' mathematics achievement. The analysis found a small positive effect on students in the Primary Years 5-7, equivalent to two months learning gain. The analysis also found a small negative effect on students in the Secondary Years 8-10, equivalent to two fewer months of learning progress. These sub-group results were statistically significant but in the proximity of the estimated minimum detectible effect size of 0.15, and in isolation, could be dismissed as being within the range of random error. Indeed, these findings are supported by the reports of teachers who acknowledged that learnings from the program were only partially implemented in their classrooms prior to post-testing.

The Teaching Maths program is a professional learning program designed to build teachers' pedagogical content knowledge. As such, it was anticipated that the level of impact of the program may be greatest on teachers and their teaching practices, but less-so on students and their mathematics achievement, particularly given that most students were tested two weeks after the intervention. This is evident in the results presented in the evaluation logic model (Figure 5), which shows a progressive reduction in effect size from teacher outcomes to student achievement.

Triangulating wider evidence gathered through the secondary analyses and process evaluation, a clear line of diminishing influence emerges – from the program inputs, to changes in teachers' knowledge and classroom practice, through to student mathematics outcomes. The evidence suggests that a goal of Thinking Maths – to improve teachers' pedagogical content knowledge – was achieved. Thinking Maths teachers, reportedly, were more likely to identify students' learning challenges, support creative and critical thinking, use questioning to diagnose students' conceptual misunderstandings, and differentiate teaching of the Australian Curriculum Mathematics. Students were more likely to gain higher mathematics outcomes in Primary classes where the teacher reported strong pedagogical content knowledge (r=0.23), and in Secondary classes where the teacher reported high professional identity and self-efficacy (r=0.40).

The differences in the Primary and Secondary teacher cohorts suggest that the impact was more positive for Primary teachers and students. A possible reason for these differences was attributed by the Thinking Maths facilitators to the structural differences in the teaching of mathematics. The constraints typically experienced by Secondary school teachers (timetabled lessons, common tests, set text-books, less flexibility), were reported as challenges to implementing Thinking Maths practices effectively. This, in turn, may have generated uncertainty in the mathematics classroom and impacted upon students' anxieties about mathematics and lower outcomes. How the structures of Secondary schools might hinder teachers implementing their learnings from the Thinking Maths program warrants further investigation.

While this evaluation focussed on the delivery of the program to teachers, what teachers actually embodied and implemented in the classroom, was beyond the scope of this evaluation, and in fact, beyond the scope of the program itself. Accordingly, the extent to which these differences led to the small negative effect on mathematics outcomes for students in Secondary Years 8-10 cannot be attributed only to Thinking Maths and may be due to many other factors specific to the teaching of mathematics in the Secondary context. To be clear, while Secondary students' outcomes in the Thinking Maths group were lower than the outcomes of their 'business-as-usual' peers, the students in both groups showed positive learning gain over the year.

5.4 Limitations

The efficacy trial was undertaken within South Australian government schools and did not include Catholic or independent schools. However, their distribution geographically and other demographic characteristics suggests that results may be applicable across this education system. As South Australia is the only state to still retain Year 7 in Primary school, findings are not generalisable to other Australian states or territories. The current findings may also have a time limitation because of the anticipated move of Year 7 into Secondary school to align with the other states and territories.

The process evaluation mainly focused on the quality of delivery (program inputs) and its direct impact on teachers, at the school level. Less focus was given to how Thinking Maths teaching differed from business-as-usual teaching, to explain the impact evaluation, focused on outcomes at the student level. Should the Thinking Maths program be expanded to scale, it would need more facilitators and a new program of facilitator training, leading to another level of evaluation, at the system level.

The close timing of the post-test, only two to four weeks after teachers completed the Thinking Maths training, did not allow sufficient time for teachers to fully implement the Thinking Maths strategies in their teaching practice, or have a chance to use the activities and resources in their classroom prior to the testing. A longer period may have resulted in larger impact. There would be value in analysing the PAT Maths 2018 scores of the current students of the 2017 Thinking Maths teachers to assess if changes in teaching practice are lasting.

While participants (schools, teachers, students) were blind to treatment allocation until after completion of pre-assessment at baseline, they were not blind to their allocation during the evaluation period or at post-assessment. The extent to which the control group maintained business-as-usual mathematics teaching practices was not assessed.

Moreover, while we were not aware of other programs running concurrently in schools to support student mathematics achievement, we cannot disentangle any impact from the broader mathematics education and STEM agendas.

In order to minimise impact on teachers and school processes, the Student Survey was administered online by the classroom teacher using a generic survey link provided by ACER. In a small number of schools with internet access issues, paper-based surveys were administered. These were collected by the teacher and sent to ACER for data extraction. There was no suggestion that teachers influenced the way students responded, however, this cannot be completely ruled out.

5.5 Future research and recommendations

The current findings are promising. Given the large effect size on teachers' pedagogical development and their positive feedback, particularly in the Primary school context, we recommend a larger-scale effectiveness RCT of the Thinking Maths program. This could include schools in other sectors (Catholic, independent) and in other states and territories where Year 7 is part of Secondary school. A parallel efficacy trial of a facilitator training model (e.g. train-the-trainer) could also be included to ensure that quality of program delivery is monitored at all levels.

As suggested by a participant, the number of professional learning days could be extended to include other content. In particular, Thinking Maths does not currently include the important area of assessment of student learning. Given that the Department have mandated the annual use of PAT Maths, it would be prudent to build the capacity of all teaching staff in how to use PAT data and the reporting platform, in further efforts to differentiate learning, identify learning gaps, and improve student numeracy outcomes.

Further investigation into how Thinking Maths is implemented by teachers in Secondary school would be helpful to better understand the current findings. There would be value in further investigating:

- how 'transferable' Thinking Maths strategies (e.g. purposeful questioning) are to other subject areas;
- the extent concrete resources are lacking in Secondary classrooms; and
- the impact of structural differences and system constraints.

The efficacy trial was conducted under optimal conditions, with the highest quality of delivery, fidelity and dosage provided by the Thinking Maths team, received by a highly motivated and receptive group of participating teachers. Under these conditions, all existing elements of the program were considered to be important, although the online professional learning community much less so. Accordingly, consideration could be given to the effectiveness of the online professional learning community Moodle as a discussion form, which was not well used. A content management system (CMS), where resources can be shared and information updates provided, may be a better approach for organising and providing access to a rich and growing library of classroom mathematics resources. If made public, this would also overcome the need to 'login', which presented as a barrier for some teachers.

Further research could also be undertaken to better understand and meet the needs of diverse schooling contexts, such as schools in remote communities with high proportions of Aboriginal or Torres Strait Islander populations.

This evaluation was Evidence for Learning's first project. As such, the processes and approaches evolved as the project progressed. ACER developed the initial proposal and evaluation protocol in the absence of final reporting templates and requirements. In effect, we collected far more data than was required and beyond the scope of this report. This affords the opportunity to conduct further analyses using three-level hierarchical linear models that take a more exploratory approach to identify the factors that influence the effectiveness of Thinking Maths and student mathematics outcomes. Likewise, the qualitative data has only undergone preliminary thematic analysis, with the opportunity to more deeply identify thematic trends in sub-groups and triangulate these against the models. Further exploiting the high-quality, rich and extensive data-set will value-add to the project, its outcomes, and potentially its impact in shaping policy and practice.

References

ACER 2011, *Interpreting ACER Test Results*, Melbourne: Australian Council for Educational Research. Available at: <https://www.acer.org/tsa/progressive-achievement-tests-pat>

Bloom, HS, Richburg-Hayes, L & Black, AR 2007, 'Using covariates to improve precision for studies that randomize schools to evaluate educational interventions', *Educational Evaluation and Policy Analysis*, *29*(1), pp.30-59.

Boaler, J 2002, *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning*, Routledge.

Boaler, J 2015, *The elephant in the classroom: Helping children learn and love maths*, Souvenir Press.

Boaler, J & Humphreys, C 2005, *Connecting mathematical ideas: Middle school video cases to support teaching and learning*, Portsmouth, NH: Heinemann.

Claxton, G, 2012, *Building Learning Power,* Victoria: Hawker Brownlow.

Cunningham, T 2010, *Power and Sample Size for Three-Level Cluster Designs,* Unpublished Doctoral Thesis, Virginia Commonwealth University. http://scholarscompass.vcu.edu/etd/148/

Daraganova, G, Edwards, B & Sipthorp, M 2013, *Using National Assessment Program Literacy and Numeracy (NAPLAN) Data in the Longitudinal Study of Australian Children (LSAC),* Department of Families, Housing, Community Services and Indigenous Affairs.

DECS 2010. *South Australian Teaching for Effective Learning (TfEL) Framework Guide,* Adelaide: South Australian Department of Education and Children's Services. https://www.decd.sa.gov.au/sites/g/files/net691/f/tfel_framework_guide_complete.pdf

Dix, KL, Keeves, JP, Slee, PT, Lawson, MJ, Russell, A, Askell-Williams, H, Owens, L, Skrzypiec, G & Spears, B 2010, *The KidsMatter Primary Evaluation: Technical Report and User Guide*, Shannon Research Press, Adelaide.

Dweck, CS 2000, *Self-theories: Their role in motivation, personality, and development*, Psychology Press.

Dweck, CS, Walton, GM & Cohen, GL 2014, *Academic tenacity: Mindsets and skills that promote long-term learning,* Bill & Melinda Gates Foundation.

Education Evaluation Commission (EEC) 2016, 'Chapters 7 and 8: Student and Teacher views', *Kingdom of Saudi Arabia National Assessment Program Report to the Ministry of Education 2015 National Assessment Program Mathematics and Science Grade 3 and Grade 6* (pp.137-185), Prepared by the Australian Council for Educational Research, ACER Press, Melbourne.

EEF 2015, *Policy on analysis for EEF evaluations,* Education Endowment Fund, UK. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf

Evidence for Learning 2017, *Guidance on cost evaluation*, Evidence for Learning, Sydney.

Eldridge, SM, Ashby, D & Kerry, S 2006, ‘Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method’, *International Journal of Epidemiology*, *35*(5), pp.1292-1300.

Estimation 180 2016, *Estimation 180: Building number sense one day at a time,* www.estimation180.com

Flay, BR, Biglan, A, Boruch, RF, Castro, FG, Gottfredson, D, Kellam, S, Mościcki, EK, Schinke, S, Valentine, JC & Ji, P 2005, ‘Standards of evidence: Criteria for efficacy, effectiveness and dissemination’, *Prevention Science*, *6*(3), pp.151-175.

Fredricks, JA & McColskey, W 2012, ‘The measurement of student engagement: A comparative analysis of various methods and student self-report instruments’. In *Handbook of research on student engagement* (pp. 763-782), Springer, Boston, MA.

Hayes, RJ & Moulton, LH 2017, *Cluster Randomised Trials, 2nd Edition*, Chapman and Hall/CRC.

Hedges, LV & Hedberg, EC 2007, ‘Intraclass correlation values for planning group-randomized trials in education’, *Educational Evaluation and Policy Analysis*, *29*(1), pp.60-87.

Hedges, L V 2007, ‘Effect sizes in cluster-randomized designs’, *Journal of Educational and Behavioral Statistics, 32*(4), 341-370.

Higgins, S, Katsipataki, M, Kokotsaki, D, Coe, R, Major, LE & Coleman, R 2013, *The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit: Technical Appendices*, London: Education Endowment Foundation.

Hill, CJ, Bloom, HS, Black, AR & Lipsey, MW 2008, ‘Empirical benchmarks for interpreting effect sizes in research’, *Child Development Perspectives*, *2*(3), pp.172-177.

Humphrey, N, Lendrum, A, Ashworth, E, Frearson, K, Buck, R & Kerr, K 2016, *Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook*, Education Endowment Foundation (Ed.).

Hummel-Rossi, B & Ashdown, J 2002, ‘The state of cost-benefit and cost-effectiveness analyses in education’, *Review of Educational Research*, *72*(1), pp.1-30.

Hutchison, D & Styles, B 2010, *A guide to running randomised controlled trials for educational researchers*, Slough: NFER.

Killian, S 2015, ‘8 Strategies Robert Marzano and John Hattie Agree On’, *Australian Society of Evidence Based Teaching*. www.evidencebasedteaching.org.au/robert-marzano-vs-john-hattie/

Levin, HM 1995, ‘Cost-effectiveness Analysis’, In M Carnoy (Ed) *International encyclopaedia of economics of education* (Vol. 2), Pergamon Press.

Little, RJ 1988, ‘A test of missing completely at random for multivariate data with missing values’, *Journal of the American Statistical Association*, 83, pp.1198-1202.

Meyer, D 2016, *Dy/dan* [Blog]. http://blog.mrmeyer.com/

Miller, G 2009, ‘Visible Learning by John Hattie. Summary by Gerry Miller’. http://www.esc5.k12.in.us/index.php/inside-wvec/documents-and-forms/visible-learning/741-summary-by-gerry-miller/file

Moher, D, Hopewell, S, Schulz, KF, Montori, V, Gøtzsche, PC, Devereaux, PJ & Altman, DG 2010, 'CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials', *BMJ*, 340, c869.

Murray, DM 1998, *Design and analysis of group-randomized trials*, New York: Oxford University Press.

NRICH 2016, *NRICH enriching mathematics.* http://nrich.maths.org/

OECD 2014, *PISA 2012 technical report*, OECD, Paris.

Pintrich, PR & DeGroot, E 1990 April, 'Quantitative and qualitative perspectives on student motivational beliefs and self-regulated learning', In *Annual Meeting of the American Educational Research Association, Boston, MA* (Vol.128).

Redfield, DL & Rousseau, EW 1981, 'A meta-analysis of experimental research on teacher questioning behavior', *Review of educational research*, *51*(2), pp.237-245.

Scootle,2016, *Scootle Community*. www.scootle.edu.au

Shriver, EK 2009, *The Evaluation of the Media-Smart Youth® Curriculum*, National Institute of Child Health and Human Development, US. https://www.nichd.nih.gov/publications/pubs/documents/msy_evaluation_report_final.pdf

Sullivan, P 2011, *Teaching mathematics: Using research-informed strategies*. Camberwell Victoria: ACER Press.

Sullivan, PA, Clarke, DM & Clarke, BA 2013, *Teaching with tasks for effective mathematics learning*, New York NY USA: Springer.

Timperley, H 2008, *Teacher Professional Learning and Development*, Educational Practices Series – 18, International Bureau of Education, International Academy of Education. http://www.ibe.unesco.org/en/document/teacher-professional-learning-and-development-educational-practices-18

Torgerson, CJ & Torgerson, DJ 2013, *Randomised trials in education: An introductory handbook,* London: Education Endowment Foundation.

Torgerson, C, Ainsworth, H, Buckley, H, Hampden-Thompson, G, Hewitt, C, Humphry, D, Jefferson, L, Mitchell, N & Torgerson, D 2016, *Affordable online maths tuition: Evaluation report and executive summary*, Project Report, Education Endowment Foundation. Durham Research. http://dro.dur.ac.uk/19387/

Tymms, P 2004, 'Effect sizes in multilevel models', In I Schagen & K Elliot (Eds), *But what does it mean? The use of effect sizes in educational research*, Slough: National Foundation for Educational Research, pp.55-66.

Van de Walle, JA 2014, *Teaching students-centred mathematics: Developmentally appropriate instruction for Grades 6-8*. United States of America: Pearson Education.

You Cubed 2016, *You Cubed at Stanford University*. www.youcubed.org/

Zopluoğlu, C 2012, 'A cross-national comparison of intra-class correlation coefficient in educational achievement outcomes', *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 3*(1), pp.242-278.

Appendix A: Evidence for Learning cost rating and effect size to months progress

Cost ratings are based on the approximate cost per student per year of implementing the intervention over three years. Cost ratings are awarded as follows:

Cost rating	Description
$	Very low: up to about $4000 per year per class of 25 students, or less than $160 per student per year.
$$	Low: $4001 to $8000 per year per class of 25 students, or up to about $320 per student per year.
$$$	Moderate: $8001 to $30000 per year per class of 25 students, or up to about $1200 per student per year.
$$$$	High: $30001 to $50000 per year per class of 25 students, or up to $2000 per student per year.
$$$$$	Very high: over $50000 per year per class of 25 students, or over $2000 per student per year.

The following table of effect size to months progress was provided and used in accordance with Evidence for Learnings' requirements, and is based on that used by the Education Endowment Fund, recently revised from the original table established by the EEF (Higgins et al., 2013). Source: educationendowmentfoundation.org.uk/help/projects/the-eefs-months-progress-measure. Also note that it differs from the conversion table used in the Toolkit: see evidenceforlearning.org.au/the-toolkit/about/#months-impact

Effective size: from	to	Months impact	Description
-0.04	0.04	0	Very small or no effect
0.05	0.09	1	Small
0.10	0.18	2	Small
0.19	0.26	3	Moderate
0.27	0.35	4	Moderate
0.36	0.44	5	Moderate
0.45	0.52	6	Large
0.53	0.61	7	Large
0.62	0.69	8	Large
0.70	0.78	9	Very large
0.79	0.87	10	Very large
0.88	0.95	11	Very large

Appendix B: Security Padlock rating

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	Power	Attrition*					
5 🔒	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%	5	➡	Adjustment for Balance [ ] Adjustment for threats to internal validity [ ]		
4 🔒	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%					4
3 🔒	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%					
2 🔒	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%					
1 🔒	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%					
0 🔒	No comparator	MDES > 0.6	>50%					

Addendum

See separate document, available at: evidenceforlearning.org.au/lif/our-projects/thinkingmaths/

Overview of secondary outcome measures

- HLM model specifications
- Overview of process measures
- Student Survey
- Teacher Survey
- Professional Learning Feedback Form
- Thinking Maths Professional Learning Session, Day 1